



East West Encounter: Second International Conference on Meaning ⇔ Text Theory
 Восток – Запад: Вторая международная конференция по модели «Смысл ⇔ Текст»



Alonso Ramos M.
 Anctil D.
 Apresjan Ju.
 Babina O.
 Barque L.
 Bélanger P.
 Boguslavskaya O.
 Boguslavsky I.
 Bojalkowska K.
 Bolshakov I.
 Dikovsky A.
 Dobrovol'skij D.
 Feldstein R.
 Gavrilova V.
 Gerdes K.
 Gladky A.
 Hi-yon Yoo.
 Iomdin L.
 Iordanskaja L.
 Janssen M.
 Kahane S.
 Kibrik A. A.
 Kibrik A. E.
 Kittredge R.
 Kopotev M.
 Krylov S.
 Lareau F.
 Lesmo L.
 Letuchiy A.

East West Encounter:

Second International Conference on
 Meaning ⇔ Text Theory

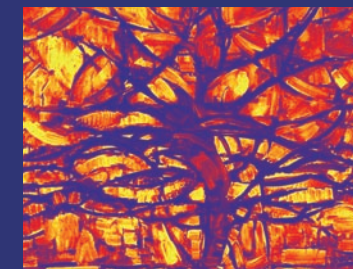
Levontina I.
 Lyashevskaya O.
 Lönngren L.
 Mel'čuk I.
 Mikaelian I.
 Milićević J.
 Nasr A.



Восток – Запад:

Вторая международная конференция
 по модели «Смысл ⇔ Текст»

Paducheva E.
 Panina A.
 Partee B.
 Plungian V.
 Polguère A.
 Pounder A.
 Rakhilina E.
 Reuther T.
 Robaldo L.
 Rudnitskaya E.
 Saloni Z.
 Sanromán B.
 Serdobolskaya N.
 Shaljapina Z.
 Sheremetyeva S.
 Shmelev A.
 Smilga B.



Specht V.
 Steinlin J.
 Trub V.
 Tutin A.
 Uryson E.
 Wanner L.
 Xrakovskij V.
 Yanko T.
 Zabokrtsky Z.
 Zalizniak Anna.

East West Encounter:

Second International Conference on Meaning $\Leftrightarrow$ Text Theory

Восток – Запад:

Вторая международная конференция по модели «Смысл $\Leftrightarrow$ Текст»

VINOGRADOV RUSSIAN LANGUAGE INSTITUTE OF RAS
LINGUISTICS INSTITUTE OF RAS
INSTITUTE FOR INFORMATION TRANSMISSION PROBLEMS OF RAS

East West Encounter:

Second International Conference on
Meaning $\Leftrightarrow$ Text Theory

Edited by
Ju. D. Apresjan and L. L. Iomdin



SLAVIC CULTURE LANGUAGES PUBLISHING HOUSE
MOSCOW 2005

ИНСТИТУТ РУССКОГО ЯЗЫКА ИМ. В. В. ВИНОГРАДОВА РАН
ИНСТИТУТ ЯЗЫКОЗНАНИЯ РАН
ИНСТИТУТ ПРОБЛЕМ ПЕРЕДАЧИ ИНФОРМАЦИИ РАН

Восток – Запад:

Вторая международная конференция
по модели «Смысл $\Leftrightarrow$ Текст»

Ответственные редакторы:
Ю. Д. Апресян, Л. Л. Иомдин



ЯЗЫКИ СЛАВЯНСКОЙ КУЛЬТУРЫ
МОСКВА 2005

ББК 81.031
В 76

В 76 East — West Encounter: Second International Conference on Meaning $\Leftrightarrow$ Text Theory / Vinogradov Russian Language Institute of RAS; Linguistics Institute of RAS; Institute for Information Transmission Problems of RAS; Ed. by Ju. D. Apresjan and L. L. Iomdin. — М.: Slavic Culture Languages Publishing House, 2005. — 592 p.

Восток — Запад: Вторая международная конференция по модели «Смысл $\Leftrightarrow$ Текст» / Ин-т русского языка им. В. В. Виноградова РАН; Ин-т языкознания РАН; Ин-т проблем передачи информации РАН; Отв. ред. Ю. Д. Апресян, Л. Л. Иомдин. — М.: Языки славянской культуры, 2005. — 592 с.

ISBN 5-9551-0094-6

This volume comprises all the contributions presented at the Second International Conference on the Meaning-Text Theory. They range over a vast domain of linguistic issues treated within the Meaning-Text framework. Most of the papers handle various semantic problems (definitions of lexical items, semantic actants and semantic roles, diatheses and their alternations, metaphor in the Meaning-Text perspective and the like). Other frequented topics are syntax and morphology, especially valency theory and the problems of syntactic representations for minor type sentences; lexicographic treatment of various word classes and the interaction between dictionary and grammar; the theory of lexical functions, paraphrase relations and rules; computer-aided investigation of lexical units, in particular, their dictionary definitions; the interrelations between linguistics and mathematics; linguistic typology; the place and role of the MTT in modern linguistics.

Этот том включает все доклады, прочитанные на Второй международной конференции по теории «Смысл — Текст». Они охватывают широкую область лингвистических проблем, рассматриваемых с точки зрения этой теории. Большинство работ посвящено различным вопросам семантики (толкование лексических единиц, семантические актанты и семантические роли, диатезы и мены диатез, метафора в модели «Смысл — Текст» и т. п.). Другие часто обсуждаемые темы относятся к области синтаксиса и морфологии. Это теория валентностей и проблемы синтаксического представления конструкций малого синтаксиса; лексикографическая разработка различных классов слов и взаимодействие словаря и грамматики; теория лексических функций, перифрастические отношения и правила; исследование лексических единиц, в частности, их словарных толкований, с помощью компьютера; отношения между лингвистикой и математикой; лингвистическая типология; место теории «Смысл — Текст» в современной лингвистике.

ББК 81.031

ISBN 5-9551-0094-6

© Авторы, 2005

© Языки славянской культуры, 2005

Contents

<i>Anctil D.</i>	
The notion of lexical error in the framework of <i>Explanatory combinatorial lexicography</i>	9
<i>Apresjan Ju.</i>	
Prolegomena to systemic lexicography	20
<i>Barque L., Nasr A.</i>	
Computing semantic relations on structured lexical definitions	30
<i>Bélanger P., Kittredge R.</i>	
Paraphrasing with space constraints: Linguistic operations in journal abstracting	39
<i>Boguslavskaya O.</i>	
Противопоставление объективной и субъективной причины в семантике причинных глаголов	53
<i>Boguslavsky I.</i>	
Actant properties of adverbial derivatives	65
<i>Bojalkowska K., Saloni Z.</i>	
On the “subject” of Polish constructions with “adverbial participles” in the framework of Meaning-Text theory	77
<i>Bolshakov I.</i>	
Stable coordinated pairs as a specific resource of language	86
<i>Dikovsky A., Smilga B.</i>	
Semantic roles and diatheses for functional discourse plans	98
<i>Dobrovol'skij D.</i>	
Semantic motivation of lexical functions and its limits (The combinatorial profile of the Russian degree modifier <i>črezvyčajno</i>)	110
<i>Feldstein R.</i>	
On the aspectual derivation of “dual simplexes”	122
<i>Gavrilova V.</i>	
Взаимодействие категорий переходности и залога в русском языке (Производящие и инициирующие переходные глаголы и их возвратные корреляты)	132
<i>Gerdes K., Kahane S, Hi-yon Yoo.</i>	
Of the descriptive adequacy of topology	146
<i>Gladky A.</i>	
Размышления о взаимодействии лингвистики и математики	159
<i>Iomdin L.</i>	
A hypothesis of Two Syntactic Starts	165
<i>Iordanskaja L., Polguère A.</i>	
Hooking up syntagmatic lexical functions to lexicographic definitions	176

<i>Janssen M.</i>	
Between inflection and derivation:	
Paradigmatic lexical functions in morphological databases.....	187
<i>Kahane S., Lareau F.</i>	
Meaning-Text unification grammar: modularity and polarization.....	197
<i>Kibrik A. A.</i>	
On the use of “quasi-grammeme” in Athabaskan	208
<i>Kibrik A. E.</i>	
Игорь Мельчук и наши экспедиции	219
<i>Kopotev M.</i>	
ГУЛЯТЬ ТАК ГУЛЯТЬ: between grammar and dictionary	225
<i>Krylov S.</i>	
Влияние МСТ на общую лингвистику (в свете подведения	
предварительных итогов «семантической революции»).....	237
<i>Lesmo L., Robaldo L.</i>	
Underspecification of quantifier scope in MTT	248
<i>Levontina I., Shmelev A.</i>	
The particles one cannot do without	258
<i>Lönngren L.</i>	
Русские прилагательные-определения: валентностный анализ.....	268
<i>Lyashevskaya O.</i>	
Наши соседи по старой даче: предложные актанты	
и модификаторы у предметных и реляционных имен.....	277
<i>Mel'čuk I.</i>	
MTT and the conceptual apparatus of modern linguistics:	
towards a better description of case and voice in Maasai	285
<i>Mikaelian I., Zalizniak Anna.</i>	
The Russian conjunction <i>A</i> revisited.....	304
<i>Milićević J.</i>	
Semantic paraphrasing in Meaning-Text linguistic models	316
<i>Paducheva E.</i>	
Соответствие «смысл ⇔ текст» в исторической перспективе.....	330
<i>Panina A.</i>	
К описанию валентностных свойств ЛФ семейства	
Oper _i , Func _i , Labor _{ij} (на японском материале).....	350
<i>Partee B.</i>	
Diathesis alternations and NP semantics.....	361
<i>Plungian V., Rakhilina E.</i>	
Приземлиться и промахнуться: семантические	
механизмы синтаксических ограничений.....	374
<i>Pounder A.</i>	
Process-and-paradigm morphology:	
A separationist sign-based framework	383
<i>Reuther T.</i>	
Метафора в модели «Смысл ⇔ Текст».....	394

<i>Rudnitskaya E.</i>	
An attempt of implementing the Generative Grammar	
Island Constraints in the «Meaning — Text» Model	405
<i>Sanromán B.</i>	
Individual-level and stage-level predicates:	
The Spanish emotion nouns	417
<i>Serdobolskaya N., Letuchiy A.</i>	
Object or situation: the factors of expressing	
the arguments of mental verbs in Russian.....	432
<i>Shaljapina Z.</i>	
Многофункциональность аппарата лексических функций	
и некоторые возможности его применения и развития	447
<i>Sheremetyeva S., Babina O.</i>	
Meaning-Text theory for textual input analysis	
and proofing in a generation system	458
<i>Specht V.</i>	
A MTT-based parser for German.....	467
<i>Steinlin J., Kahane S., Polguère A.</i>	
Compiling a “classical” explanatory combinatorial	
lexicographic description into a relational database	477
<i>Trub V.</i>	
Противопоставительная структура как средство	
семантической интерпретации отрицания.....	486
<i>Tutin A.</i>	
Annotating lexical functions in corpora:	
Showing collocations in context	498
<i>Uryson E.</i>	
Connective particles and theory of valence	510
<i>Wanner L., Alonso M.</i>	
What type of entity is a lexical function?.....	518
<i>Xrakovskij V.</i>	
The concept of diatheses and voices in early XXI century	529
<i>Yanko T.</i>	
Интонация иллокуций. Грамматика и словарь.....	538
<i>Žabokrtský Z.</i>	
Resemblances between Meaning ⇔ Text theory	
and functional generative description	549
Сведения об авторах	558

The Notion of Lexical Error in the Framework of Explanatory Combinatorial Lexicology

Dominic Anctil

Département de didactique,
OLST — Département de linguistique et traduction, Université de Montréal
dominic.anctil@umontreal.ca

Lexical mastery constitutes a central component of language mastery. Nevertheless, few studies dealing with linguistic errors — notably in written language — consider the lexical aspect of such errors. Furthermore, the actual notion of lexical error seems vague, both in linguistics and language teaching, and no precise definition of it has been proposed. As a result, many errors considered as syntactic or morphological are in our opinion lexical.

Using *Explanatory Combinatorial Lexicology*'s principles of lexical description as a guideline, we will try to characterize clearly the notion of lexical error. Based on an exhaustive presentation of the properties of a lexical unit, we can observe in which ways they can be violated or transgressed to cause lexical errors. This approach will provide us with all the necessary tools to propose a descriptive typology of lexical errors. We will finally put our typology to test by describing the lexical errors found in a text corpus.

Понятие лексической ошибки в рамках толково-комбинаторного словаря. Владение лексикой составляет центральный компонент владения языком. Однако лишь немногие исследования, посвященные языковым ошибкам, — в частности, в письменном языке, затрагивают лексический аспект этих ошибок. Более того, само понятие лексической ошибки остается весьма расплывчатым, как в лингвистике, так и в преподавании языков; до сих пор отсутствует его точное определение. В результате многие ошибки, которые по нашему мнению являются лексическими, обычно квалифицируются как синтаксические или морфологические. Основываясь на принципах описания лексики, выдвинутых в толково-комбинаторной лексикологии, мы пытаемся более точно охарактеризовать понятие лексической ошибки. Исходя из исчерпывающего представления свойств лексических единиц, можно установить, какими способами эти свойства могут нарушаться, что ведет к лексическим ошибкам. Такой подход дает все необходимые инструменты, чтобы установить типологию лексических ошибок. Эта типология проверяется на экспериментальном материале — на множестве лексических ошибок, найденных в корпусе текстов.

Introduction

In 1995, a study by Roy, Lafontaine & Legros revealed that more than a third of Québec's undergraduates studying in French didn't have sufficient mastery of the written language, especially regarding grammatical relations and orthography. A more recent study (Lefrançois, 2004) also showed that syntactic, lexical and textual coherence problems are, among all types of errors, those less often identified and those most often corrected inaccurately in both writing and revision tasks. Because the authors were more interested in the grammatical mastery of the subjects, they supply very little information concerning the three specific aspects of the language mastery mentioned above. Although we know that syntactic, lexical and textual coherence problems are present in students' essays, the types of errors made within these dimensions haven't yet been determined.

Therefore, we decided to investigate the lexical errors made by university students when writing. But even before we could start formulating a research methodology, many theoretical problems arose. Is there such a thing as *lexical error*? Is there a norm governing the usage of the lexicon — as there is one for grammar —, a norm from which we could judge whether the use of a word is correct or not? If we accept that there is more to lexical errors than spelling mistakes, what are the limits of what we could call a lexical problem?

Looking at the literature, both in linguistics and language teaching, we couldn't find any definition of the lexical error. If some authors mention correctors' difficulty to separate syntax and vocabulary in many cases (Denance, 1999), no one, to the best of our knowledge, proposed a clear characterization of lexical error. The first step in our attempt to elaborate a typology of errors made by students was therefore to define the limits of lexical error. The Meaning-Text Theory (MTT), which gives a central role to meaning and lexicon, provided a perfect framework for our task. Of course, we concentrated on the lexicological branch of the theory: the *Explanatory Combinatorial Lexicology* (ECL). This particular approach to lexicology supplied us with all the tools necessary to propose a clear and complete characterization of lexical error.

This article is divided into four parts. The first section defines the notions of *error* and *norm*. The second constitutes the core of the article and is devoted to the characterization of the lexical error. The third part presents a validation of our typology on a text corpus and the fourth one discusses some problems encountered. The conclusion will mention possible applications of the typology proposed.

1. Error and norm

First of all, we must define the concept of error, in its linguistic meaning. We could say that an error is the result of the transgression of a norm. This definition might be right, but it is also completely useless until we know to which norm we should refer. This is the main problem in trying to define the lexical error; if the “rules” are clear in such fields as orthography and verbal morphology, the norm governing the usage of the lexicon is vague and difficult to delimit. Therefore, a purely prescriptive approach in our endeavour would lead to a dead end. So, even if the concept of error is somehow “prescriptive”, we will have to refer to different norms, depending on the dimension of the word in which we are interested.

The written language is, in many aspects, subject to the **prescriptive norm**, and this is true for some facets of the lexicon as well. We could say that a norm is “prescriptive” when it is a set of instructions specifying which usage, among all those linguistically possible, is the “right” one. The best example of how the prescriptive norm applies to the lexicon is the orthography. Opposed to the prescriptive is the **descriptive norm**, the point of view generally adopted by linguists. The descriptive norm considers “normal” a usage common to a vast majority of speakers: the norm is then statistical. This norm will be essential to describe certain lexical phenomena for which there are no “normative” rules, collocation for instance. As we also need to consider the pragmatic dimension of communication, we will refer to the **functional norm** as a set of conditions which exercise some lexical choices, the level of language for example.

2. Characterization of the notion of lexical error¹

With these three types of norms, we will be able to characterize the lexical error. Our approach is quite simple: considering each aspect of the mastery of a lexical unit, we will determine in which ways the appropriate norm could be violated, leading to a lexical error. But what are the dimensions involved in the mastery of a lexical unit? This is where ECL brings us some answers. We know that ECL has a specific and rigorous method of lexical description which materializes in *Explanatory Combinatorial Dictionaries* (ECD). An ECD entry contains all the information necessary to correctly use the lexical unit described in all contexts. In

¹ Because our typology has been elaborated with a French corpus, all examples will be in French. We will provide a translation that shows the lexical error where possible.

other words, an ECD entry contains everything a speaker should know to have complete mastery of a specific lexical unit. Therefore, it is easy for us to see which problems could possibly occur when using a lexical unit: any violation of specific information appearing in an ECD entry. To elaborate our typology of lexical errors, we will consider the three aspects of the lexical unit: the signifier (form), the signified (meaning) and the combinatorial properties.

Signifier:

The signifier is the form of a word or, if we consider written language, the way it is spelled. This aspect of the lexical unit is, as we mentioned before, subject to the prescriptive norm. Apart from some rare cases for which variants exist such as *clé* ~ *clef* (= ‘key’), there is usually only one spelling accepted for a word. It is then easy to see how this aspect of the lexical unit could be transgressed: by not respecting the spelling imposed by the prescriptive norm. This disrespect of the signifier is our first category of lexical errors, within which we will distinguish two types. The first one is the **spelling mistake**, which indicates ignorance of the “right” spelling of the lexical unit. For example, **restaurant* and **échalotte* (= ‘shallot’) are spelling mistakes. The second type of error results from a deeper lack of knowledge of the form of a lexical unit. A **barbarism** is a distorted form of a word that could be noticed not only in writing, but even orally, **crocodile* and **aréoport* (= ‘airport’) for example.

Signified:

The signified of a lexical unit is surely its central component, as words exist precisely to express meanings. The meaning of a lexical unit is set synchronically and thus subject to the prescriptive norm. ECL has a precise and rigorous method of definition: semantic decomposition. We effectively describe a specific meaning in terms of simpler meanings that we can call *meaning components*. For our description to be complete, we also have to take into account the actantial structure of the lexical unit. Finally, we have to define the semantic constraints that limit the type of lexical units that can actualize the actants. ECL specifies these constraints with semantic labels. The following definition of the French verb COURIR (= ‘to run’) satisfies ECL’s principles of definition.

« (Êtres animés) Aller, se déplacer rapidement par une suite d’élans, en reposant alternativement le corps sur l’une puis l’autre jambe, l’une puis l’autre patte. »²

² *Le Petit Robert. Dictionnaire alphabétique et analogique de la langue française, version électronique 2.1.*

The meaning ‘courir’ was decomposed into simpler meanings and is presented as an analytical definition. This definition also tells us implicitly that COURIR is a predicate with one semantic actant that can be actualized by a lexical unit denoting an animate being (*Êtres animés*). Therefore, two types of transgressions are possible with the meaning. The first one is the **disrespect of the constraints on actants**. For example, ARBRE (= ‘tree’) couldn’t function as the subject for COURIR as it doesn’t denote an animate being. The second case would be the **disrespect of any other semantic component**. An example of this would be the sentence [#]*Le ver courait sur la feuille.* (= [#]*The worm was running on the leaf.*)³. Although the lexical unit VER (= ‘worm’) satisfies the constraints on the actant as it denotes an animate being, there is still a problem with this sentence. The problem is due to a transgression of the semantic component ‘en reposant alternativement le corps sur l’une puis l’autre patte’ (= ‘the body being supported alternatively by one leg then the other’), as a worm doesn’t have legs. Of course, the misunderstanding of the meaning of a lexical unit is impossible to detect if the lexical unit appears by itself; it is from semantic incompatibilities with other words that we will be able to diagnose lexical errors related to meaning.

Combinatorial properties:

The third aspect of a lexical unit is what the MTT calls the *syntactics*, which contains all combinatorial properties. Polguère (2003: 37) defines combinatorial properties as “the constraints inherent to a sign that limit its capacity to combine with other linguistic signs and which can’t be deduced from its signified nor its signifier” [Translation is ours. — DA]. We will categorize combinatorial properties according to two axes: grammatical and lexical. The prescriptive norm will help us to determine errors regarding grammatical combinatorial properties, but we will have to refer to the descriptive norm for lexical properties.

The combinatorial properties related to government pattern are those listed in the syntactic zone of an ECD entry. This section specifies how to express in syntax the actantial structure of the lexical unit. These structures can’t be fully explained by the general syntax rules of a language; they are inherent to specific lexical units that all govern their “micro-syntax”. Therefore, a student writing **Il aide à sa mère à ranger la cuisine.* (= **He’s helping to his mother to clean the kitchen.*) would make a lexical error because he would transgress a fundamental property of the

³ The symbol # indicates that the expression is semantically deviant, while * indicates that it is ungrammatical.

verb AIDER, the fact that its second syntactic actant must be a direct object. Problems related to government pattern are often considered by teachers as syntax errors, which is not completely wrong. But, since the source of the problem is the lack of knowledge of a lexical property (even if its manifestation is syntactic), we prefer to talk about **lexico-syntactic errors**.

A lexical unit has many other grammatical properties that are essential in its mastery, such as gender (for a noun) and part of speech. These characteristics of the lexical unit have great impact on how it combines with other words. In French, for instance, the gender of a noun influences the inflection of the article and the adjective. Therefore, if a student writes **une grosse avion* (= ‘a_{fem} big_{fem} plane’), we will consider that he’s making a lexical error, as the noun AVION is masculine. This kind of error is often considered as a grammatical problem by teachers, which is wrong in our opinion. In fact, the student who writes **une grosse avion* perfectly applied the grammatical rules of agreement between the noun, its article and its adjective. The problem was prior to the application of grammatical rules; the student didn’t know that AVION is a masculine noun. Therefore, it is more accurate to talk about a **lexico-grammatical error**, since the source of the problem is the ignorance of a lexical property and its manifestation is grammatical. Grammatical properties are various (defective verbs, invariability, etc.) and usually appear in dictionaries. But one type of lexical unit for which grammatical properties are not always clear should be mentioned: idiomatic expressions, or *phrasemes*. The set character of these expressions is difficult to define and not all “linguistic manipulations” are tolerated. For example, the verbal phraseme FORCER LA MAIN [à qqn] (= ‘to force one’s hand’) can’t be put in the passive voice. Phrasemes are not described thoroughly in dictionaries and we may need to refer to corpuses (cf. descriptive norm) to understand how they can be used.

The second group of combinatorial properties correspond to the information listed in the lexical zone of an ECD entry. There is no doubt, for anybody who works within the MTT, that the mastery of collocations constitutes a central point in the mastery of a lexical unit. If we accept this, we will consider that the **use of a wrong collocate** constitutes a lexical error. For example, somebody who writes **Le jury a pris son choix*. (= **The jury took its choice*.) makes an error, as the “right” collocate for CHOIX is FAIRE [Oper₁(choix) = *faire*, Oper₁(choice) = *make*]. Of course, no “official” rules exist to determine whether a collocate is correct or not, and dictionaries don’t provide exhaustive lists of collocations. Therefore, our only way to decide if a collocation is acceptable will be our own judgement as a native speaker, or the use of corpuses. So, in regard to restricted lexical cooccurrence, we will refer to the descriptive norm.

Use of a fictitious lexical unit:

All the lexical errors discussed above are consequences of the violation of a property of the lexical unit. When such a violation reaches a certain degree, it becomes difficult to speak of an actual lexical unit. We will then consider that the lexical unit used is fictitious, as it doesn't exist in language. We will distinguish three types of fictitious lexical units.

Impropriety results from the association of a signified with an existing signifier whose signified shares no semantic component with the new lexical unit created. In other words, an impropriety is a wrong association between a meaning and a form that already exists in the language. For example, someone using the word *ÉVASION* (= 'escape') to express the meaning 'invasion' (= 'invasion').

The **creation of a new lexical unit of a *vocable*** in accordance with a productive polysemy pattern is the second type of fictitious lexical unit. For instance, it is very common for French verbs to be polysemous so that a verbal *vocable* contains the primary sense and its causative. An example is the verbal *vocable* *BRÛLER* (= 'to burn') that contains *BRÛLER*₁[*La feuille brûle.* (= 'the sheet is burning')] and *BRÛLER*₂[*Il brûle la feuille.* (= 'He's burning the sheet')]. In this case, *Caus(brûler*₁) = *brûler*₂. So, following this polysemy pattern, one could write **Il cliqueta ses clefs pour attirer l'attention de Lili.* (= 'He rattled his keys to attract Lili's attention'). But in French, the causative sense of *CLIQUETER* (= 'to rattle') doesn't exist and you must say *faire cliqueter quelque chose* (= 'make something rattle') to express it. The person using *CLIQUETER* in this sense is therefore using a fictitious lexical unit. Metonymy and metaphor are also very common polysemy patterns and their application in some cases could occasion fictitious lexical units. Of course, we have to be very careful when we describe as an error a metonymic or metaphoric use of a lexical unit, as these two lexical phenomena can also be used as stylistic devices.

The last case is the **use of a nonexistent signifier to express a meaning**. This fictitious form is created by morphological derivation rules that are productive in other cases. Following the model of *BANQUE* ~ *BANQUIER* (*BANK* ~ *BANKER*), one might "generate" the word **brocantier* (= 'someone who works in a secondhand shop') from *BROCANTE* (= 'secondhand shop'). Since this form doesn't exist in French, we have to consider it a fictitious lexical unit.

Lexical errors according to the context:

So far, we have seen lexical errors resulting from the transgression of inherent properties of a lexical unit. But sometimes, lexical errors will re-

sult from the use of a lexical unit in an inappropriate context. The following problems are subject to the functional norm.

Different **levels of language** exist and our choice of register is conditioned by the context of communication. Lexical units belonging to an informal register could then be considered as lexical errors in a given context.

When two languages are in contact, it is inevitable that some mixes occur. Some words of the other language might be accepted while the use of others will be criticized. **Anglicisms** in French are a good example of this phenomenon; if a certain word is considered an anglicism by dictionaries (cf. prescriptive norm), it will be considered a lexical error. For instance, **PARKING** is considered an anglicism since the French equivalent **STATIONNEMENT** exists. Lexical units typical of certain dialects could also be perceived as lexical mistakes in some contexts. For example, using words belonging to French as spoken in Québec might be inappropriate if you were to write an article for a French magazine.

The cooccurrence of two words expressing the same meaning is called a **pleonasm** and also constitutes a type of lexical error. *Se réunir ensemble* (= *to unite together*) is an example of a pleonasm.

To evaluate lexical mastery, we must finally take into account the range of vocabulary. **Excessive repetition** of words in a text and **use of periphrases or generic terms** (such as **STUFF**) indicate a lack of vocabulary and could therefore be perceived as lexical problems.

3. Testing our typology of lexical errors on a text corpus

After reviewing the information essential to the mastery of a lexical unit, we are able to identify 19 different types of lexical errors. The next step is to verify if this typology of lexical errors actually helps to diagnose lexical problems in texts. For this purpose, we constituted a corpus of 103 texts written by students in their first year of university. All of them have been studying in French since primary school and can be considered as native speakers. Our correction revealed 391 lexical problems, among which 260 could be considered as “real” errors and 131 are perceived as “awkward uses” of lexical units for which teachers wouldn’t necessarily deduct marks.

Our analysis revealed that the most common lexical errors⁴ are those affecting the combinatorial properties (30.4% of all lexical problems). If this conclusion might seem predictable for linguists, it is more surprising

⁴ For the following statistics, we have considered all 391 problems as “errors”.

in the context of first language teaching, where this aspect of the lexical mastery is generally neglected. Our work also showed that errors relating to inappropriate use of lexical units according to the context are as common as those attributable to an alteration of the signifier (mostly spelling mistakes): they each represent 20.9% of all lexical problems. Once again, this observation is interesting if we compare the importance lent to spelling in language teaching with the one given to the presentation of such notions as levels of language and lexical variation. The use of fictitious lexical units represents 15.8% of all lexical problems and impropriety is by far the most common type (14%). This fact reveals that students don't seem to doubt their understanding of word meanings and probably don't use dictionaries, or at least not efficiently. Finally, lexical problems resulting from an alteration of the signified constitute 8.4% of the 391 errors.

4. Problems encountered

If our typology of lexical errors looks exhaustive and precise, its application to diagnose problems found in actual texts occasioned some difficulties.

First observation: different perspectives on specific errors could lead to different characterizations. We quickly realized that interpretation is inherent to the activity of "labelling" errors. We experienced this by asking a colleague to use our typology to characterize some of the lexical problems found in our corpus. If the diagnosis was often the same, we have seen cases for which it differed. The interesting fact is that both interpretations were valid. We therefore had to admit that lexical errors don't always fit into just one category and to allow ourselves to double code in some cases. A reason for this might be the actual structure of our typology, as some of our categories seem to be overlapping. For example, anglicisms can often be perceived as improprieties, some errors related to the government pattern are typical of the spoken language, etc.

The second big problem we met was the characterization of lexical problems related to meaning. Our typology includes three types of such errors: those caused by a *transgression of constraints imposed by actants' type*, those ensuing from the *disrespect of a meaning component* and *improprieties*. If the definitions of these concepts are clear, the limits between each category of errors are, in their application, very difficult to determine. Each time we've met a problem that was attributable to a misunderstanding of the meaning of a lexical unit, we've tried to identify which component was concerned, which was almost impossible in many

cases. We see two reasons for that. First, the commercial dictionaries we used as references don't provide exhaustive definitions that would allow us to see all meaning components with the semantic decomposition. Second, the fact that meaning problems can only be detected from semantic incompatibilities between lexical units forces us to infer the "altered meaning" present in the writer's head from the limited context we see. In other words, we are rebuilding the meaning from the text, which is the opposite of the approach favoured by the Meaning-Text Theory.

Many questions regarding the level of sophistication of the typology arose during our errors analysis. We first built this typology in a didactic perspective — as we ultimately expected it to be used by language teachers who are not necessarily familiar with the MTT Theory —, but we also wanted it to be valid from a linguistic point of view. The thing we realised while using it to label errors was that the categories proposed were too vague in some cases, and errors of different natures were identically labelled. For example, among the errors labelled **impropriety**, some present a signified ~ signifier association that seems "accidental", for instance the adjective EXPRESSIF (= 'expressive') used to express the meaning 'turbulent' (= 'unruly'). Other errors also labelled **impropriety** seem phonetically motivated, for instance the verb CORROBORER (= 'corroborate') used to express the meaning 'collaborer' (= 'collaborate'). But our typology does not show this distinction between "accidental" improprieties and those attributable to confusion between paronyms. In other words, some of our categories should be divided into subcategories to be able to allow a proper linguistic analysis of errors. On the other hand, refining the typology in order to satisfy thorough linguistic description would render it very complex and difficult to use by teachers who don't necessarily have the time to really understand and assimilate our linguistic approach. The simple solution to this problem would be to make our typology more precise and exhaustive with all the subcategories needed before proposing a simplified version of it to be used by teachers.

The last and most fundamental problem has to do with the "descriptive" character of our typology. We wanted this typology to be "descriptive", which means we expected to be able to characterize errors by their structure only, with no speculation on the thought processes that led to them. After labelling our 391 lexical errors, we came to the conclusion that it is not always possible to characterize errors without referring to the potential sources of the problem. This fact implies a part of subjectivity when coding errors, as we speculate about the thought processes of the person who made them. Of course, this observation is not really shocking. In fact, error characterization doesn't constitute a goal in itself; it is a

necessary step in the endeavour of finding ways to prevent these errors. Therefore, it is understandable that the analysis of errors can't always be dissociated from the question of the possible sources of these errors.

Conclusion

This article demonstrates that the characterization of the notion of lexical error is possible and that ECL provides a valuable framework in achieving this task. Of course, the typology proposed could be improved, and testing it with other languages or on a second language corpus would insure a better validation, although we don't think it would lead to major changes. Nonetheless, its application in labelling actual lexical problems showed that this typology is a useful tool to stimulate reflection on lexical errors.

We elaborated our typology in order to describe the lexical problems of students, but we are convinced that a typology similar to the one proposed above has numerous applications, particularly in language teaching. Such a tool would help teachers to better understand and explain errors usually difficult to diagnose. By giving them new avenues to explore, it would bring them beyond the “*we don't say it this way*” correction often typical of lexical errors. But, most of all, this typology shows all aspects of the lexical unit and proves that there is more to lexical mastery than knowing many words.

Bibliography

- Denance, M. (1999). «L'évaluation des fautes de vocabulaire à l'évaluation de l'épreuve uniforme». *Correspondance*, vol. 4. № 4.
- Lefrançois, P. (2004). «Les stratégies de résolution de problèmes orthographiques d'étudiants à l'université». In C. Barré-De Miniac, C. Brissaud & M. Rispail (eds.), *La littératie : conceptions théoriques et pratiques de l'enseignement de la lecture-écriture*. Paris: l'Harmattan, 2004. P. 233—244.
- Mel'čuk, I., Clas, A., & Polguère, A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve, France: Duculot, 1995.
- Polguère, A. (2003). *Lexicologie et sémantique lexicale : Notions fondamentales*. Montréal : Les Presses de l'Université de Montréal, Coll. Paramètres, 2003.
- Roy, G.-R., Lafontaine, L. & Legros, C. (1995). *Le Savoir grammatical après treize ans de formation*. Sherbrooke : Éditions du CRP, 1995.

Prolegomena to Systemic Lexicography¹

Juri D. Apresjan

Institute for Information Transmission Problems, Russian Academy of Sciences /
Institute of the Russian Language, Russian Academy of Sciences

The basic principle of systemic lexicography requires the unification of grammar and dictionary within an integrated linguistic description, i. e. a description in which grammar and dictionary are fully coordinated with regard to the types of linguistic information included and the formal ways of its representation. In such a description linguistic rules should be geared to the dictionary entries and dictionary entries should be made sensitive to the rules. The implementation of this principle brings about a fundamental change in the approach to lexicographic description, necessitating the introduction of two new concepts, those of an integrated lexicographic representation of a lexeme and of a lexicographic type. An integrated lexicographic representation of a lexeme is composed of (a) its analytical definition in a special metalanguage, (b) a set of interrelated syntactic, combinatorial, morphological, communicative, prosodic and other characteristics deducible to a large extent, though not fully, from its semantic core and (c) a list of relevant linguistic rules with lexically constrained scopes. Lexicographic types are multiply intersecting classes of lexemes with very similar or identical lexicographic representations; they should be uniformly treated throughout the dictionary, with, of course, allowances made for the individual properties of lexemes.

Пролегомены к системной лексикографии. Главный принцип системной лексикографии требует интегрального описания грамматики и словаря, т. е. такого описания, в котором грамматика и словарь полностью согласованы друг с другом по типам помещаемой в них информации и по формальным способам ее записи. В таком описании лингвистические правила должны быть настроены на словарные статьи, а словарные статьи должны быть чувствительны к правилам. Реализация этого принципа приводит к изменению самого подхода к лексикографическо-

¹ The present work has been supported with the President of the Russian Federation's grant for support of the leading scientific schools No. НШ-1576.2003.6, the Program for Fundamental Research of the Department of History and Philology of the Russian Academy of Sciences (section 4.15) and three grants of the Russian Foundation for the Humanities (Nos. 05-04-04190a, 03-04-00046a and 04-04-00263a).

му описанию и требует введения двух новых понятий — интегрального лексикографического представления лексемы и лексикографического типа. Интегральное лексикографическое представление лексемы складывается из (а) ее аналитического толкования на специальном метаязыке, (б) множества взаимосвязанных сведений о ее синтаксических, сочетаемостных, морфологических, коммуникативных, просодических и иных свойствах, в большой степени, хотя и не полностью выводимых из ее значения, и (в) списка релевантных для нее правил с лексически ограниченной областью действия. Лексикографические типы — это многократно пересекающиеся классы лексем с очень похожими или идентичными лексикографическими представлениями; они должны единообразно описываться в словаре, разумеется, с учетом индивидуальных свойств лексем.

0. Introductory

As can be seen from the title, the present paper purports to outline an approach to lexicographic research. This approach has been practised by the author over forty years now and has been implemented in a number of lexicographic products, the most important being synonym dictionaries of English and Russian. It owes very much to Igor Mel'čuk's idea of an ECD as part of the MTT.

I am going to discuss and illustrate with representative examples the following three ideas underlying research in systemic lexicography:

1) The necessity of an integrated linguistic description, i. e. a description in which grammar and dictionary are fully coordinated with regard to the types of linguistic information included and the formal ways of its representation. Such a description requires that linguistic rules should be geared to the entries of an integrated dictionary and that the entries of such a dictionary should be made sensitive to linguistic rules.

2) The emphasis on integrated descriptions of lexemes. The term "lexeme" here stands for a word considered in one of its meanings. Within such a description a lexeme may be thought of as the sum of (a) its analytical definition in a special metalanguage, (b) a set of interrelated syntactic, combinatorial, morphological, communicative, prosodic and other characteristics deducible to a large extent, though not fully, from its semantic core, c) a list of relevant linguistic rules with lexically constrained scopes.

3) The search for systematicity in lexicon as manifested in predicate classification, regular polysemy, lexicographic types etc. For considerations of space we shall be able to discuss below only the concept of lexicographic types.

1. The unified, or integrated approach to linguistic description

Every complete linguistic description is ultimately made up of a grammar and a dictionary. It is reasonable to expect that these two documents should be mutually adjusted to each other, i. e. coordinated with regard to the types of information included and the formal devices (metalanguages) employed to record them.

Unfortunately, up till quite recently these natural principles have not been clearly formulated, much less adhered to. Originally dictionaries and grammars were produced by different people. The result was basically discordant grammars and dictionaries that did not give a coherent picture of the language at large. Below I shall quote one of the most intriguing examples of such a discrepancy that I know.

English grammar has always recognized (cardinal) numerals as a part of speech in its own right, distinct from nouns and adjectives. Indeed, their semantic, morphological, syntactic, combinatorial, derivational, and accentual properties are very different from those of true nouns and adjectives².

(1) Semantics. Numerals have a unique and well-defined meaning of number which is almost totally alien to genuine adjectives and nouns, with the exception of such isolated examples as *dozen*, *score*, *couple*, *both*, *several* etc. Nouns and adjectives have much fuzzier meanings not reducible to a single prototype. Prototypically nouns refer to things, but they can also denote actions (*payment*), activities (*trade*), occupations (*Don't interrupt his reading*), processes (*rise of the prices*), states (*joy*), properties (*courage*), human attributes (*name*), parameters (*weight*) and so on. Adjectives in this respect are very much like nouns.

(2) Morphology. Numerals have no number marking — the basic morphological category of genuine nouns; forms like *twos* or *tens* (*in twos*, *by tens*) are nouns, not numerals.

(3) Syntax. Numerals form several interesting constructions, none of which can be assigned to nouns or adjectives: (a) the normal quantitative construction with the defined noun in postposition in the plural form, as in *five books*; (b) the numerative-appositive construction with the defined noun in preposition, as in *room five*; (c) the quantitative-auxiliary construction with an additive meaning, comprising two numerals in the “descending order”, as in *twenty five* = ‘20 unities plus 5 unities’; (d) the quantitative-auxiliary construction with a fractional meaning, comprising

² Some of the observations that follow are due to I. A. Mel'čuk's account of Russian numerals; see Mel'čuk 1985.

two numerals in the “ascending” order, as in *five twenty* (= ‘five large units plus twenty smaller units which are fractions of the former’). Semantically and syntactically all these constructions are entirely different from comparable adjectival and substantive constructions; in particular, constructions like *twenty five* or *five twenty* conjoining two numerals have almost nothing in common with the typically substantive or adjectival constructions conjoining two nouns or adjectives, like *cannon ball*, *computer system*, *dark blue*, *English-Russian* and so on. On the other hand, such characteristically substantive constructions as *dozens of people*, *a score of people* and the like are totally inadmissible for numerals.

(4) Lexical co-occurrence (combinatorial profile). There are at least four patterns of lexical co-occurrence which are specific for numerals. (a) Numerals combine with a unique lexeme of the preposition *by* in the multiplicative construction of the type (*a room*) *ten by five*. (b) A closely related pattern is the homonymous “divisive” construction (*to divide*) *ten by five* featuring another unique meaning of the same preposition. (c) The third pattern is co-occurrence with a unique “subtractive” meaning of the preposition *off* which is also possible only in the context of numerals, cf. *at 10% off the regular price* = ‘less than the regular price’³. (d) The last numeral-specific pattern manifests itself in the combinations of numerals with the word *o’clock* in the indications of current time, cf. *It’s five o’clock now*.

(5) Derivation patterns. Derivationally numerals are set off from nouns and adjectives by such patterns as ‘*X + teen*’, ‘*X + ty*’, ‘*X + th*’, ‘*X + fold*’, as in *fifteen*, *fifty*, *fifth*, *fivefold*.

(6) Accentuation. Compound numerals are set off against comparable substantive combinations in that they bear two equal stresses, as in *’twenty ’five*, or *’five ’twenty*. In combinations of two conjoined nouns the first often (though not always) bears the principal stress and the last — the secondary stress, cf. *cannon ball*, *stone wall*, *computer system*.

Let me note in conclusion that all the non-semantic properties of numerals considered above (except stress) are motivated by one central property — their unique lexical meaning.

This account shows that numerals have a greater claim to being an independent part of speech than any other class of words. Within a rigorous description of English, classing numerals as nouns or adjectives in the dictionary is bound to play havoc with the grammatical rules geared to

³ Let me note in passing that that rounds off the set of “arithmetic” meanings peculiar to constructions with numerals — additive, subtractive, multiplicative and divisive. Cf. also combinations with unique lexemes *plus* and *minus*.

genuine nouns and adjectives. However, there is virtually no comprehensive dictionary of British, American, Australian or any other variety of English that has a special grammatical label for numerals. In some most influential families of dictionaries (Oxford, Webster, Chambers, Random House, The American Heritage, Thorndike and Barnhart, Wyld, Macquarie, and a host of others) numerals are labelled as nouns (*twenty five* type of construction) or as adjectives (*five books* type of construction)⁴.

As can be seen from this account, traditional grammar and dictionary at this point are glaringly incompatible. Insistence on *integrated* linguistic descriptions, with perfectly coordinated dictionary and grammar, is thus the first basic idea of systemic lexicography.

2. Integrated lexicographic representation of a lexeme

From the principle proposed above it immediately follows that within an integrated linguistic description every lexeme in the dictionary should be assigned all linguistically relevant properties, with particular emphasis on the semantic motivation of its more formal properties. A certain property is considered to be linguistically relevant if there is a rule of grammar or some other sufficiently general rule (semantic rules included) that accesses this property. Once the given lexeme is viewed against the whole set of linguistic rules, an entirely novel point of observation is created. It highlights new facets of lexemes and helps to uncover a number of their lexicographically relevant and often semantically motivated properties that have never been recorded in dictionaries.

Consider the following two uses of the adverb *alone*: (1) *The house stands on the hillside all alone, He likes living alone* (I shall call this *alone 1*); (2) *Smith alone knows what happened, Time alone will help, Man shall not live by bread alone* (I shall call this *alone 2*).

Current dictionaries confine themselves to specifying just their meanings. *Alone 1* is defined as ‘by oneself, without the company or help of others’, while *alone 2* — as ‘and no other, no one else, the only one’. All the other relevant properties are practically ignored, although the sets of such properties are conspicuously different for each of the two lexemes. I shall try to do fuller justice to those other properties in the account that follows.

⁴ Some of the more noticeable exceptions are Macmillan’s dictionaries (the label “number”), Collins Cobuild (“num” meaning ‘number’), Longman (“determiner, noun, pronoun”), Oxford Advanced Learner’s (“pronoun, determiner”).

(1) Semantics. I propose the following more explicit definitions of the lexemes at issue: *XP alone 1* = ‘X does P; no one else does P simultaneously or together with X’ (cf. *He went there alone*); *X alone 2 P* = ‘X does, is or has P; there is no one else that does, is or has P’ (cf. *He alone is courageous enough to tell you the truth*). *Alone 1* is synonymous to the syntactic phraseme “*by + self-pronoun*”, cf. *He went there alone* ≈ *He went there by himself*; *alone 2* is synonymous to the quantifier adverbs *only* and *just* and in fact is itself a quantifier, cf. *Smith alone knows it* = *Only Smith knows it*.

(2) Syntax. There are several interesting syntactic properties that should be recorded in the dictionary entries for *alone 1* and *alone 2*. (a) *Alone 1* is an adverbial modifier, that is, a verbal dependent (*Don't go there alone*), whereas *alone 2* is a noun attribute (cf. *Smith alone, time alone*). (b) The normal or preferable position of *alone 1* is at the end of the verbal group: *He went there alone* is better than *?He went alone there*. (c) *Alone 2* is always placed in postposition to the defined head, which distinguishes it from its closest synonym *only*; cf. *Smith alone can do it* VS. *Only Smith can do it*, while **Alone Smith can do it* is absolutely ungrammatical. (d) As a noun attribute *alone 2* can define not only prototypical substantives but personal pronouns as well: *He alone can do it*.

(3) Lexical co-occurrence. *Alone 1* and *alone 2* display several notable co-occurrence distinctions. (a) *Alone 2* (the noun attribute) is more or less indifferent to the semantic class of the predicate in whose context it occurs. The latter may denote an action (*Smith alone will go there*), an activity (*Smith alone fights corruption*), a current occupation (*Smith alone is reading*), a spatial position or location (*Smith alone is sitting at table, Smith alone was in the yard*), a mental or emotional state (*Smith alone knows it, Smith alone is ashamed of what he has done*), a property (*Smith alone is brave enough*) etc. (b) *Alone 1*, for semantic reasons (see its explication above), combines mostly with actional predicates and does not occur in the context of states or properties: **Smith knows that alone, *Smith believed that alone, *Smith was ashamed alone, *Smith is brave alone* are all ungrammatical. (c) *Alone 1* may be modified by such particles and adverbs as *all, almost, practically* etc., cf. *He stood there all alone, The kids were left there almost alone, The widow raised her three children practically alone*. *Alone 2* does not take any such modifiers; phrases like **He all alone stood there, *The kids almost alone were left there, *The widow practically alone raised her three children* are ungrammatical.

(4) Theme — rheme (topic — comment) structure. *Alone 1* has no permanent value in that respect. It may mark off the verbal group as the

theme (topic) of the utterance, as in *Living alone 1* [theme] *is a nuisance* <a pleasure>. On the other hand, it may serve as the rheme (comment), as in *The house stands on the hillside all alone 1* [rheme]. Unlike it, *alone 2* always marks off the nominal group to which it refers as the rheme of the utterance; cf. *Smith alone 2 knows what happened*.

(5) Prosody. Distinction (4) is mirrored in the prosodic properties of the two lexemes. *Alone 1* can bear the main phrasal stress, as in *The house stands on the hillside all ↓ alone 1*, or it may be left phrasally unstressed, as in *He ↓ likes living alone 1*. Contrary to that *alone 2* always bears the main phrasal stress, cf. *Smith ↓ alone 2 knows what happened*, *You ↓ alone 2 can help me*.

To sum up, an integrated lexicographic representation of a lexeme requires storing in its entry a much greater amount of linguistic information than the current dictionaries do.

3. Lexicographic types

I use this term to refer to a group of lexemes with a shared property or properties which are sensitive to the same linguistic rules and should therefore be uniformly described in the dictionary. This concept is exemplified with very well-known classes of factive and putative predicates denoting mental states. So my point in this section will be to illustrate rather than to report any new findings, although a couple of new observations are added.

(1) Semantics. Following Vendler 1972, the label of “factive” is assigned to verbs *to know* <to understand, to guess, to remember, ...> (*that P*) and the like which govern propositions denoting facts. All of them are decomposable into semantic structures with the sense ‘to know’ at the bottom and presuppose the truth of the subordinate clause. That means that irrespective of whether the knowledge of *P* is asserted or denied, *P* always remains true. Such sentences as *He knew that he was under police surveillance* and *He didn’t know that he was under police surveillance* are alike in asserting that he was under police surveillance.

The label of “putative” is assigned to verbs *to think* <to believe, to consider, to find, to hold, to doubt, to suspect, ...> (*that P*) and similar predicates which denote opinions. Opinions, unlike knowledge, are not necessarily true. In other words, it cannot be deduced either from the sentence *He thought that he was under police surveillance*, or from the sentence *He didn’t think that he was under police surveillance* whether he was in fact under surveillance or not.

Both groups of verbs share the common feature of all statives noted in Vendler 1967: 99—103, namely, a specific relation to the idea of duration. It manifests itself above all in the inability of *to know* (<to understand, to guess, to remember, ...> (*that P*) and *to think* (<to believe, to consider, to find, to hold, to doubt, to suspect, ...> *that P*) to occur in the progressive tenses (in the senses under consideration). Indeed, the phrases **When I entered he was knowing* (<understanding, guessing> *that the meeting had been cancelled*) or **When I entered he was believing* (<considering, doubting, suspecting> *that the meeting had been cancelled*) are ungrammatical.

On other points factive and putative statives differ from one another. All of their formal differences are quite systematic, i. e. semantically motivated, so that two well-defined lexicographic types emerge which should be uniformly described throughout the dictionary.

(2) Syntax. There are a number of well-known syntactic properties, most of which are semantically motivated, that distinguish factives from putatives. (a) Factives can govern an oblique question introduced by the *wh*-words: *He knew what it was* (<where to look for the mistake, how to do the job>). Putatives do not govern oblique questions. (b) Knowledge has a source but not a reason. Therefore factives can govern nominal groups denoting sources of information and cannot subordinate adverbial modifiers of cause. Compare the well-formed phrases *How do you know it?*, *I know it from the newspapers* and the ungrammatical phrase **Why do you know it?* (c) By contrast, opinions have a reason, but never a source. Therefore putative verbs can subordinate adverbial modifiers of cause but not those denoting a source of information. We can say *Why do you think so?* But not **I think so from the newspapers*. (d) Putatives allow of neg-transportation, with only a slight change of emphasis: *I didn't think he would cope with the task* $\approx$ *I thought he would not cope with the task*. With factives neg-transportation is impossible for semantic reasons: there is a fundamental difference between *I didn't know he had coped with the task* $\neq$ *I knew he had not coped with the task*. (e) Putative verbs denote opinions, that is, evaluative judgements. Therefore most of them can in some way or other govern assessment constructions with an obligatory second complement denoting the essence of evaluation: *I think* (<believe, consider, find>) *her young*, *I regard* (<look upon>) *this marriage as a mistake* etc. It should be noted that assessment constructions presuppose simultaneity of the time of opinion and the time of the situation assessed (i. e., 'she is young', 'this marriage is a mistake'). Propositional factives have no such pattern. Superficially similar phrases like *I knew* (<remember>) *her young* feature a different syntactic construction with a different — tempo-

ral — meaning: *I knew her young* means ‘I knew her at the *time* when she was young’. The lexical meaning of the verb in such constructions also changes; in the last example *to know* denotes acquaintance, not propositional knowledge.

(3) Lexical co-occurrence. (a) Factive verbs rather freely co-occur with positive evaluation adverbs like *well*, *perfectly well*, and so on. To qualify knowledge in this way is just to emphasize its truth. Cf. *I know that you are against rigid measures* $\approx$ *I know perfectly well that you are against rigid measures*. Putatives in such contexts are ruled out: phrases like **I think perfectly well that you are against rigid measures* are totally ungrammatical. (b) In their turn, putative verbs co-occur freely with truth-adverbs like *correctly*, *rightly* and the like: *He rightly thought that it would be pointless to continue the conversation*. The respective phrases with factives are pleonastic and therefore ungrammatical: propositional knowledge cannot be wrong by definition.

(4) Theme-rheme structure and prosody. I shall subsume those two interrelated topics within one section. Putative words express somebody’s opinion about something. But saying something under ordinary circumstances is equivalent to expressing an opinion, so that the propositions *He is wrong* and *I think he is wrong* convey approximately the same meaning; cf. Ch. Bally’s idea that every proposition opens up with the implied modus ‘Je crois’. Since the modus ‘I think’ is implicit in every proposition, unless the opposite is overtly stated, it has very little information value and is therefore reserved for the thematic part of the utterance. Therefore putatives are never marked off with the main phrasal stress. The only type of phrasal stress they can bear is the so-called logical, or contrastive stress marking the contrastive rheme of the utterance, as in the sentence *Do you ↑↑ think you are under police surveillance, or do you ↓↓ know it?* The contribution of factive words to the meaning of the utterance is much greater: it is always important to know something for a fact. Therefore factives often occur in the rhematic part of the utterance and can bear the main phrasal stress, as in the phrase *I ↓ knew she would marry him*, *I ↓ remember how it all ended*.

These distinctions are so strong that they occur even within a single word if it happens to have a factive and a putative sense. Note the difference in the interpretation of the verb *to understand* in such sentences as *I ↓ understand he is in trouble* (He is in trouble) and *I understand he is in ↓ trouble* (I am doubtful about whether he is in trouble or not and am asking for information rather than asserting anything).

Up till now I have insisted on systematicity in the description of lexicographic types. But blind allegiance to systematicity can easily lead the

lexicographer into the trap of overgeneralization. In particular, it is unreasonable to expect that all stative factives and putatives will uniformly display the prototypical properties of those two lexicographic types.

For example, various stative putative verbs display varying degrees of incompatibility with the idea of duration mentioned above. The verb *to think* (*that P*), which is a close synonym of *to believe* and *to consider*, can be used in the progressive tenses, especially when it is conjoined with a genuinely actional verb: *As I lay down thinking that my book was quite close to completion, I heard the phone ring*. Neither *to believe*, nor *to consider* can replace *to think* in such contexts.

The explanation is to be sought in the semantic structure of *to think* as a whole. The basic sense of *to think* is purely actional: *I was thinking about tomorrow's session when the phone rang*. Now, closely related senses of a single word are apt to "grow" into one another and impart to the neighbouring senses at least some of their properties. This is especially true of the basic senses. In such cases deviation from the prototype becomes highly probable.

This adds a new dimension to the facts discussed so far. It appears that in lexical description one should give equal attention to the shared properties of lexemes (the problem of lexicographic types, or *unification*) and to what distinguishes them (the problem of lexicographic portraits, or *individuation*).

References

- Mel'čuk, I. A., Poverxnostnyj sintaksis russkix chislovyx vyrazhenij [Surface Syntax of the numerical constructions of Russian] // Wiener Slawistischer Almanach, Sonderband 16, Wien 1985. 509 s.
- Vendler, Z., Verbs and Times. In: *Linguistics in Philosophy*. Ithaca, Cornell University Press, 1967. P. 97—121.
- Vendler, Z., *Res Cogitans*. Ithaca, Cornell University Press, 1972. 225 p.

Computing Semantic Relations on Structured Lexical Definitions

Lucie Barque, Alexis Nasr

LATTICE-CNRS (UMR 8094)

Université Paris 7

`lbarque@linguist.jussieu.fr`, `alexis.nasr@linguist.jussieu.fr`

1. Introduction

The BDéf was introduced in (Altman and Polguère, 2003) as a formal database derived from the Explanatory Combinatorial Dictionary of Contemporary French (ECD). One of the major features of the BDéf entries is the formal description of the internal structuring of lexical meanings. The need to precisely describe such a structuring led to the definition of a formal metalanguage for lexicographic definitions. The work presented in this paper is based on the BDéf approach and shows how the BDéf definitions can be used in order to compute semantic relations between lexical units.

The structure of the paper is the following: in section 2 we describe the grammar of the BDéf metalanguage, in section 3 we show how semantic relations between lexical units can be modelled using the decomposition proposed by the BDéf. We illustrate this, in section 4, on the antonymy relation. Section 5 concludes the paper.

2. Formalizing the BDéf definitional metalanguage

In this section we quickly review the structure of a BDéf definition, as introduced in (Altman and Polguère, 2003); we then propose to describe the BDéf definitional metalanguage by means of a formal grammar and to represent the entries as typed feature structures (Carpenter, 1992), (Copestake and Briscoe, 1995). The grammar allows us to pro-

duce feature structures as the output of a parsing process that takes as input the definitions. The feature structures produced are then used in a calculus.

We will illustrate the BDéf metalanguage on an example, the French verb APPLAUDIR I (*to applaud*), whose definition, translated into English, is represented below, in a linear form, as it might appear in the ECD (Mel'čuk et al. 1984, 1988, 1992, 1999). Its BDéf definition appears on the left in figure 1.

X applaud Y for Z $\equiv$ ‘the person X produces a meaningful sound by means of clapping hands that express that X congratulates the person Y for his performance Z’

The BDéf definition of the verb APPLAUDIR I decomposes the ECD definition into smaller parts (propositions and groups of propositions) and describes explicitly the relations that hold between these parts.

APPLAUDIR I	
Propositional Form	
$X \sim Y \text{ pour } Z$	
Semantic label	
son.expressif	
DEFINITION	
Central Component	
/*son expressif*/	
1: X produire son.expressif	
Definition body	
/*Contenu*/	
2: *1 exprimer *3	
3: X complimenter Y pour Z	
/*Manière*/	
4: *1 se.faire.au.moyen.de *5	
5: X taper.dans.les.mains	
Actant typing	
X: personne	
Y: personne	
Z: performance	

[LU APPLAUDIR I	
P-FORM	[1 [NAME X TYPE personne], [2 [NAME Y TYPE personne], [3 [NAME Z TYPE performance]
SEM-LAB	son.expressif
DEF	[CENTRAL-COMP [LABEL son.expressif FIRST-PROP [4 [PRED son.expressif PRED-ARG [1]
	[BODY [LABEL contenu FIRST-PROP [5 [PRED exprimer PRED-ARG [4, [6]
	[CONTENT [6 [PRED complimenter PRED-ARG [1, [2, [3]
	[LABEL manière FIRST-PROP ...]

Fig. 1. The BDéf description of the French verb APPLAUDIR I (left) and its representation as a feature structure (right)

More precisely, a BDéf entry describing the meaning of a lexical unit L consists of four parts, a propositional form, which introduces the actants of L , a semantic label, which encodes the general meaning of the unit, the definition proper, which constitutes the main part of the semantic description of the unit and the actant typing section which attributes a type

to each actant introduced in the propositional form¹. The feature structures defining the types **bdef entry** and **lu argument** are presented below.

bdef entry			lu argument
LU		lexical unit	NAME
PROP-FORM		list_of lu argument	TYPE
SEM-LAB		semantic label	
DEF		definition	

The definition itself comprises two parts: the central component and the definition body, which respectively correspond to Aristotle notions of *genus* and *differentiae* (Aristote, ed. 2004). Briefly, the central component represents the general meaning of the defined lexical unit (also represented, but more concisely, by the semantic label). The definition body represents components (*differentiae*) which specify further the central component and distinguish the lexical unit from other units that share the same general meaning. Both the central component and the definition body are made of blocks (one for the central components and several for the definition body), as shown below in the feature structure that defines the type **definition**.

definition			block
CENTRAL-COMP		block	LABEL
BODY		list_of block	FIRST-PROP
			CONTENT
			block label
			first proposition
			list_of proposition

A **block** represents an «autonomous» component of the definition, which means that one can remove a block from a definition and keep the definition well-formed whereas removing a part of it does not. Each block is tagged with a block label (between /* */ signs in the definition, see figure 1). Such a label accounts for the informational purpose of the block. In our example, the second block of the definition body specifies *how* an applause is performed.

A block itself contains a list of propositions. Each of them is numbered so that it can be referred to anywhere in the definition. A **proposition** is made of a predicate and its arguments which can be a semanteme, a reference to another proposition, or an actant introduced in the propositional form. For instance, in figure 1 the proposition number 2 contains

¹ A BDéf entry contains another section describing the semantic relations (if any) that hold between the actants. For instance, the actant Y of APPLAUDIR is the first actant of the predicate Z that denotes a performance. Due to space limitations we will ignore this part of the description here.

the predicate **exprimer** and its two arguments: a pointer to (the predicate of) proposition 1 and a reference to (the predicate of) proposition 3. Propositions may also include modifiers of the predicate (for example a negation adverb) and support verbs, when the predicate is not a verb (for example, in proposition number 2, **produire** is a support verb of the main predicate **son_expressif**). The **first proposition** of a block plays a special role: its predicate is tightly linked to the block label.

first proposition		proposition	
PRED	block indicator predicate	PRED	semanteme
PRED-ARG	list_of pred argument	MOD	semanteme
		ARG-STRUCT	list_of pred argument

The «lexical units» of the metalanguage are described in another database² where the semantemes used in the propositions are associated with a lexical unit. For example the configuration of semantemes (also called BDéf word) **se_faire_au_moyen_de** will be associated with the lexical unit **MOYEN** in order to access its meaning. The description also accounts for the function of the unit in the definition. For instance, **se_faire_au_moyen_de** is a block indicator predicate of the block **manière**. The labels (the semantic label and the block labels), which belong to the label hierarchy, are described in another part of the BDéf lexicon (Polguère, 2003a).

The conditions of structural well-formedness of a definition are represented as a context free grammar, more specifically as an XML Document Type Definition (DTD). This DTD defines a set of XML tags as well as the rules that describe the structure of the definitions. It is therefore possible to check the structural well-formedness of a given definition using standard XML parsers. A definition is represented as an XML document³ and it is also possible, using standards tools, to produce the original representation for the entry, as it appears on the left side of figure 1, from its XML representation.

To be semantically well-formed, a definition must verify some «semantic» constraints. Let's mention two of them: first, the labels of the blocks that can appear in the definition of *L* are controlled by its semantic label. In our example, the semantic label **son expressif** accounts for the occurrence of the two blocks **Contenu** (what does X mean when he pro-

² The development of this database is in progress at OLST, Université de Montréal.

³ Space limitations preclude us from presenting here an example of a definition in the XML format.

duces this sound) and **Manière** (how does X produce this sound). Second, the label of a block controls the predicate of its first clause: the predicate must express the informational purpose of the block.

The semantic tags, block labels, first proposition predicates and the relations that hold between them are represented in another XML document whose structure is described in another DTD.

As noted, the lexical entries are represented as typed feature structures in order to carry out computations on them. The feature structure representation of the definition⁴ of **APPLAUDIR I** is represented on the right in figure 1. This means of representation allows us to use the operation of unification for our calculus, as in (Pustejovsky, 1995) and (Copestake and Briscoe, 1995). The following sections present one of them: to check whether a given semantic relation exists between two given lexical units.

3. Semantic relations

3.1. Definition

Semantic relations are defined in (Polguère, 2003b) in terms of set-theoretic operations on lexical definitions. A lexical unit L is seen as a structure based on a set of simpler lexical meanings (the *semanemes* that appear in the definition of L). A semantic relation holds between two lexical units L_1 and L_2 if there is either identity, inclusion or a non null intersection between the definitions of L_1 and L_2 . Such relations are seen as basic semantic relations in terms of which more complex relations, as hyperonymy/hyponymy, synonymy, antonymy, ... are built.

As we have seen in section 2, the BDéf definition of a lexical unit describes precisely its lexical meaning in terms of other lexical units, uncovering the «organization» that is mentioned in the definition above. This explicit decomposition allows for a precise definition of lexical relations: the decomposition of the definitions in terms of blocks, propositions and predicates, and their representation as feature structure enable the description of a lexical relation as a pair of underspecified feature structures. Such feature structures describe the properties that two lexical units must have for a relation to hold between them. The fine grained level of decomposition of the definitions allows for a perspicuous defini-

⁴ The representation of a definition as a typed feature structure can be produced automatically from the XML representation of a definition.

tion of relations. One can, for example, define a relation between two lexical units by referring to specific blocks, propositions or predicates of their definitions.

3.2. Defining lexical relations as feature structure pairs

We define a lexical relation R as a couple of feature structures S_1 and S_2 (we note $R = \langle S_1, S_2 \rangle$)⁵. The relation R holds between two lexical units L_1 and L_2 (we note $R \langle L_1, L_2 \rangle$) if $*L_1$ unifies with S_1 and $*L_2$ with S_2 , or the opposite, where $*L$ denotes the feature structure that represents L 's definition. In other words, the operator ‘*’ is used to access the semantic decomposition of a lexical unit.

$$R \langle L_1, L_2 \rangle \Leftrightarrow (S_1 \vee *L_1 \wedge S_2 \vee *L_2) \vee (S_1 \vee *L_2 \wedge S_2 \vee *L_1)$$
⁶

The type of relation defined above is called direct relation since it is directly defined on two lexical units. Indirect relations between lexical units L_1 and L_2 imply that another relation R (direct or indirect) exists between components of L_1 and L_2 (we note $IR = \langle S_1, S_2 \rangle \wedge R \langle S_1.x, S_2.y \rangle$ where $S_1.x$ and $S_2.y$ are paths in the feature structures S_1 and S_2 leading to semantemes and R is a semantic relation). Checking for the definition of an indirect relation IR between L_1 and L_2 is defined below:

$$IR \langle L_1, L_2 \rangle \Leftrightarrow ((S_1 \vee *L_1 \wedge S_2 \vee *L_2) \vee (S_1 \vee *L_2 \wedge S_2 \vee *L_1)) \wedge R \langle L_1.x, L_2.y \rangle$$

4. Computing antonymy

Two lexical items L_1 and L_2 are in an antonymy relation if they share the same part of speech and if some components of the definition of L_1 are negated or, more generally «put in opposition» in the definition of L_2 or the opposite (Polguère, 2003b).

When formally defining (as a pair of feature structures) the antonymy relation, many subcases can be distinguished according to the position that the negation occupies in the two feature structures. We describe below three types of antonymy which correspond to three different positions

⁵ R is actually defined as a feature structure having S_1 and S_2 as substructures. We have chosen to represent it here as a couple for readability reasons.

⁶ Unification is seen here as a logical predicate. The expressions $A \vee B$ is true if A and B unify.

of the negation in the patterns⁷. Note that we only consider the cases where negation is located in the central component of the definiton⁸.

Subtype 1: Inclusive antonymy In this case, a lexical unit appears, modified by a negation, in the definition of another one⁹, as in the example below, where the definition of RÉUSSIR¹⁰ (to succeed) is included in the definition of ÉCHOUER (*to fail*).

$$\left[\begin{array}{c} \text{LU} \quad \text{ÉCHOUER} \\ \text{DEF} \left[\begin{array}{c} \text{C-COMP} \left[\begin{array}{c} \text{FIRST-PROP} \left[\begin{array}{c} \text{PRED} \quad \text{réussir} \\ \text{MOD} \quad \text{ne\_pas} \end{array} \end{array} \right] \end{array} \right] \end{array} \right] \end{array} \right]$$

The definition of this type of antonymy is represented below:

$$Anti_1 = \left\langle \left[\text{LU} \begin{array}{c} \boxed{1} \end{array} \right], \left[\text{DEF} \left[\text{C-COMP} \left[\text{FIRST-PROP} \left[\begin{array}{c} \text{PRED} \quad \boxed{1} \\ \text{MOD} \quad \text{ne\_pas} \end{array} \right] \right] \right] \right] \right] \right\rangle$$

Subtype 2: Intersective antonymy In this case, two lexical units share a predicate, that is negated in one of them. An example of this subtype is presented below with the pair

PROVOQUER (to cause) ~ EMPÊCHER (to prevent)¹¹.

Lexical unit X provoquer Y	Lexical unit X empêcher Y
Central component	Central component
1: X causer *2 (1: X cause *2)	1: X causer *2 (1: X cause *2)
2: Y avoir lieu (2: Y happen)	2: Y ne pas avoir lieu (2: Y not happen)

⁷ This typology is mainly inspired by the typology of antonymy presented in (Apresjan, 1992) and (Milićević, 2003). The reader should note that antonymy has been chosen here as a case in point. For the detailed literature on this type of semantic relation, see among others, (Apresjan, 1992), (Cruse, 1986), (Lehrer and Lehrer, 1982), (Lyons, 1995).

⁸ For instance, pairs of antonyms like APPLAUDIR ~ SIFFLER (*applaud/boo*) are not taken into account here. Indeed, they share the same central component (X produire son expressif) while the opposition takes place in the body of the definition, more precisely in the Contenu block (X expresses that he is happy vs X expresses that he is not happy).

⁹ More precisely, the first lexical unit must appear as the predicate of the first proposition of the central component of the other lexical unit.

¹⁰ We have only represented the relevant part of the definition.

¹¹ Due to lack of space, we represent the central component of these two lexical units in the BDéf formalism and not as feature structures.

The definition of intersective antonymy is represented below:

$$Anti_2 = \left\langle \left[\text{DEF} \left[\text{C-COMP} \left[\begin{array}{l} \text{FIRST-PROP} \left[\begin{array}{l} \text{PRED } [4] \\ \text{PRED-ARG } [1] \end{array} \right] \\ \text{PROP } [1] \end{array} \right] \right] \right], \left[\text{DEF} \left[\text{C-COMP} \left[\begin{array}{l} \text{FIRST-PROP} \left[\begin{array}{l} \text{PRED } [4] \\ \text{PRED-ARG } [3] \end{array} \right] \\ \text{PROP } [3] \end{array} \right] \right] \right] \right] \right\rangle$$

Subtype 3: Indirect antonymy In this case, the central component of one lexical unit includes a predicate which is in an antonymy relation with a predicate of the other lexical unit¹². These two predicates must be located at the same place in the definition. If they appear in a subordinate proposition, they should share the same governor. The pair EXHIBER (*to exhibit*) ~ ESCAMOTER (*to make sth vanish*) illustrates this third subtype.

Lexical unit X exhibe Y

Central component

1: X causer *2 (1: X cause *2)

2: Y apparaître (2: Y appear)

Lexical unit X escamoter Y

Central component

1: X causer *2 (1: X cause *2)

2: Y disparaître (2: Y disappear)

The definition of indirect antonymy is represented below:

$$Anti_3 = \left\langle \left[\text{DEF} \left[\text{C-COMP} \left[\begin{array}{l} \text{FIRST-PROP} \left[\begin{array}{l} \text{PRED } [5] \\ \text{PRED-ARG } [1] \end{array} \right] \\ \text{PROP } [1] \end{array} \right] \right] \right], \left[\text{DEF} \left[\text{C-COMP} \left[\begin{array}{l} \text{FIRST-PROP} \left[\begin{array}{l} \text{PRED } [5] \\ \text{PRED-ARG } [3] \end{array} \right] \\ \text{PROP } [3] \end{array} \right] \right] \right] \right] \right\rangle$$

5. Conclusions and perspectives

We have presented in this paper a methodology that consists in modelling semantic relations by means of BDéf definitions. The formal patterns that result of this modelling are closely related to lexical functions (LF), another formalism developed by the *Explanatory Combinatorial Lexicology* to model semantic relations in the lexicon. Formalization based on BDéf descriptions may help to tackle issues like the control of LF's values, the organization of the set of an LF's values, etc. In this paper we have used the relation patterns to check for the occurrence of a relation *R* between two lexical units *L*₁ and *L*₂. Such patterns can also be used in another way: having the definition of *L*₁ and the relation *R*, we can automatically produce an (underpecified) definition of *L*₂. This technique could be used by the lexicographer when producing new entries.

¹² Note that this third subtype has to be taken into account with respect to the computations on definition but may be not relevant linguistically.

Acknowledgements

We are grateful to Sylvain Kahane and Alain Polguère for helpful comments on the preliminary draft of this paper.

References

- Altman J., Polguère A. (2003), La BDéf: base de définitions dérivée du Dictionnaire explicatif et combinatoire, *Proceedings of the First International Conference on Meaning-Text Theory*, Paris, p. 43—54.
- Apresjan J. (1992), *Lexical Semantics: User's Guide to Contemporary Russian Vocabulary*, Ann Arbor.
- Aristote (2004), *Les Topiques*, Librairie Philosophique J. Vrin, Paris.
- Carpenter B. (1992), *The logic of typed feature structure*, Cambridge University Press, Cambridge, England.
- Copestake A., Briscoe E. (1995) Semi-Productive Polysemy and Sense extension *Journal of Semantics*, 12(1), p. 15—67.
- Cruse D. A. (1986) *Lexical semantics*, Cambridge University Press, New York.
- Lehrer A., Lehrer K. (1982), Antonymy, *Linguistics and philosophy*, 5, p. 483—501.
- Lyons J. (1995), *Linguistics semantics*, Cambridge University press, London.
- Mel'čuk I. et al. (1984, 1988, 1992, 1999), *Dictionnaire explicatif et combinatoire du français contemporain*, vol. I—IV, Les Presses de l'Université de Montréal, Montréal.
- Milićević, Jasmina (2003), *Étude des aspects sémantiques et syntaxiques de la paraphrase: application à la génération automatique des phrases françaises*, Thèse de doctorat, Département de linguistique et de traduction, Université de Montréal.
- Polguère A. (2003a), Étiquetage sémantique des lexies dans la base de données DiCo, *TAL*, 44, 2, p. 39—68.
- Polguère A. (2003b), *Lexicologie et sémantique lexicale*, Presses de l'université de Montréal, Montréal.
- Pustejovsky J. (1995), *The Generative Lexicon*, MIT Press, Cambridge, Mass.

Paraphrasing with Space Constraints: Linguistic Operations in Journal Abstracting

Pascale Bélanger

Département d'Informatique et
Recherche Opérationnelle,
Université de Montréal /
Nstein Technologies, Inc.
Belangpa@iro.umontreal.ca.

Richard Kittredge

Département de Linguistique et
Traduction,
Université de Montréal /
CoGenTex, Inc.
Richard.Kittredge@umontreal.ca.

Recent research into text summarization provides an opportunity to test, expand, and eventually implement MTT-based rules for paraphrasing and other linguistic operations as part of a general model of text condensation. Our studies of French and English journal articles in biology and medicine show that lexical functions are important in modeling the transformation of selected full-text (FT) sentences into the sentences used for an article's author-written abstract (AB).

Перифразирование в условиях ограничений на длину текста. Лингвистические операции в реферировании журналов. Исследования в области автоматического реферирования, проводимые в последнее время, предоставляют обширное поле для проверки, дальнейшей разработки и последующей компьютерной реализации правил перифразирования и других лингвистических трансформаций системы «Смысл-Текст» в контексте общей модели сжатия текста. Наши исследования французских и английских статей из академических журналов по биологии и медицине показывают, что лексические функции играют важную роль в моделировании трансформации предложений, выбранных для включения в автореферат, в их формулировку в автореферате.

1. Introduction

For the past half-century a major goal of computational linguistics has been to simulate the human capacity to summarize written texts in various genres. As early as the 1950s, considerable attention was devoted to the problem of automating the production of abstracts of scientific and tech-

nical journals. Various techniques, both linguistic and quantitative, were proposed for selecting and trimming sentences from full texts to produce informative text *extracts*. Further linguistic processing was required, however, to condense the extracts and make them into fluent *abstracts*. Techniques for doing so were dependent on reliable parsing and semantic analysis, which was then beyond the state of the art. In a landmark review of prior research, Paice (1990) called for renewed attempts to apply computational models of syntax and semantics to journal abstracting.

During the 1990s, statistical approaches showed promising results on the problem of selecting salient sentences from news articles (Kupiec et al. 1995), but such techniques were clearly inadequate for creating a fully coherent and cohesive abstract from a handful of extracted non-contiguous FT-sentences. At about the same time, workers in text generation began to apply insights gained from text planning to the problem of summarization in general and abstracting in particular. Some of the new techniques pertained to sentence selection (e.g., picking salient FT-sentences using rhetorical structure trees (Marcu 1997) and “lexical chains” running through the whole text (Barzilay & Elhadad 1997)). A few researchers (Maybury 1995; Jing & McKeown 2000; Saggion 2000) began to address the issues of conceptual generalization and fusing of selected sentences into new wholes. But there was still a missing part of the puzzle — detailed techniques for semantic and syntactic paraphrasing, models of communicative structure, and methods for simplifying content, often with domain-dependent rules.

To improve the contribution of linguistics to abstracting, there has been a recognized need to observe more closely what it is that human authors do when they compose an abstract based on a journal article. Chuah (2001), for example, analyzed 57 journal articles in entomology and behavioral biology, along with their author-written abstracts. Each AB sentence was traced back to one or more source sentences in the full text, resulting in a categorization of linguistic processes for transforming selected FT sentences into AB sentences¹. Chuah described in detail the variety of lexical substitution processes observed, showing the utility of WordNet semantic relations (e.g., synonymy, hyperonymy, meronymy, holonymy, etc.) in linking the FT item with its AB substitute from the same lexical category (e.g., Meronym(*summer*) = *August*). In addition,

¹ This model makes the assumption that authors work entirely from the full text, which is clearly not true, since one occasionally finds new material (often evaluative words) in the abstract. But this simplification is close enough to the truth to allow empirical work to proceed, with interesting results.

Chuah introduced a limited use of MTT lexical functions to describe lexical relations not found in WordNet (e.g., $\text{Mult}(\text{larva}) = \text{brood}$), and in describing some derivational changes that accompany nominalization or denominalization (e.g., *dramatic differences* $\Leftrightarrow$ *differ dramatically*).

In this paper we report on our investigation into some linguistic processes used in abstracting which can be described conveniently and powerfully within the MTT framework. These include:

- paraphrase operations, roughly as defined in the ECD vol. III (Mel'čuk 1992);
- synonymic phrasal equivalences, some of which are specific to a subject matter domain;
- simplification or compression operations which preserve major aspects of meaning;
- operations on longer stretches of text which use lexical functions to paraphrase rhetorical structures as described in Rhetorical Structure Theory (RST, Mann & Thompson 1988) and other features of paragraphs.

The following section summarizes our corpus studies of French and English journal articles in biology and medicine, and outlines the overall view of the abstracting process in which MTT could play a role, both as descriptive mechanism, and in support of practical implementations in the distant future. In Section 3 we present corpus examples illustrating some of the most frequent or interesting phenomena observed in the two languages. Section 4 raises some questions and challenges for MTT that we see as resulting from our investigation.

1. Corpus studies of English and French

1.1. English research journals in medicine and biology

Kittredge (2002) analyzed several on-line medical journal articles in English from the viewpoint of MTT paraphrase operations, attempting to show the plausibility of a linguistic model for reformulation and reduction where lexical functions play a central role. He also sketched an overview of the whole FT-to-AB transformation process which fits Spark-Jones' (1999) three-stage model of text summarization: (1) **interpretation** of the source text into some representation, (2) **transformation** of the FT interpretation into a representation of the summary or abstract (e.g., by selection, deletion, aggregation, etc.), and (3) **generation** of the summary or abstract from the representation.

In our view of this model as it applies to journal abstracting, part of the interpretation stage should involve steps such as anaphora resolution, word-sense disambiguation, detection of rhetorical structures (at the level of clause, sentence, and full text), and communicative structures, etc. Also included here would be some marking of the FT-sentences and their major segments with respect to information salience, by whatever means, statistical and linguistic. Note that the interpretation stage can benefit greatly from the fact that journal articles are *formatted* according to a standardized division that is common to many journals, typically the four sections titled Introduction, Materials and Methods, Results, Discussion. During the second stage, transformation of FT to AB, the sequence of interpreted FT-sentences undergoes a selection process, whereby key sentences or key clauses (less than 10% of the whole text) are chosen, based on rhetorical, lexical chain, statistical and other techniques. We have little to say about the details of this part, except that research into selection techniques seems to be ahead of research into the means of reformulation and condensation, where MTT has an important role to play. Reformulation must somehow ensure that the future AB-sentences are paraphrased in a way that allows syntactic and semantic *aggregation* processes to combine similar clauses and eliminate redundancies. It seems clear that communicative structures inherited from the FT will help guide the reformulation, but some will be overridden by textual constraints that operate within the emerging representation of the AB. If the transformation stage can be carried out at the deep syntactic or semantic level, then the generation stage will involve realization of the final AB, as in any text generation process. This will include re-creating certain pronouns and other anaphors as a function of the newly reduced text.

Kittredge (2002) also found that, among the many kinds of lexical replacements are some that are apparently specific to the domain's special sublanguage. As reported by Lee and Evens (1996), a specific medical domain makes use of a particular subset of lexical functions (LFs) and may require the introduction of non-standard LFs. It also appears that values for some standard LFs need to be recalculated or their argument conditions specially restricted. Furthermore, Kittredge (2002) describes an opportunity to combine some of the insights of RST with those of MTT to account for transformative paraphrases that reformulate material from two or more sentences of FT (see Section 4 for more on this point). His further analysis of texts in biology and medicine has expanded upon the 2002 results. Some examples of these observations for English are given in Section 3.

1.2. French research journals in botany

Bélanger (2003) examined 30 journal articles in French dealing with botanical research and found one or more plausible source sentences in the full article (FT) for each sentence in the article's author-written abstract. Her analysis focuses on the variety of linguistic processes by which source sentence(s) are transformed into abstract sentence(s) (FT → AB transformations). It first confirms the basis of Chuah's taxonomy of condensation processes, and then goes beyond that distinction to describe a subset of the linguistic processes found in terms of MTT, among them various kinds of lexical substitutions, phrase substitutions and other paraphrasing operations.

The study shows that about 70% of the lexical substitutions found in FT → AB transformations could be encoded using a subset of standard LFs, plus a few non-standard LFs. In addition to various types of synonyms and antonyms, the most frequent LFs involved were (a) derivatives between the categories Verb, Noun, Adjective and Adverb (LFs **V₀**, **N₀**, **A₀**, **Adv₀**), and (b) superordinate replacement (LF **Gener**). On the other hand, in the sample analyzed, occurrences of paraphrases involving support verb constructions were relatively infrequent.

In addition, some productive operations were found which are not yet described in detail as standard MTT paraphrasing. Other examples show that LFs need to be re-evaluated and their values need to be described specifically for sublanguages of botany, confirming observations by Lee and Evens (1996), Chuah (2001) and Kittredge (2002). Bélanger also considered another type of transformation — simplification processes, i.e., approximate paraphrases similar to those described by Kittredge. Unlike paraphrases, simplifications tolerate minor semantic change throughout the FT → AB process. In fact, simplifications often combine the characteristics of substitution and deletion processes, but within an apparently specific and constrained area. Since quite frequent and productive rules seem to underlie some of the simplification processes observed in Bélanger's study, it seems reasonable to attest the regularity of a subset of those processes and to consider their similarities with paraphrases (as proposed in section 3).

Bélanger also found that some of the transformations observed between FT and AB sentences extracted from author-written abstracts are beyond the scope of linguistic transformation, and therefore cannot be expressed in term of LFs, paraphrasing rules or simplification rules. We omit discussion of these in this paper.

2. General features of the transformation of FT to AB sentences

The current section will set out some general features of lexical substitution, compression², and paraphrase rules using examples from our French and English corpora. As often as possible, the transformations presented will be described in MTT terms.

Typically, in the passage from FT to AB sentence, more than one operation is being carried out. Deletions or simplifications of FT material usually accompany otherwise paraphrastic changes, and presumably must be described as a preliminary step. Example (1a-b) shows the simultaneous operation of paraphrase with deletion (~~striethrough~~) and other operations in transforming a FT sentence into an AB sentence³. Note that the FT sentence of example (1a) appears in the Results section of the article, fourth paragraph and second sentence, (R-4—2), and the corresponding AB sentence (1b) is the ninth of the abstract (A-9). This indexing convention will be followed for every example, along with the source journal reference.

- (1a) *Overall 20 (60%) patients, ~~14 with NVE and six patients with PVE~~, underwent valvular surgery 6±9 (range: 0—38) days after admission to the ICU. [R-4—2] →*
- (1b) *Surgical intervention was performed in 60% of the patients... [A-9; Critical Care 2002, 5: 149—154].*

While the previous example shows a direct correspondence between a single FT sentence and a single AB sentence, the material for an AB sentence occasionally comes from aggregating two different FT sentences, or clauses thereof. In this case, deletions, synonym substitutions and sentence paraphrasing operations may need to be applied to set up the argument clauses in a sufficiently parallel structure for the aggregation. Note that in example (2a-b) below there must be one synonym substitution involving a LF on the first clause (rule 1, Mel'čuk 1992: 37), and passive operation applied to both clauses; this shows aggregation at sentence level, but combined with replacement and passivization processes.

² We use the term 'compression' for an elementary operation that reduces total length, and 'condensation' as a more general term, including any simple operation or complex sequence that figures in the process of re-expressing "essential" content in less space.

³ The derivation of (1b) from (1a) involves several steps including a substitution of type $Oper_1 \Leftrightarrow Oper_2$, accompanied by actant permutation (cf. Lexical Rule 38 in Mel'čuk 1992). This example is analyzed in Kittredge (2002: 272—274); a slightly different analysis is proposed in Milicevic (2003: 264—266).

- (2a) *Les feuilles ont un effet inhibiteur de l'efflux et peuvent, de ce fait...* [R-5—4]
L'apex, au contraire, a un effet stimulateur de l'efflux. [R-5—5] →
- (2b) *...l'efflux est freiné par les feuilles et stimulé, au contraire, par l'apex.* [A-8; R.C.B. 2001, 79: 398—403]
- Oper**₁(effet) : avoir;
 —**S**₀(stimuler) : stimulation; —**Able**₁(stimuler) : stimulateur;
 —**S**₀(inhiber) : inhibition; —**Able**₁(inhiber) : inhibiteur, —**Syn**_c(inhiber) : freiner

Notice that the parallelism of clause structures needed for aggregation to operate can very well be defined in terms of DSyntS dependency trees; also, the lexical identity of the common elements can be determined quite simply in dependency terms (as compared with phrase structure trees, for example).

Sometimes, the observed synonymic replacements or other paraphrase operations do not in themselves lead to shorter sentences; indeed, a paraphrastic operation on a FT sentence may lead to a longer AB sentence. The longer structure may be necessary for various reasons, such as emphasis, preserving communicative structure of the abstract as a text, etc. A simple example of emphasis addition is provided in (3a-b), where the author specifies the object of shortening — the length. Knowing that *shorten* means “to reduce the length”, the extra information is completely redundant.

- (3a) *...était bien liée à un raccourcissement de l'entre-noeud.* [I-9—2] →
- (3b) *Un raccourcissement temporaire de la longueur de l'entre-noeud est donc lié...* [A-4; R.C.B. 2000, 78: 68—76]

The space constraint inherent in abstract writing appears to be smoother than first expected, thus expanding the range of transformations likely to be encountered. But it still exists, and is represented by the high number of compressions and substitutions by merged LFs, as shown in example (4a-b).

- (4a) *...et l'augmentation brutale de la longueur du pétiole...* [D-1—2] →
- (4b) *...et allongement du pétiole...* [A-1—7; R.C.B. 2000, 78: 68—76]
- IncepPredPlus**(longueur): augmenter // allonger also, note deletion of *brutale* = **Magn**(*augmentation*)

Examples of classical full-sentence paraphrase operations are not infrequent (especially use of passive, conversive, **Oper**₁), but are not as frequent as more local synonymic replacements. Examples are provided

in (5a-b) and (6a-b), which respectively show the application of a standard fission rule involving a support-verb LF (Rule 18, Mel'čuk 1992: 39), and an application of a standard conversive rule (Rule 2, Mel'čuk 1992: 37).

- (5a) *Seule la variable épaisseur totale des horizons organiques présente des variations spatiales...* [D-8—11] →
 (5b) *L'épaisseur totale des horizons organiques fluctue dans l'espace...* [A-9; R.C.B. 1998, 76: 804—817]
 –**Oper1**(variation) : présenter [ART ~] | pluriel
 –**QSyn**(variation) : fluctuation; **V0**(fluctuation) : fluctuer
 –**S0**(spatial) : espace; **Locin**(espace) : dans [ART_{def} ~];
 (6a) *...les fèces d'équins contiennent 17 fois moins de graines...* [D-3—2] →
 (6b) *... [peu] de graines se trouvaient dans les fèces [...] des équins.* [A-4 ; R.C.B. 2000, 78: 1430—1439]
 –**Conv₂₁**(contenir) : se trouver

Gener fission rules are quite frequent, but the **Gener** value may be slightly different from the expected LF value in the general language. Examples (1a-b), shortened here, and (7a-b) correspond to **Gener** fission transformation (Rule 31, Mel'čuk 1992: 40) as long as we are willing to accept **Gener**(base): zone and **Gener**(surgery): intervention.

- (1a) *...underwent... surgery* [R-4—2] →
 (1b) *Surgical intervention was performed...* [A-9; Critical Care 2002, 5: 149—154.]
 (7a) *...le fonctionnement de la base_N du méristème apical...* [D-1—2] →
 (7b) *...du fonctionnement de cette seule zone_N basale_A du méristème apical.* [A-6; R.C.B. 2000, 78: 68—76]

Standard and non-standard LFs are also useful to express lexical compressions involving simple and merged values of FLs. Examples (8a-b) and (8c-d) show the replacement of two simple Loc values by their merged form. Note that if Loc_{in} is a well-known and accepted FL, Loc_{between} is less frequent and may be considered as a non-standard Loc FL.

- (8a) *...by injecting human recombinant TF (rTF) into joint cavities of healthy mice.* [I-3—1] →
 (8b) *...TF was injected intra-articularly into healthy mice.* [A-3; Arthritis Research 2002, 4: 190—191.]
 –**Loc_{in}**(joint cavities): into [~], // intra-articularly
 (8c) *...des corrélations entre bourgeons déterminent...* [D-10—1] →

- (8d) ...des corrélations intergemmaires qui déterminent... [A-5; R.C.B. 1998, 76: 1385—1403]
 –**Loc**_{between}(**bourgeons**): entre [~], // *intergemmaire*

Some observed paraphrase operations involve a local lexically-based phrase-level replacement which is not always easy to describe in terms of LFs; many of these appear to be specific to the sublanguage or to a family of related sublanguages. Examples (9a-b) illustrate a paraphrase which one could formulate as:

X be (more/most) closely related to Y (than Z) ⇔ X is evolutionarily close(-er/-est) to Y (than Z).

The adjective *close (to)* can also be replaced by *distant (from)*, but in either case, this appears to be an equivalence that holds only in biological discussions. Note also the equivalence of the two generic forms, one with plural, the other with definite article: *sea urchins = the... sea urchin*.

- (9a) *Sea urchins ... are thus more closely related to humans than are other invertebrate model organisms such as Drosophila and C. elegans.* [I-1—2] →
 (9b) *The Pacific coast sea urchin ... is evolutionarily closer to humans than Drosophila and C. elegans, ...*
 [A-1—2; BMC Devel. Biol. 2005, 5: 3]

Many observed transformations include simplifications along with paraphrastic operations. In the transformation of (10a) to (10b), for example, the segment *environ 3 fois* is simplified to *fortement* or perhaps just **Magn** in the context of the comparison. In addition, there appears to be a paraphrastic operation at work which could be represented as:

$[X \text{ } V_0(C) \text{ } \text{moins} \{ \text{de } W \} \{ \text{que } Y \}] \Leftrightarrow [C \{ \text{de } W \} \text{ IncepPredMinus}(C) \text{ chez } X \{ \text{comparativement au } Y \}]$.

The degree modifier, in whichever form, depends from the comparative element on either side (*moins* or **IncepPredMinus**(C)). The exact conditions which allow this simplification are probably domain-sensitive. Paraphrases of quantitative comparison are productive in the botanical corpus and often combined with such simplification rules.

- (10a) *Ces dernières exsorbent ~~done environ 3 fois~~ moins de NO₃ — que les plantes intactes ...* [R-1—3] →
 (10b) *...l'exsorption des nitrates est fortement diminuée chez les racines excisées comparativement aux plantes témoins.* [A-2; R.C.B. 2001, 79: 398—403]

–**S₀** (exsorber) : exsorption ; –**IncepPredMinus**(exsorption) : diminuer
 –**Syn** (NO3) : nitrate
 also, note the replacement *intactes* → *témoins* that are contextually considered as exact synonyms.

Other simplifications are entailments instead of paraphrases (cf. Apresjan et al. 2003). The aggregation of two sentences (11a) into (11b) accompanies a transformation of type: *Y is found* → *Y* that might not involve standard LFs, but that is very similar to Mel’čuk 1992’s 53rd implicative rule: *Incep X* → *X*; also, (11a-b) includes a transformation of type: *Y appears* → *Y*.

- (11a) *Aucune relation significative n’apparaît entre la longueur de leur pousse porteuse et celles des pousses florifères monocycliques...* [R-8—1]
Une corrélation positive significative au seuil de 5 % a été trouvée entre la longueur des pousses bicycliques et celle de leur pousse porteuse_x ($r = 0,64$). [R-8—2] →
 (11b) *La longueur des pousses porteuses n’est pas corrélée à la longueur des pousses monocycliques nouvellement émises, mais elle l’est avec la longueur des pousses bicycliques portées.*
 [A-3; R.C.B. 1998, 76: 1253—1264]

3. Some questions and a challenge for MTT

In the course of our research we have encountered phenomena whose description would seem to pose some general questions for MTT. One fundamental issue concerns the relationship between LF description for the general language and for a particular variant of that language used for scientific or technical discourse (special purpose language or sublanguage). Is it likely that some or all lexical LFs active in a sublanguage will need to be described as if in a separate lexicon? If true for only certain LFs, are some more likely to have specialized values than others (e.g., the specialized values for **Magn** described by Lee & Evens)? Our evidence is still scanty, but indicates that this is the case, especially for a domain’s frequently used LFs. Another aspect of this specialization issue regards new varieties of LFs. Should we expect, as it appears at this point, that specialized sublanguages will often make use of one or more non-standard or unusual compositions of LFs for lexical relations that play a key role in the domain? Similar questions arise for standard vs. special-

ized paraphrase operations. We also know very little about the domain of application of paraphrase operations (including semantic constraints on lexical categories used to describe the rule, such as verb aspect). It appears likely that paraphrase operations defined for the general language will apply with somewhat different or specialized argument restrictions in specialized language.

Another set of questions arises concerning the use of MTT representations when describing processes of simplification, implication and deletion needed for abstracting. If we limit our discussion to the deep syntactic level, could we say that **Magn** nodes are good candidates for deletion in simplification operations? Could we say the same thing about **Quant** specification, or $\mathbf{A}_0(\mathbf{Real}(\mathbf{S}_0))$ specification?

The examples of implication operations given in Apresjan et al. (2003) suggest that patterns exist in this type of operation, which is clearly of interest for abstracting. What is the relationship between simplification and implication operations? The two types intersect, but have distinct cases and probably different uses in a detailed model of abstracting.

Finally, we would like to add further support to Kittredge's (2002) claim that there is an opportunity for MTT to interact productively with Rhetorical Structure Theory (RST) of Mann and Thompson (1988). RST postulates a set of two dozen standard rhetorical relations to represent the communicative intention of the author as to how segments of text are to be understood as connected into larger segments to make up a coherent whole. Minimal segments may be single clauses, or clause nominalizations, so that a RST tree structure for the whole text represents its rhetorical (i.e., logical, persuasive, etc.) building blocks, and the specific explicit or implicit connections between them. (Sometimes there is no explicit word or phrase indicating a rhetorical link between two successive sentences, but their mere juxtaposition is interpreted by the reader as signaling a rhetorical intent of the author.) RST thus provides one aspect of semantic representation for multi-sentential chunks of text that can complement MTT's representations of individual sentences. RST relations have been used for computational modeling of text planning from communicative goals (Hovy 1988), and have been proposed as a useful element in models of summarization (Marcu 1997) and automatic translation (Mitkov 1993).

In our work we have found cases where two FT sentences in adjacent segments have been fused into a single AB sentence according to a pattern that could be generalized nicely by using RST relations and lexical functions. For example, the introduction to an article on heart muscle hypertrophy contains the following two adjacent sentences:

(12a) *Thus, whether cardiac hypertrophy in response to pressure overload is adaptive or maladaptive remains a fundamental issue in the pathogenesis of heart failure.* [I-2—3]

(12b) *To address this issue, we studied the effect of long-term pressure overload on the development of heart failure using two murine models that have an attenuated hypertrophic response to pressure overload:...* [I-3—1]

The corresponding single sentence in the abstract preserves a RST rhetorical link called **Purpose**, reflected explicitly in the *to*-clause, which also links (along with anaphora) the two sentences of the full text.

(12c) *To determine whether cardiac hypertrophy is necessary to preserve cardiac function, we used two genetically altered mouse models...* [A-3; Circulation 2002; 105: 85]

This reflects a frequent pattern in science article summarization whereby an early FT sentence poses a general unanswered question, issue or need. Sometimes the lexical item expressing the question, issue or need is verbalized in a sentence using a **Func₀** lexical function. A following FT sentence states the focus of the research, linking it explicitly or implicitly via a **Purpose** relation to the first sentence. This FT segment is then summarized in the abstract via a single compound sentence using a **Real₂** verb calculated from the underlying question-type noun associated with the first FT sentence⁴. The general pattern can be loosely stated as:

AB sentence with content:

[*For the researchers to Real₂(N) the N which is not yet Real₂(N)-ed, they did X*],

is derived from the two FT sentences with respective contents:

S-1: [*N Func₀(N), which is not yet Real₂(N)-ed*]

linked explicitly or implicitly as the **Purpose** or cause of

S-2: [*The researchers did X*]

In the example cited in (Kittredge 2002) the AB sentence speaks about *developing a method* (= X) in order to *fulfill* (Real₂) a *need* (= N), while the two FT sentences speak about a *need* which *exists* (Func₀) and the development of the method, respectively.

⁴ In the case here, the explicit noun from the first FT sentence is *issue*, but the choice of **Real₂(N)** = *determine* would seem to indicate that N = (*answer to the question* (whether...)). A plausible solution which accounts for all the details would include several more steps, including a shift in emphasis in the abstract from addressing the issue to answering the particular question raised.

To conclude, we have presented an overview of the range of condensation processes which one can find in an author-written abstract. We have highlighted the advantages of using LFs and paraphrasing rules in order to formalize substitution, compression and deletion processes used by author-abstractors. Nonetheless, we need to study larger corpora in order to be able to make solid generalizations concerning these processes, identify more precisely the range of LFs involved, extract domain-specific LFs, and develop a calculus which includes both simplification and paraphrase operations. In order to arrive at a more complete model of condensation, we finally have to consider the contributions of text linguistics, including RST, to be able to constrain aggregation operations and other multi-sentence paraphrases, as an author-abstractor would process them. Knowing that it is no simple task to automatically obtain an MTT representation of text or to determine implicit rhetorical relations between sentences, we do not pretend to propose an easily computable model, and recognize that our research is in its early stages.

References

- Apresjan Ju. D., Boguslavski I. M., Iomdin L. L., Tsinman L. L. Lexical Functions as a Tool of ETAP-3. *Proceedings of MTT 2003*, Paris, 2003.
- Barzilay R., Elhadad M. Using Lexical Chains for Text Summarization. In: ACL'97/EACL'97. Madrid: Association for Computational Linguistics, 1997. P. 82—88.
- Bélanger P. *Exploration des procédés de condensation pour le résumé de texte grâce à l'application des formalismes de la Théorie Sens-Texte* MA thesis, Department of Linguistics, Université de Montréal, 2003.
- Chuah C.-K. *Linguistic Processes for Content Condensation in Abstracting Scientific Texts* PhD thesis. Université de Montréal, 2001.
- Hovy E. H. Planning Coherent Multisentential Text. In *Proceedings of 26th ACL Conference*. Buffalo, NY, 1988.
- Jing H., McKeown K. Cut and Paste Based Text Summarization. *Proceedings of ANLP 2000*. 2000. P. 178—185.
- Kittredge R. Paraphrasing for Condensation in Journal Abstracting. *Journal of Biomedical Informatics* 2002. 35(4). P. 265—277.
- Kupiec J., Pedersen J., Chen F. A Trainable Document Summarizer. In: Fox, E. A.; Ingwersen, P.; Fidel, R., eds. 18th Annual International ACM SIGIR Conference Proceedings: Research and Development in Information Retrieval. Seattle, WA. New York, NY: The Association for Computing Machinery, 1995. P. 68—73.
- Lee W., Evens M. Generating Cohesive Text Using Lexical Functions. In: Wanner L., editor. *Lexical Functions in Lexicography and Natural Language Processing*. John Benjamins, 1996. P. 299—306.

- Mann W., Thompson S. Rhetorical Structure Theory: Towards a functional theory of text organization. *TEXT* 1988. 8(3). P. 243—281.
- Marcu D. From Discourse Structures to Text Summaries. In: ACL'97/EACL'97. Madrid: Association for Computational Linguistics, 1997. P. 82—88.
- Maybury M. Generating Summaries from Event Data. *International Journal of Information Processing and Management: Special Issue on Text Summarization*. 31(5). 1995. P. 735—751.
- Mel'čuk I. A. Lexical Functions: A Tool for the Description of Lexical Relations in a Lexicon. In: Wanner L., editor. *Lexical Functions in Lexicography and Natural Language Processing*. John Benjamins, 1996. P. 37—102.
- Mel'čuk I. A. Paraphrase et Lexique : La Théorie Sens-Texte et la Dictionnaire Explicatif et Combinatoire. In: *Dictionnaire explicatif et combinatoire du français contemporain*. Vol. III. P. 9—66. Les presses de l'Université de Montréal, 1992.
- Miličević J. *Modélisation sémantique, syntaxique et lexicale de la paraphrase*. Ph. D thesis. Université de Montréal, 2003.
- Mitkov R. How could rhetorical relations be used in machine translation? *Proceedings of the ACL Workshop on Intentionality and Structure in Discourse Relations*, Ohio State University, Columbus, Ohio, 1993.
- Paice C. Constructing Literature Abstracts by Computer: Techniques and Prospects. *Info. Processing and Management*. 1990. 26(1). P. 171—186.
- Saggion H. *Génération automatique de résumés par analyse selective*. Ph. D thesis. Université de Montréal, 2000.
- Sparck-Jones K. Automatic Summarizing: Factors and Directions. In *Advances in Automatic Text Summarization*, Many I. & Maybury M. (eds.) MIT Press, 1999. P. 1—12.

Противопоставление объективной и субъективной причины в семантике причинных глаголов¹

О. Ю. Богуславская

Институт русского языка им. В. В. Виноградова РАН
boguslavskaya@inbox.ru

Семантика нескольких русских глаголов со значением причины описана на фоне других манифестаций значения причины. Показано, что основные семантические противопоставления, характерные для причинной и целевой лексики, релевантны и для этих глаголов. Речь идет о противопоставлении объективной и субъективной причины (то есть, причины-основания) и непосредственной и опосредованной причины. Выявлены ограничения на не-действия в роли исходной ситуации. Обсуждается форма толкования.

The «objective cause» vs. «subjective cause» opposition in the semantics of Russian verbs. The paper is devoted to semantics of several Russian verbs which include the component of causation. It is described against a background of other manifestations of this meaning. It is shown that the major semantic oppositions of objective vs. subjective cause and immediate vs. mediate cause, which are characteristic of other types of causatives and telic words, are also relevant for the words in question. Constraints on non-actions in the capacity of the initial situation have been discovered. The format of the semantic definition is discussed.

0. Введение

Работа продолжает цикл исследований причинно-следственных отношений и способов их выражения в русском языке, которые ве-

¹ Работа выполнена при финансовой поддержке программы фундаментальных исследований отделения историко-филологических исследований РАН «История, языки и литературы славянских народов в мировом социокультурном контексте», раздел 4.15 «Системообразующие смыслы в русском языке», Российского гуманитарного научного фонда, грант 02-04-00306а, а также гранта Президента РФ на поддержку ведущих научных школ № НШ-1576.2003.6.

дуются автором совместно с И. Б. Левонтиной в русле традиций Московской семантической школы. Научный фон и основные принципы, лежащие в их основе можно почерпнуть из вводных материалов к словарю НОСС 2004, написанных Ю. Д. Апресяном. Более узкие теоретические рамки исследования изложены в статье Богуславская, Левонтина 2004.

В ходе работы над выпусками НОСС автором были исследованы и так или иначе описаны некоторые причинные существительные (*основание, резон, мотив, мотивация, мотивировка*, а также *повод, предлог*), И. Левонтина описала причинные предлоги. Были выявлены основные семантические противопоставления, организующие семантическое поле причины; см. Богуславская, Левонтина 2004².

В настоящей работе группа глаголов с чисто причинным значением рассматривается на фоне других лексических единиц, так или иначе указывающих на причину. Показано, что для них релевантны те же противопоставления.

Причинно-следственные отношения в общем случае предполагают наличие двух ситуаций, исходной и производной. Тот или иной тип каузальных отношений между исходной и производной ситуацией может накладывать определенные ограничения на тип каждой из этих ситуаций, а также на ее участников.

Приведем список исследуемых глаголов с краткими поясняющими примерами. Все рассматриваемые глаголы многозначны. Они рассматриваются в дальнейшем лишь в одном из своих значений.

Вести [несобственная форма совершенного вида (СВ) *привести*]: *Трусость ведет к предательству*.

Приводить [СВ *привести*]: *Это ни к чему не приведет*.

Влечь (за собой) [СВ *повлечь за собой*]: *Малая ложь влечет за собой большую*.

Вызывать [СВ *вызвать*]: *Такой ответ мог вызвать только недоумение*.

Порождать [СВ *породить*]: *Недостаток информации порождает слухи*.

Рождать [СВ *родить*]: *Спрос рождает предложение*.

Обуславливать (морфологический вариант *обуславливать*) [СВ *обусловить*]: *Распад государства обусловлен несколькими причинами*.

Можно было бы ожидать, что эти глаголы будут иметь одинаковое значение, чисто причинное, отличаясь только сочетаемостью.

² Многие из выделяемых нами семантических противопоставлений содержатся также в работе Иорданская, Мельчук 1996, посвященной описанию семантики русских причинных предлогов.

Это, однако, неверно. Хотя все рассматриваемые глаголы близки друг к другу и часто взаимозаменяемы, между ними имеются и существенные семантические различия.

1. Противопоставление объективной и субъективной причины

Первый признак, различающий рассматриваемую группу глаголов — противопоставление объективной и субъективной причины. Это семантическое противопоставление подробно описано в работе Богуславская 2003, посвященной многозначности существительного *причина*.

Существительное *причина 1* указывает на то, что при условии существования исходной ситуации *обязательно* реализуется и определенная производная ситуация.

- (1) *По той же причине — сильный жар и головная боль — я плохо помню, что я делал на Сухаревке, хотя и провел там почти целый день* (В. Каверин, Два капитана);
- (2) *Причина смерти — удар ножом в сердце* (М. Булгаков, Собачье сердце).

Причина 1 описывает, таким образом, объективную связь между двумя ситуациями.

Причина 2 (а также ее синонимы *основание, резон, мотив*), напротив, указывают лишь на то, что исходная ситуация так или иначе воздействует на субъекта производной ситуации, но реализация производной ситуации зависит от него. Поэтому условно мы говорим в этом случае о субъективной причине.

- (3) *В год своего сорокалетия Ахимас Вельде стал подумывать, не пора ли удалиться от дел. Нет, он не пресытился работой — причина была другая. Из работы ушел смысл* (Б. Акунин, Смерть Ахиллеса).

Лексема *причина 1*, указывающая на объективную причинно-следственную связь между двумя ситуациями, имеет две валентности, валентность исходной ситуации и валентность производной ситуации. Валентность производной ситуации выражается существительным в родительном падеже (*причина смерти, причина землетрясения*).

У лексемы *причина 2* — три валентности: валентности исходной и производной ситуаций и валентность субъекта.

Валентность субъекта при лексеме *причина 2* не может быть выражена непосредственным синтаксическим зависимым, но может выражаться предложно-именной группой *у + РОД*, подчиненной лексико-функциональному глаголу.

- (4) *У домашних тоже были причины — у каждого своя, относиться к новорожденному более внимательно, чем к его двухлетнему брату Дмитрию* (М. Горький, Жизнь Клима Самгина).

Валентность производной ситуации при лексеме *причина 2* может быть выражена несколькими способами. Во-первых, как и в случае лексемы *причина 1*, она может быть выражена существительным в родительном падеже.

- (5) *Я совершенно не понимаю — причин такого резкого обращения со мной* (М. Булгаков, Мастер и Маргарита).

Если *причина 2* подчинена лексико-функциональным глаголам типа *быть* (в экзистенциальном значении или в функции связи), *иметь, давать, видеть, послужить, найти, найтись*, то валентность производной ситуации может быть выражена кроме того инфинитивом, придаточным предложением, вводимым союзом *чтобы*, а также предложно-именными группами *для + РОД* и *к + ДАТ*.

- (6) *Если у тебя есть серьезные причины остаться у себя, хотя на Сивцевом вам будет в тысячу раз удобнее, ты должен ей сказать, и баста* (В. Каверин, Два капитана);
(7) *У меня были свои причины, чтобы избежать встреч с нею*;
(8) *Это еще не причина для разговоров о смерти и тому подобном* (А. и Б. Стругацкие, Улитка на склоне);
(9) *Словом, причин к войне против Наполеона нашлось достаточно даже и помимо желания как-нибудь отплатить за аустерлицкий разгром и позор* (Е. Тарле, Наполеон).

Во всех этих конструкциях лексема *причина 1*, указывающая на объективную причину производной ситуации безотносительно к какому бы то ни было субъекту, невозможна.

Разный состав валентностей — самый сильный аргумент в пользу различения двух лексем, объективной и субъективной причины. Прочие языковые свидетельства в пользу выделения двух таких значений у слова *причина* (морфологические, сочетаемостные, парадигматические) см. в работе Богуславская 2003.

Среди рассматриваемых нами глаголов три лексемы *вести*, *приводить* и *влечь за собой* указывают на то, что при условии существования исходной ситуации *объективно* реализуется и определенная производная ситуация³.

- (10) *Начавшееся окисление ведет к образованию эфирных масел, отвечающих за запах и вкус конечного напитка* (ИТАР-ТАСС Экспресс, 1996, вып. 36);
- (11) *А не приводят ли мои рассуждения о юморе к некоторой опасной потере чувства юмора?* (Ф. Искандер, Сандро из Чегема);
- (12) *Хвастовство не влечет за собой никаких неприятных последствий* (В. Аксенов, Пора, мой друг, пора).

Ср. также неприемлемые сочетания **вести* <**приводить*> к *беспокойству*, **влечь за собой чувство страха*.

Глаголы *вызывать*, *порождать*, *рождать* и *обуславливать*, в отличие от рассмотренных выше, указывают либо на *объективную* связь между ситуациями, не зависящую от воли людей, либо на то, что исходная ситуация так или иначе воздействует на *субъекта* производной ситуации, который в принципе может сопротивляться этому воздействию. Во втором случае, таким образом, реализация производной ситуации зависит как от исходной ситуации, так и от субъекта ситуации производной. Ср. пары (13)—(16), где примеры серии а) иллюстрируют объективную причинную связь, а примеры серии б) — субъективную.

- (13) а) *Увы, ранения, вызвавшие ее смерть, не соответствовали очень легким повреждениям, которые потерпел автомобиль* (В. Набоков, Лолита);
б) *Это тонкое замечание вызвало полное сочувствие у женской части аудитории* (Б. Акунин, Пелагия и белый бульдог);
- (14) а) *Вегетарианство не только обедняет человеческий организм, не предоставляя ему необходимых питательных веществ, но и даже несет прямую угрозу его существованию, порождая различные заболевания* (ИТАР-ТАСС Экспресс, 1996, вып. 7);

³ В конструкциях типа *Спокойствие жены приводило его в ярость* представлено близкое, но другое значение глагола *приводить*. Его формальное отличие от рассматриваемого здесь значения — управление предложно-именной группой *в* + СУЩ ВИН.

- б) *Эти постоянные искания приюта порождали странное чувство бездомности* (В. Набоков, *Другие берега*);
- (15) а) *Минувшие десятилетия оснастили наш политический и экономический язык терминологией фронта. Иначе, видимо, и быть не могло: казарменные методы управления рождали соответствующий стиль* (Политическая метафора);
- б) *Произведение литературы только тогда становится близким читателю, когда оно рождает отклик в душе* (Б. Панкин, *Человек начинается с утра*);
- (16) а) *Радуга первого порядка, которую обычно наблюдают после дождя, обусловлена резким увеличением интенсивности рассеянного света при угле 138°; она намного ярче радуг более высоких порядков* (Уппсальский корпус);
- б) *Самооценка поэта, вернее его отношение к своим стихам в период работы, всегда пристрастна и обусловлена множеством сложных причин* (Н. Мандельштам, *Воспоминания*).

Е. В. Падучева, исследуя типы каузальных отношений в семантической структуре лексемы (Падучева 1994), выделяет в качестве особого случая ситуацию неполного контроля субъекта исходной ситуации над производной ситуацией. Она относит к этому классу глаголы типа *решить (задачу)*, *уговорить*, *убедить*, *поймать*, «семантика которых предполагает объект или лицо, оказывающее, в той или иной мере, сопротивление» каузирующему воздействию (Падучева 1994: 4).

Несмотря на некоторое сходство ситуации неполного контроля в такой формулировке и ситуации субъективной причины, нужно подчеркнуть, что они отнюдь не совпадают.

Ситуация неполного контроля предполагает, что каузатором является субъект действия или деятельности, имеющий определенную цель (Падучева 1994: 11—12). То есть ситуация неполного контроля накладывает ограничения на исходную ситуацию. Между тем, субъективность причины накладывает ограничения только на производную ситуацию. В частности, о субъективной причине мы говорим в том числе и тогда, когда исходная ситуация лишена целеполагающего субъекта. Ср. (17).

- (17) *Ночное небо рождало тревогу, книги на прилавках вызвали решительные чувства: я лучше могу, я все могу!* (В. Аксенов, *Пора, мой друг, пора*).

2. Непосредственная и опосредованная каузация

Глаголы *вести* и *приводить* в соответствии со своей внутренней формой обычно предполагают, что производная ситуация следует за исходной не непосредственно, а через промежуточные звенья причинно-следственной цепочки или в результате постепенного развития производной ситуации.

- (18) *Монополия ведет к загниванию* (Уппсальский корпус);
- (19) *Кто же мог предположить, что перемены в политике и экономике приведут к постепенному умиранию спорта* (А. Маринина, Мужские игры);
- (20) *К чему приводит ранняя судимость? 60 % осужденных совершают повторные преступления* (Уппсальский корпус).

В отличие от этих двух глаголов лексемы *вызывать*, *порождать*, *рождать*, *обуславливать*, а также (вопреки внутренней форме!) *влечь за собой* допускают одновременное существование исходной и производной ситуации.

- (21) *Русская угрюмость вызвана иным поведенческим этикетом — меньшей приветливостью и ослабленной социальной ролью вежливости* (Б. Акунин, Алтын-Толобас);
- (22) *Всеобщая невиновность порождает и всеобщее бездействие* (А. Солженицын, Архипелаг ГУЛАГ);
- (23) *Таким образом, нервный ребенок худеет по двум причинам: нервность, с одной стороны, снижает его аппетит, а с другой — обуславливает большие затраты энергии* (Б. Спок, Ребенок и уход за ним, пер. с англ.);
- (24) *Малый тираж влечет за собой удорожание издания, а дороговизна книг в свою очередь еще более сокращает их распространение* (В. Ходасевич, Литература в изгнании).

При этом глаголы *вызывать*, *порождать*, *рождать*, *влечь за собой*, но не *обуславливать* допускают в то же время и наличие промежуточных звеньев в причинно-следственной цепочке и, как следствие этого, наличие временного зазора между исходной и производной ситуацией.

- (25) *Снижение рождаемости, может быть, и целительное для человечества, порождает множество проблем для нынешних европейских государств, в которых прекращается рост населения* («Знамя», 2002, № 5);

- (26) *Переселение было вызвано каким-то неизвестным массовым заболеванием козлотуров* (Ф. Искандер, Сандро из Чегема);
 (27) *Завоевание столицы неприятельского государства обычно влекло за собой прекращение организованного сопротивления* (Е. Тарле, Москва в Отечественной войне 1812 года).

Несмотря на это, сочетания с наречиями и наречными выражениями типа *постепенно, в конечном счете, в конце концов*, подчеркивающими опосредованность каузации и наличие промежуточных звеньев, для этих глаголов нехарактерны. Ср. неестественные примеры *Это породило "в конце концов" <постепенно> определенные проблемы; Массовое заболевание "постепенно" <"в конечном счете"> вызвало переселение.*

Между тем для глаголов *вести* и *приводить* такие сочетания как раз очень характерны.

- (28) *Реализация этой программы ведет в конечном счете к созданию безъядерного ненасильственного мира* (Уппсальский корпус);
 (29) *Любая из этих ошибок — в конце концов приводит к путанице понятий* (Ф. Искандер. Созвездие Козлотура);
 (30) *Раздвигая ткани, сдавливая и разрушая сосуды, гной постепенно приводит к некрозу тканей* (В. Кунин, Ты мне только пиши).

Для глагола *обуславливать* такая сочетаемость невозможна; ср. неприемлемое **постепенно обусловили*. Этот глагол способен указывать только на непосредственную причинно-следственную связь.

3. Активность исходной ситуации

В работе Богуславская, Левонтина 2004 показано, что для многих целевых слов существенно противопоставление действий и воздержания от действий (не-действий).

Например, слово *цель* плохо совместимо с идеей воздержания от действия. Ср. неприемлемость примеров типа (31) и (32):

- (31) **В чем цель вашего молчания <бездействия>?*
 (32) **С какой целью ты избегаешь полемики <не обращаешь внимания на замечания>?*

А вот для целевого союза *чтобы* запрета на подобные контексты не существует. Ср. стандартные примеры в (33).

- (33) *Я избегаю полемики <игнорирую его шуточки>, чтобы сохранить нормальные отношения <не накалять обстановку>.*

Аналогичное противопоставление, хотя и в гораздо более слабом виде, релевантно и для рассматриваемых глаголов.

Если исходная ситуация — статична, является не-действием, а производная ситуация — физические действия или деятельность, то предпочтительнее использование глаголов *вести, приводить, влечь за собой* и *обуславливать*, чем *вызывать* и *порождать*.

- (34) а) *Недостаток средств привел к остановке строительства;*
 б) *Недостаток средств повлек за собой остановку строительства;*
 в) *Остановка строительства обусловлена недостатком средств;*
 г) *Недостаток средств вызвал остановку строительства;*
 д) **Недостаток средств породил остановку строительства.*
- (35) а) *Не всякий Вас как я поймет, / К беде неопытность ведет (А. С. Пушкин, Евгений Онегин);*
 б) *Неопытность влечет за собой беды;*
 в) *Неопытность вызывает беды;*
 г) *Неопытность порождает беды;*
- (36) а) *Его упорное молчание на следствии повлекло за собой ужесточение наказания;*
 б) *Его упорное молчание на следствии вызвало ужесточение наказания;*
 в) **Его упорное молчание на следствии породило ужесточение наказания.*
- (37) *Вместе с тем «Душенька» значительно разнится от своего первоисточника, чем, разумеется, и обусловлена ее слава (В. Ходасевич, Богданович).*

Нужно отметить, впрочем, что появление в этих контекстах глагола *вызывать* все-таки допустимо или почти допустимо. А неприемлемость примеров с *порождать* можно объяснить другими семантическими ограничениями, на которых мы здесь останавливаться не будем⁴.

⁴ Глагол *порождать* в соответствии со своей внутренней формой используется предпочтительно в тех случаях, когда речь идет о создании чего-то: нового информационного объекта, нового ментального состояния, нового положения вещей. Ср. *Эта идея была порождена гуманизмом и демократическими тенденциями девятнадцатого века* (Н. Мандельштам, Воспоминания); *Слишком подробное алиби порождает новые подозрения*

Если производная ситуация не является физическим действием или деятельностью, то противопоставление действий и не-действий стирается или ослабляется.

- (38) *Все рассказы о виллах, принадлежавших Горькому, и о чуть ли не оргиях, там происходивших, — ложь, для меня просто смешная, порожденная литературной завистью и подхваченная политической враждой* (В. Ходасевич, Горький);
- (39) *Неявка избирателей привела к тому, что выборы были объявлены несостоявшимися* (ИТАР-ТАСС Экспресс, 1997, вып. 6) [замена на *вызывать* или *порождать* затруднена].

4. Пассивная конструкция

Рассматриваемые глаголы различаются по степени характерности для них форм пассива. Для глаголов *вызывать*, *порождать*, *обусловливать* пассивные конструкции характернее активных. Это связано, по-видимому, с тем, что для этих глаголов естественнее помещать в позицию темы производную ситуацию.

- (40) *Обсуждая это смягчение, мы решили, что оно вызвано было инструкцией «сохранить»* (Н. Мандельштам, Воспоминания);
- (41) *Миф о всезнающих удавах порожден кроличьим страхом* (Ф. Искандер, Кролики и удавы);
- (42) *Я уже сказал, что у этой женщины низкий духовный уровень, но он обусловлен не примитивностью, а дисгармоничностью ее натуры* (К. Чапек, Тайна почерка, пер. Т. Аксель и О. Молочковского).

5. Проблемы толкования

В работе Богуславская, Левонтина 2004 обсуждалась проблема выбора примитива среди различных лексических манифестаций причины. Решительно отвергнуты были предлог *из-за* и искусственный глагол *каузировать*⁵. Приводились аргументы в пользу выбора

(Ф. Искандер, Сандро из Чегема); *Сочетание воды и камня порождает здесь особую, величественную атмосферу* (С. Довлатов, Чемодан).

⁵ Веские аргументы против выбора на роль примитива глагола *каузировать* содержит, по нашему мнению, статья Апресян 1998. В работе Ю. Д. Ап-

на эту роль существительного *причина* (разумеется, в его первом, объективном значении) и союза *потому что*.

Попытка истолковать рассматриваемые глаголы через существительное *причина* наталкивается на существенные затруднения.

Во-первых, в качестве примитива может рассматриваться только первое значение существительного, то есть объективная причина. Тем самым это могло бы быть приемлемым решением для глаголов *приводить*, *вести*, *влекать за собой*, описывающих объективную причинно-следственную связь, но не для глаголов *вызывать*, *(по)рождать* и *обуславливать*, способных указывать как на объективную, так и на субъективную причину.

Кроме того, как показано в работе Богуславская, Левонтина 2004 русское слово *причина*, подобно своему конверсиву *следствие*⁶ и в отличие от союза *потому что*, небезразлично к идее целенаправленности. *Причина* (как и *следствие*) плохо подставляется в контексты, где результат совпадает с целью субъекта исходной ситуации.

- (43) а) Он победил в этом бою, потому что правильно выбрал оружие;
 б) ?? *Причина его победы — правильный выбор оружия*;
 в) ?? *Его победа — следствие правильного выбора оружия*.

Поэтому смысл лексемы *причина* 1 сложнее, чем смысл союза *потому что*.

X — причина Y -а = ' Y имеет место, потому что имеет или ранее имела место ситуация X ; причем, если у ситуации X есть субъект, то Y не является его целью'.

Как видно из толкования, слово *причина* истолковано здесь через *потому что*, но его смысл не исчерпывается смыслом 'потому что'.

Рассматриваемые глаголы, как и союз *потому что*, свободны от этого ограничения.

- (44) *Сильный шахматный ход, хитрая жертва, приводящая в дальнейшем к выигрышу* (Ф. Искандер, Сандро из Чегема);
 (45) *Уменьшение коэффициента трения повлекло за собой значительную экономию горючего*.

ресяна, написанной в 1976 году, но опубликованной только 20 лет спустя (Апресян 1998), показывается, что в реальной лексикографической практике уже в те годы глагол *каузировать* использовался в восьми разных значениях.

⁶ О семантике слов *следствие* и *последствие* и их соотношении со словом *результат* см. Богуславская, Левонтина 2001 и Богуславская, Левонтина 2002.

Поэтому рассматриваемые глаголы следует толковать через союз *потому что*, а не через существительное *причина*.

Литература

- Апресян 1998 — *Апресян Ю. Д.* Каузативы или конверсивы? // Типология, грамматика, семантика: К 65-летию В. С. Храковского. СПб.: Наука, 1998. С. 273—281.
- Апресян 2001 — *Апресян Ю. Д.* Значение и употребление // *Вопр. языкознания*. 2001. № 4.
- Богуславская 2003 — *Богуславская О. Ю.* Структура значения существительного *причина* // *Русистика на пороге XXI века: проблемы и перспективы: Материалы международной конференции (Москва, 8—10 июня 2002 г.)*. С. 102—105.
- Богуславская, Левонтина 2001 — *Богуславская О. Ю., Левонтина И. Б.* Задумается о последствиях // *Труды Международного семинара Диалог'2001 по компьютерной лингвистике и ее приложениям*. Аксаково, 2001. С. 21—27.
- Богуславская, Левонтина 2002 — *Богуславская О. Ю., Левонтина И. Б.* Подведение итогов в русском языке // *Логический анализ языка. Семантика начала и конца* / Ред. Н. Д. Арутюнова. М.: Индрик, 2002. С. 36—49.
- Богуславская, Левонтина 2004 — *Богуславская О. Ю., Левонтина И. Б.* Смыслы 'причина' и 'цель' в естественном языке // *Вопр. языкознания*. № 2. 2004. С. 68—88.
- Иорданская, Мельчук 1996 — *Иорданская Л. Н., Мельчук И. А.* К семантике русских причинных предлогов // *Московский лингвистический журнал*. Т. 2. М., 1996.
- Мельчук, Жолковский 1984 — *Мельчук И. А., Жолковский А. К.* Толково-комбинаторный словарь современного русского литературного языка. Опыты описания русской лексики. Wiener Slavistischer Almanach. Sonderband 14. Вена, 1984.
- НОСС 2004 — Новый объяснительный словарь синонимов русского языка. 2-е изд. / Под общ. руков. акад. Ю. Д. Апресяна. Москва; Вена, 2004.
- Падучева 1994 — *Падучева Е. В.* Типы каузальных отношений в семантической структуре лексики // *Russian Linguistics*. I. 18. P. 1—16.

Actant Properties of Adverbial Derivatives¹

Igor Boguslavsky

Institute for Information Transmission Problems, Russian Academy
of Sciences / Universidad Politécnica de Madrid
bogus@cl.iitp.ru / igor@opera.dia.fi.upm.es

The actant properties of adverbial derivatives (AdvD) are much more diverse than those of other types of syntactic derivatives, such as conversives, S_i , A_i , or verbal adverbs. We show that in a general case a valence slot of an AdvD can be either (a) blocked, or filled (b) by a dependent of the AdvD itself, or (c) by the dominating verb, or (d) somewhere within the clause organized by the dominating verb; in the latter case the AdvD actant may be identified based on its syntactic position, its semantic role with respect to its mother element, or its communicative function.

Актантные свойства адвербиальных дериватов. Актантные свойства адвербиальных дериватов (AdvD) гораздо разнообразнее, чем у других типов синтаксических дериватов, таких, как S_i , A_i или деепричастия. В общем случае валентность AdvD может быть либо (а) заблокирована, либо выражена (б) зависимым самого AdvD или (в) подчиняющим AdvD глаголом или (г) какой-либо именной группой, подчиненной подчиняющему AdvD глаголу. В последнем случае актанта AdvD может быть идентифицирован на основе его синтаксической позиции, семантической роли или коммуникативной функции.

1. Statement of the problem

Syntactic derivation is one of the most direct manifestations of the systemic character of the lexicon. As is well-known, language is capable of representing the same meaning or several very close meanings by means of words belonging to different grammatical classes. It is possible to transform words of a certain grammatical category to those of other grammatical categories without significant modification of their lexical meaning. A convenient notation for this process was proposed by I. Mel'čuk and

¹ The author thanks the Ramon y Cajal Program of the Spanish Ministry of Science and Technology for support of this work.

A. Zholkovsky within the system of lexical functions (Mel'čuk 1974: 84, 86, 89—91). Ju. Apresjan investigated this type of synonymy in more detail in his "Lexical semantics" (1974). Very large lexicographic material was accumulated in Russian and French EC dictionaries (Mel'čuk et al. 1984a, Mel'čuk et al. 1984b, Mel'čuk et al. 1988 and Mel'čuk et al. 1992).

Our subject will be ADVERBIAL syntactic derivatives (AdvD) of nouns. As a rule, such a derivative is expressed by a phrase consisting of a preposition and the initial noun (below, keyword). We will distinguish between two types of AdvD: "pure" derivatives, which do not contain any meaning components absent in the meaning of the keyword, and "semantically augmented" derivatives, for which the reverse is true. As an example of the latter, let us consider the phrase *pod imenem X* 'under the name of X' as represented in

- (1) *Napoleon exal pod imenem gerzoga Vičentskogo, to est' Kolenkura* (Danilevskij, Sožžennaja Moskva) 'Napoleon was travelling under the name of duke of Vicenza, that is of Colencour'.

The meaning of this phrase contains a component of concealing the true name. *Pod imenem X* 'under the name of X' does not mean that the person in question has the name of X, but rather that this person or somebody else wants other people to refer to him/her as X while the speaker knows, believes or admits that this is not the true name. Obviously, the noun *imja* 'name' has no such component (as opposed to *pseudonym* or *nickname*). It cannot even be ascribed to the preposition *pod* 'under', either, since in (1) this preposition has obviously the same meaning as in phrases *pod nazvanijem* (zagolovkom, rubrikoj) 'under the title (heading)' to which the component of concealment is completely alien. Phrases like *pod imenem* that are to a certain extent idiomatic are lexical units in their own right and have their own entries in the dictionary.

Phrases without any semantic augmentations have more rights to the status of pure syntactic derivatives. From the viewpoint of the theory of phraseology developed in Mel'čuk 1995, they belong to the class of collocations and should be represented in the ECD within the entries of their keywords (Mel'čuk 1995: 184). We will be mostly interested in the relationship between the actantial structures of the keyword and the derivative. This question is relevant both for the pure and semantically augmented derivatives.

The actantial structure of words that are not correlates of any lexical function is described in two ways. Active² valence slots are conveniently

² A valence slot of lexeme *L* is active if it is filled by an expression which is syntactically dominated by *L*. A passive valence slot is filled by an expression that syntactically dominates *L*. Cf. Boguslavsky 1996 for more details.

described by means of government patterns (cf. Apresjan 1974, Mel'čuk 1974). Actants that correspond to passive valence slots are not GOVERNED by the predicate and therefore cannot be served by a government pattern. They need special "scope rules" (Boguslavsky 1985, 1996).

For some lexical functions more information is needed. Such lexical functions as Conv_{ij} , Oper_i , Func_i , Labor_{ij} , S_i , A_i require that their actantial structure be brought into correlation with the actantial structure of the keyword. I will demonstrate this with a simple example. Like any verb, *buy* and *sell* are supplied with a government pattern that lists all their valence slots and their means of expression. On the other hand, their being conversives implies that their LF-description should indicate the correlation between their actantial structures. Namely, the first actant of *sell* ("who sells?") corresponds to the third actant of *buy* ("from whom buys?"), the third actant of *sell* ("to whom?") — to the first actant of *buy* ("who buys?"), while the second and the fourth actants ("what?" and "for how much?") occupy the same positions within government patterns of both verbs. This correlation is rendered by the numerical index attached to the LF symbol: $\text{Conv}_{3214}(\text{sell}) = \text{buy}$. The index is composed in the following way: the j th position in the index is occupied by i if the j th actant of the converse corresponds to the i th actant of the keyword³.

For AdvDs the problem of the correlation of their actantial structure with that of the keyword is particularly acute. If *sell* is undoubtedly a lexical unit in its own right and is entitled to have its own government pattern, adverbial collocations of the type *v jarosti* 'in anger', *po privyčce* 'by habit', *s akcentom* 'with an accent' or *pod predlogom* 'under the pretext of' are not usually treated as separate lexical units. It is assumed that all necessary information about their meaning and use should be formulated in the lexical entry of the keyword. To what extent does an AdvD inherit the actantial structure of the keyword? If not in full, how should its actantial structure be described in the dictionary?

A more practical question is this. If the sentence contains predicate L , its actants' position and form are specified by the government pattern of L . If, however, L is represented by its AdvD, what can we say about the position and form of its actants?

Before answering this question, we will recall how syntactic derivatives, and AdvD in particular, are treated in the theory of lexical functions.

³ Yu. Apresjan demonstrated the need to enlarge the numerical index of lexical functions of the Oper_i — Func_i — Labor_{ij} and Real_i — Fact_i — Labreal_{ij} families. The traditional index consisting of one or two figures does not fully specify the actantial structure of the corresponding verbs (Apresjan 2001).

2. Derivatives in the theory of lexical functions

Two types of syntactic derivation are distinguished: “zero” and “actant” derivation. Zero syntactic derivatives (S_0 , V_0 , A_0 и Adv_0) have the same meaning as the keyword but belong to a different part of speech. Actant derivatives (S_i ⁴, A_i and Adv_i) are oriented towards one of the actants of the keyword in the following sense.

S_i is a standard name of the i th actant of the keyword ($S_1(\text{teach}) = \text{teacher}$, $S_2(\text{teach}) = (\text{subject}) \text{matter}$ [in high school], $S_3(\text{teach}) = \text{pupil}$).

A_i also has a bearing to the i th actant, but in the adjectival syntactic status. This means that its typical syntactic function is to modify a noun that fills the i th valence slot of the keyword. A grammaticalized way of expressing A_i is participles. A_1 is equivalent to an active participle, and A_2 , to a passive participle. For example, A_1 of the verb *to burn* is either an active participle *burning* or a prepositional phrase *on fire* — cf. *a burning house*, *a house on fire* (Apresjan 1974: 49).

Things are more complicated with the adverbial actant derivatives (Adv_i). This function is defined as follows:

« Adv_i is a typical designation of the action of the i th participant of the situation by way of an adverbial: $Adv_1(\text{skorost})$ ‘speed’ = *so* ‘at’ [*skorostju* ‘a speed’], cf. *mčat’sja so skorostju* ... ‘rush at a speed...’; $Adv_2(\text{somnevat’sja})$ ‘doubt’ = *vrjad li* ‘hardly’» (Mel’čuk et al. 1984: 84).

« $Adv_1, Adv_2, \dots$: adverbe typique pour caractériser le comportement du premier, deuxième, ... actant du mot clé, autrement dit l’adverbe qui signifie ‘étant le premier, deuxième, ... actant de’». (Mel’čuk et al. 1984: 9).

« Adv_i — determining property of the action by the i -th (DSynt-) actant of L according to its role in the situation denoted by L . Adv_1 is roughly equivalent to an active verbal adverb (‘while L-ing’) and Adv_2 , to a passive verbal adverb (‘while being L-ed’).

$Adv_1(\text{anger}) = \text{with} [\text{anger}], // \text{angrily}$

$Adv_2(\text{bombard}) = \text{under bombardment}$

$Adv_1(\text{speed}) = \text{at} [\text{a speed of...}]$ » (Mel’čuk 1996: 55).

As these definitions show, the only link between the actantial structures of the keyword and the adverbial derivative of the Adv_i type is the i th actant of the keyword. Although it is not stated explicitly, one can presume that the i th actant’s position in the sentence is the position of the subject (or, to be more precise, the first deep syntactic (DSynt) actant)

⁴ The same applies to the variety of S_i — S'_i , introduced in Apresjan 1974: 165—167.

of the verb to which Adv_i is attached. In (2) the first actant of *anger* is obviously *Mary*, the first actant of the verb *reject*, and not *John* or anybody else.

- (2) *Mary rejected John's proposal with anger.*

This is understandable, since lexical function Adv_i is intended to model the behaviour of verbal adverbs — cf. the definition from Mel'čuk 1996 above. As is known, in many languages, in particular in Russian, the main peculiarity of the actant behaviour of verbal adverbs is that one of their DSynt actants cannot be expressed in the surface structure and should be co-referential with the first DSynt actant of the dominating verb. Their other DSynt actants do not have any characteristic properties and are attached to them just as they are attached to the finite form of the verb:

- (3) *Petr pokupaet odezhdu v modnyx magazinax za ogromnye den'gi* 'Peter buys clothes in fashionable shops for a lot of money'.
 (4) *Pokupaja odezhdu v modnyx magazinax za ogromnye den'gi, Petr vechno po ushi v dolgax* 'buying clothes in fashionable shops for a lot of money, Peter is always in debt up to his neck'.

Many languages have absolute constructions which allow the first DSynt actant of the verb to be not-coreferential with the actants of verbal adverbs:

- (5) *His wife buying clothes in fashionable shops, Peter is always in debt up to his neck.*
 (6) Spanish: *Habiendo pasado más de una hora, las piernas comen-zaron a flaquear* 'more than an hour having passed, his legs began to fail'.

These constructions cannot be represented as Adv_i .

3. Actants of adverbial derivatives

If we compare adverbial derivation with other types of syntactic derivation, we will encounter an important difference. Actant structure of such derivatives as $Conv_{ij}$ or S_i can be easily characterized in terms of the actant structure of the keyword. When we pass from the keyword to such a derivative, we may find that an actant either stays in its initial position (*teach mathematics* — *teacher of mathematics*), or changes its number (*the verb dominates the preposition* — *the preposition depends on the*

verb), or gets blocked altogether (*drive home* — **driver home*). However, the syntactic position of the actant can only change in a very limited way. If a valence slot of the keyword is expressible in the sentence with its Conv_{ij} or S_i at all, the actant should either be attached to the derivative directly, or through a copula or other LF-verb (*Peter teaches mathematics* — *Peter is a teacher of mathematics*).

The matters stand differently with AdvDs. Their actant properties are much more diverse than those of Conv_{ij} or S_i , or even of verbal adverbs. In some cases, the position of their actants in the sentence cannot be characterized in purely syntactic terms. In a general case, a valence slot of an AdvD can be either blocked, or:

- filled by a dependent of the AdvD itself;
- filled by the dominating verb;
- filled somewhere within the clause organized by the dominating verb; in this case the AdvD actant may be identified based on:
 - its syntactic position;
 - its semantic role with respect to its mother element;
 - its communicative function.

We will illustrate these possibilities below.

3.1. Valence slots filled by a dependent of AdvD

In the canonical case, AdvD inherits most of the governing properties of the keyword.

- (7a) *skorost' 800 km/čas* 'a speed of 800 km per hour',
- (7b) *Samolet letel so skorostju 800 km/čas* 'the aircraft was flying at a speed of 800 km per hour',
- (8a) *sovet Ivana* 'Ivan's advice',
- (8b) *po sovetu Ivana* 'at Ivan's advice',
- (9a) *Eto podarok ot Viktora* lit. 'this is a present from Victor',
- (9b) *Ja polučil eto v podarok ot brata* lit. 'I got it as a present from my brother'.

In some cases, governing properties of AdvD are different from those of the keyword. Let us consider the pair *predlog* 'pretext' — *pod predlogom* 'on/under the pretext of' that manifests an interesting correlation of actant properties. The noun *predlog* 'pretext' has three valence slots: *P is a pretext for X for doing Q* = 'wishing to do Q, which violates norms of ethics, or politeness, X uses situation P to do P; he thinks that P justifies Q' (Boguslavskaya 2003, slightly modified). When *predlog* is used

without preposition *pod*, it can attach actant Q (= the action carried out) but not P (= false motive). The latter can only be expressed outside the phrase containing *predlog*:

- (10a) *Golovnaja bol'* [P] — *xorošij predlog, čtoby ostat'sja doma* [Q]
 'headache [P] is a good pretext for staying at home [Q]'.
 (10b) **predlog golovnoj boli* [P] 'the pretext of the headache [P]'

AdvD *pod predlogom* has opposite governing properties. Actant P (= false motive) can now be a dependent of AdvD while actant Q (= the action) loses this property and moves to the position of the dominating word:

- (10c) *Ona ostalas' doma* [Q] *pod predlogom golovnoj boli* [P] 'she stayed at home [Q] on the pretext of the headache [P]'.

3.2. Valence slots filled by the dominating verb

Adverbial derivatives of many predicates which have a propositional valence slot fill it by means of the dominating verb. One of the examples is (10c) above. Let us give some examples more.

- (11a) *Otec imeet privyčku rano ložit'sja spat'* 'father has a habit of going to bed early'.
 (11b) *Otec po privyčke rano leg spat'* 'by habit, father went to bed early'.
 (12a) *Ljusja Kinžalova dokazala polnuju sdaču svoix pozicij tem, čto pocelovala Marata v nos* 'Ljusja Kinzhalova proved complete surrender by kissing Marat on the nose'.
 (12b) *V konce koncov sama Ljusja Kinžalova priznala grubost' svoego zamečanija i v dokazatel'stvo polnoj sdači svoix pozicij pocelovala Marata v nos* (Iskander, Malen'kij gigant bol'sogo seksa) 'after all, Ljusja Kinzhalova herself acknowledged that her remark had been rude, and as a proof of complete surrender kissed Marat on the nose'.
 (13a) *On otvetil mnogoznačitel'nym myčaniem* 'he responded with a significant mumble'.
 (13b) *V otvet on čto-to mnogoznačitel'no promyčal* 'in response he mumbled something in a significant manner'.
 (14a) *Ja sčitaju, čto ždat' bol'se nečego* 'I think there is nothing more to wait for'.
 (14b) *Po-moemu, ždat' bol'se nečego* 'in my opinion, there is nothing more to wait for'.

3.3. Valence slots filled by dependents of the dominating verb

If a valence slot of an AdvD is filled by a dependent of the dominating verb, the question arises as to how to specify its position among other dependents of the verb. We will show that this position can be identified based on the syntactic function (3.3.1), semantic role (3.3.2) or communicative function (3.3.3).

3.3.1 Syntactic function

As mentioned in section 2, in the prototypical case of adverbial derivation, that of verbal adverbs, one of the actants of the keyword (the first or the second) is necessarily co-referential with the first DSynt actant (subject) of the dominating verb. Since this actant is not expressible as a dependent of the AdvD, the subject of the dominating verb is its only manifestation in the sentence. In this sense, we can say that the valence slot is filled by the subject of the dominating verb. If it is the first actant of the keyword that is co-referential with the subject, we are dealing with the active verbal adverb (Adv₁, in Mel'čuk's terminology). If it is the second actant, we are facing the passive verbal adverb (Adv₂). If the co-reference requirement is not met, sentences with verbal adverbs are usually anomalous in Russian. Cf. a textbook example **Podjezžaja k stancii, u menja sletela šljapa* 'when (I was) approaching the station, my hat fell down'. With non-verbal AdvDs, this requirement can erode to various degrees. Let me give some examples in the decreasing order of obligatoriness of this requirement.

AdvD *po privyčke* 'by habit, out of habit' requires that its actant ("whose habit?") be the subject of the main verb. Therefore, the verb cannot be passivized:

(15a) *Pereselency po privyčke stroili doma u reki* 'out of habit, resettlers built their houses at the river bank'.

(15b) ^{??}*Doma po privyčke stroilis' u reki* 'out of habit, the houses were built at the river bank'.

AdvDs *po prikazu* 'by smb's order', *po rasporjaženiju* 'by smb's instruction', *po sovetu* 'according to smb's advice', etc. are usually used when the addressee of order, instruction, advice, etc. (who is at the same time the agent of the corresponding action) acts as the subject:

(16a) *Po prikazu verxovnogo komandovanija vojska perešli v nastuplenie* 'by order of high command the troops passed to the offensive'.

However, this time the verb can be passivized and the addressee of the order then loses the subject position:

- (16b) *Po prikazu Šujnskogo byla postroena plotina na rečke Uže, kotoraja protekala čerez Tulu* ‘by Shujskij’s order, a dam was constructed on the Uzha river, which flowed through Tula’.

If instead of passivizing the verb we replace it with its conversive, which is in a sense very similar to the passive, the sentence seems to become less appropriate:

- (17a) *Po prikazu verxovnogo komandovanija sojuznye vojska osuš-čestvili massirovannyj artobstrel protivnika* ‘by order of high command, the allied troops carried out concentrated bombardment of the enemy’
 (17b) ³*Po prikazu verxovnogo komandovanija protivnik podvergsja massirovannomu artobstrelu sojuznymi vojskami* ‘by order of high command, the enemy was subjected to concentrated bombardment by the allied troops’.

The next step of weakening the co-reference requirement is sentence (18) in which the addressee of the order is the first actant of the first actant of the main verb: ‘the command ordered the troops that they should begin the offensive at 3 a.m.’

- (18) *Po prikazu verxovnogo komandovanija nastuplenie načalos’ rovno v 3 časa utra* ‘by order of high command, the offensive began at 3 a.m. sharp’

3.3.2. Semantic role

Yet further along this road are AdvDs *v podarok* ‘as a present’, *v dar* ‘as a gift’, *v nagradu* ‘in reward’. Nouns of the *present / gift / reward* type have three valence slots: the subject of presenting (X), the object (Y) and the recipient (Z). These AdvDs co-occur with a large set of verbs concentrated around the meaning of ‘transfer’: *polučat* ‘receive’, *prinimat* ‘accept’, *trebovat* ‘demand’, *prosit* ‘request’, *prinosit* ‘bring (on foot)’, *privozit* ‘bring (by transport)’, *dostavljat* ‘deliver’, *posylat* ‘send’, *otpravljat* ‘dispatch’, *prednaznačat* ‘intend for’, *žalovat* ‘grant’, *podnosit* ‘offer’, *peredavat* ‘pass (to)’, *vručat* ‘hand over’, *davat* ‘give’, *otdavat* ‘give back’, etc. It is impossible to associate the subject of the dominating verb with any single actant of AdvD, since each of the three actants can perform the role of the subject:

- (19a) *Otec (X) privez žene v podarok ožerelje* ‘Father (X) brought a necklace as a present to his wife’.
 (19b) *Maria (Z) prinjala ožerelje v podarok* ‘Mary (Z) accepted a necklace as a present’.
 (19c) *Ožerelje (Y) dostalos’ ej v podarok ot babuški* ‘the necklace (Y) came to her as a present of her grandmother’.

It is not syntactic constraints that regulate the position of the actants of these AdvD with respect to the main verb but semantic ones. The correlation between the valence slots of AdvD and the main verb can be formulated IN TERMS OF SEMANTIC ROLES as follows: if a valence slot of AdvD which corresponds to semantic role R (agent, objet, recipient) is instantiated in the sentence, it is either filled by an AdvD dependent (as in *v podarok synu* ‘as a present to one’s son’, *v podarok ot otca* ‘as a present from one’s father’), or by a dependent of the main verb which performs the role R with respect to the predicate of transfer within the meaning of the main verb. For example, in (19b) Maria is the recipient of the present because she is the recipient of the predicate ‘give’ in the meaning of the main verb (‘accept’ ≈ ‘agree to be given’). It should be stressed that the semantic role of a noun phrase with respect to the dominating verb may be different from its semantic role with respect to an inner predicate of this verb. For example, in (20)

- (20) *Ivan potreboval polcarstva* ‘Ivan demanded half of the kingdom’
 [= ‘demanded that he were given half of the kingdom’]

Ivan is the agent of demand and at the same time the recipient of giving. What is important for AdvD of the *v podarok* type is the role of the actant with respect to giving. Therefore, in (21) Ivan is the recipient and not the agent of reward:

- (21) *Ivan potreboval sebe v nagradu polcarstva* ‘Ivan demanded half of the kingdom as a reward’

3.3.3. Communicative function

In Boguslavsky 2005 we discussed actantial properties of noun *bol’sinstvo* ‘majority, most of’ and its AdvD *v bol’sinstve* ‘in a majority, mostly’. Now we will only be interested in one of its valence slots — that of the whole, expressed prototypically by preposition *iz* ‘of’ as represented in phrases *the majority of cases*, *most of the students*. In sentences with AdvD *v bol’sinstve* this valence slot is filled, as a rule, by the subject of the dominating verb:

- (22) *Oni byli arestovany i podverglis' v bol'sinstve svoem ssylke v Gvianu i na Sejšel'skie ostrova* 'they were arrested and mostly exiled to Guiana or Seychelles' [= 'most of them were exiled...']

However, this is not the only possible syntactic role for this actant. In (23) it is the direct object *inostrannye knigi* 'foreign books' that fills the valence slot of the whole:

- (23) *Russkie knigi byli sobrany pokojnym mužem knjagini..., inostrannye že — v bol'sinstve vyvezla sama Anna Arkadjevna iz Pariža* lit. 'Russian books were collected by the late husband of the princess..., while foreign books (dir. object) mostly Anna Arkadjevna brought from Paris herself' [= 'most of the foreign books'].

What is essential here is the communicative organization of the clause. The valence slot of the whole should be filled by the Theme. Since the position of the Theme is most often held by the subject, it is clear why it is the subject that for the most part fills the valence slot of the whole.

Conclusion

The data presented above show that the actantial properties of AdvDs are much more diverse than it was believed before. They are not fully inherited from the keyword and cannot be accounted for by the index assigned to the LF symbol (Adv_i) without its complete revision. The information on the AdvD actants should be explicitly entered in the entry of the keyword or in the entry of the AdvD itself.

References

- Апресян 1974 — Апресян Ю. Д. Лексическая семантика (синонимические средства языка). М.: Наука, 1974.
 Апресян 1992 — Апресян Ю. Д. О новом словаре синонимов русского языка // Известия РАН. Сер. литературы и языка. Т. 51. № 1. С. 18—39.
 Апресян 2001 — Апресян Ю. Д. О лексических функциях семейства REAL — FACT // Nie bez znaczenia... Prace ofiarowane Profesorowi Zygmuntowi Saloniemu z okazji jubileuszu 15000 dni pracy naukowej. Białystok, 2001. S. 23—40.
 Boguslavskaya 2003 — Богуславская О. Ю. Словарная статья ПРЕДЛОГ // Новый объяснительный словарь синонимов. Вып. 3. 2003.

- Boguslavsky 1985 — *Богуславский И. М.* Исследования по синтаксической семантике. М.: Наука, 1985.
- Boguslavsky 1996 — *Богуславский И. М.* Сфера действия лексических единиц. М.: Школа «Языки русской культуры», 1996.
- Boguslavsky 2003 — *Богуславский И. М.* Замечания об актантной структуре адвербиальных дериватов. Die het kleine eert, is het grote weerd. Pegasus Oost-Europese Studies 1. Uitgeverij Pegasus. Amsterdam, 2003. S. 23—40.
- Boguslavsky 2005 — *Богуславский И. М.* Валентности кванторных слов. (В печати.)
- Mel'čuk 1974 — *Мельчук И. А.* Опыт теории лингвистических моделей «Смысл $\Leftrightarrow$ Текст». М., 1974.
- Mel'čuk I. et al. 1984a — *Мельчук И. А., Жолковский А. К., Апресян Ю. Д. и др.* Толково-комбинаторный словарь современного русского языка: Опыты семантико-синтаксического описания русской лексики. Wien, 1984.
- Mel'čuk I. et al. 1984b — DEC: Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques I. Les Presses de l'Université de Montréal, Montréal (Québec), 1984.
- Mel'čuk I. et al. 1988 — DEC: Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques II. Les Presses de l'Université de Montréal, Montréal (Québec), 1988.
- Mel'čuk I. et al. 1992 — DEC: Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques III. Les Presses de l'Université de Montréal, Montréal (Québec), 1992.
- Mel'čuk 1995 — Phrasemes in Language and Phraseology in Linguistics // M. Everaert, E.-J. van der Linden, A. Schenk and R. Schreuder (eds). Idioms. Structural and Psychological Perspectives, 1995, Hillsdale, N. J. — Hove: Lawrence Erlbaum Associates, 1995. P. 167—232.

On the «Subject» of Polish Constructions with «Adverbial Participles» in the Framework of Meaning-Text Theory

Krystyna Bojałkowska

Nicolaus Copernicus University
in Toruń
kristinas@wp.pl

Zygmunt Saloni

University of Warmia and Masuria
in Olsztyn
saloni@uw.edu.pl

This paper deals with Polish constructions with forms of verbs called AdvPart, used as subordinates to another verb form. Its central problem is the interpretation of the 1st (deep) actant of AdvPart, which sometimes cannot be found on the surface-syntactic level. On the deep-syntactic level it should be identified with some actant of the other verb.

О субъекте польских конструкций с деепричастиями в рамках теории «Смысл — Текст». В докладе обсуждаются конструкции с деепричастиями. Главная тема — это интерпретация 1-го (глубинного) актанта деепричастия. Он иногда не может быть определен на поверхностно-синтаксическом уровне. На глубинно-синтаксическом уровне его можно отождествить с некоторым актантом другого глагола.

1. Introduction

The aim of this paper is to discuss the Polish constructions with forms of verbs called «adverbial participles» (literal translation of the Polish term «imiesłowy przysłówkowe»), used as subordinates to another verb form («main verb»). There are two types of adverbial participles: (a) contemporary or present («współczesny»), ending with *-ąc* (e.g. *pisząc*, approx. ‘writing’, *siedząc* ‘sitting’) for imperfective verbs and (b) previous or past («uprzedni»), ending with *-szy* (e.g. *napisawszy* ‘having written’, *usiadłszy* ‘having sat’) for perfective verbs¹. «Adverbial participles»

¹ Similar verb forms exist in other Slavonic languages, e.g. in Russian (*деепричастие*) and in Czech (*přechodník*). Such forms of verb exist also in Germanic languages, e.g. in English participles are used in compound forms of the verb and adjectivally or adverbially, cf. Mel’čuk, Pertsov (1987).

always express the relative tense according to the main verb, cf. Wróbel (1975). We will call them in this paper «AdvPart».

The «adverbial participles» are neutralized as regards person, number and gender, therefore cannot combine directly with the nominative subject², cf.

- (1) **Jan śpiąc*. (John sleeping.)

They belong to non-finite verb forms, cannot constitute a complete sentence, and are always dependent on other verb forms (henceforth: MV), as a rule — finite ones, e.g.:

- (2) *Jan odpoczywa, śpiąc*. (John is resting, sleeping.)

We shall not discuss here complicated problems of temporal and semantic relations between two actions (like those in Russian analyzed by Boguslavskij (1977)). We shall limit ourselves to consideration of the relation of actants of the AdvPart and the MV in the terms of Meaning-Text Theory (Mel'čuk (1974, 1988)).

Indeed, AdvPart forms belong to the verbal lexemes, having determined syntactic requirements. The AdvPart form *śpiąc* ('sleeping') in (2) belongs to the lexeme SPAC ('sleep') which has the following government pattern (henceforth: GP):

SPAC
X
NP _{nom}

The verb SPAC has one deep-syntactic actant slot³. In a typical situation (when the verb under consideration is used in finite form), this actant corresponds to a nominative subject in a surface-syntactic structure (= SSyntS). We will call it the 1st actant (even if it is the only actant of a given verb). For the AdvPart the 1st actant slot in SSyntS is systematically blocked. Thus in order to interpret the construction it is necessary to analyze deeper levels of representation. We propose to take into account the deep-syntactic structure (= DSyntS).

² The basis of the description of surface-syntax of Polish is for us Saloni, Świdziński (2001). According to this description we regard as the subject the part of a sentence represented by the noun phrase in the nominative.

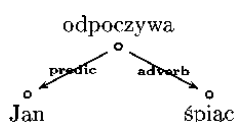
³ The distinction between actants and actant slots was proposed by Mel'čuk (2004), who explored the problem of three types of actants — semantic, deep-syntactic, surface-syntactic and the correspondences between them.

A rule of the interpolation of the 1st actant, missing in the AdvPartP (adverbial participle phrase), is traditionally formulated in Polish grammar: the underlying «subject» of AdvPartP should be coreferential with the nominative subject of the main clause, cf.: «the blocked slot of the nominative subject in the participle form is filled in a default way: it is understood as referentially identical (referred to the same object) with the nominative subject in the higher-level structure» (Saloni 2000, p. 13). Such interpretation, as we will show lower, leads to considerable oversimplification.

Nevertheless, it explains basic facts: the structure of a typical sentence with AdvPartP, and also of somewhat more sophisticated ones: the impossibility of formation of AdvPart from so-called impersonal verbs, i.e. verbs which do not have the nominative actant slot in their surface structure at all (cf. Saloni 2001). Moreover, it is possible to explain by means of this rule the impossibility of introducing AdvPartP into a sentence in which such a verb (e.g. ŚWITAĆ ('dawn'), GRZMIEĆ ('thunder'), KŁUĆ ('stab'), MDLIĆ ('nauseate')) is used as a predicate (MV) and does not introduce any other verbal form.

2. Standard constructions — correspondence of the 1st actants

Let us begin the analysis with most standard constructions, e.g. (2), in which SSyntS appears as follows:

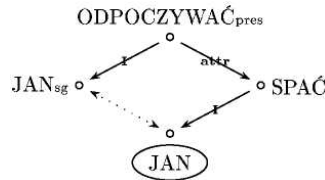


The traditional grammar called this situation the «subjects' identity». Of course, this is a broad and vague use of the term «subject». In the MTT framework the interpretation is very simple. Let us have a look at the GPs of both verbs:

ODPOCZYWAĆ	SPAĆ
X	X
NP _{nom}	NP _{nom}

There is only one possibility of interpretation of the 1st actant of *śpiąc* in the DSyntS ⁴:

⁴ We have adopted the basic notational distinction between the SSyntS and DSyntS (and we write the names of lexemes in DSynt trees in capitals).



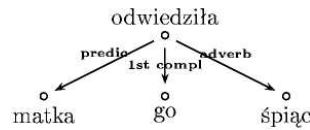
The situation is somewhat more complicated if the MV has more valency slots, e.g.:

- (3) *Śpiąc, odwiedziła go matka.* (Sleeping, the mother visited him.)

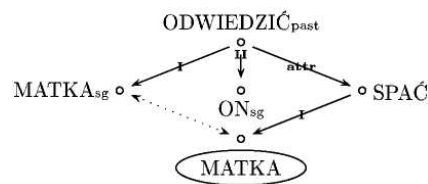
The GPs of the verb ODWIEDZIĆ ('visit') and SPAĆ ('sleep') are the following:

ODWIEDZIĆ		SPAĆ	
X	Y	X	
NP _{nom}	NP _{acc}	NP _{nom}	

The SSyntS, expressed by the dependency tree is unambiguous:



However, the interpretation of the 1st actant of AdvPart is still open. The deep-syntactic interpretation, referring to structural factors (identification of 1st actants)⁵ is the following:

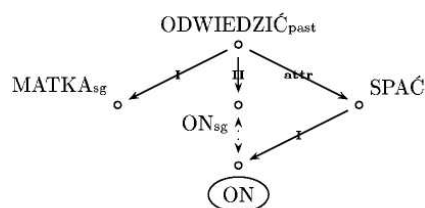


This interpretation seems to be absurd: it is not possible to sleep and to visit somebody simultaneously (although we cannot exclude the possibility of constructing a situational model that makes it reasonable).

From the viewpoint of the presumed communication purpose and the probable semantic interpretation (taking knowledge of non-linguistic re-

⁵ We call «identification of actants» the procedure analogous to unification of variables in logical formulae and programs.

ality into consideration) the DSyntS of the sentence (3) would be the following:



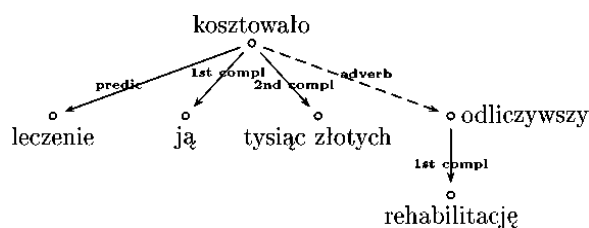
In this tree the rule of identification of the 1st actants has been broken. As a result of the discrepancy between the structural rule of actants' identification and the natural semantic interpretation this sentence is evidently incorrect. It is treated this way both by native speakers and grammar handbooks.

3. Absolute («unrelated») constructions

The sentences with absolute use of AdvPartP (of some groups of verbs for which such constructions are frequent), when the 1st actant refers to somebody generalized, indeterminate, are considered to be correct (by native speakers as well as linguists), e.g.:

- (4) *Leczenie kosztowało ją tysiąc złotych, odliczywszy rehabilitację.*
(The treatment cost her a thousand zlotys, having deducted rehabilitation.)

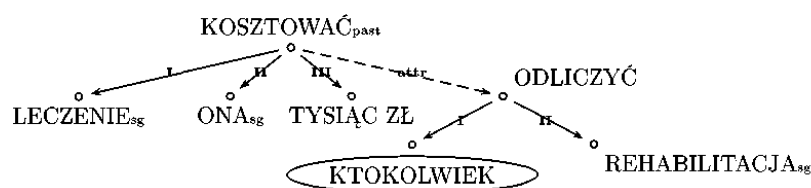
Its SSyntS and the GPs of both verbs are as follows:



KOSZTOWAĆ		
X	Y	Z
NP _{nom}	NP _{acc}	NP _{acc}

ODLICZYĆ	
X	Y
NP _{nom}	NP _{acc}

The identification of the 1st actants in the DSyntS gives absurd results: «treatment» can «cost», but it cannot «deduct» anything. Thus, it is necessary to accept the following interpretation:



In the DSynt tree the AdvPartP is treated as «absolute», it is interpreted from the outside, irrespective of the higher-level structure (*ktokolwiek* ‘anybody’).

It is worth noting that the lack of the 1st actant of MV in the lexical form does not mean its total absence. In Polish in the indicative the finite verb form is its sufficient exponent, as well as in the imperative:

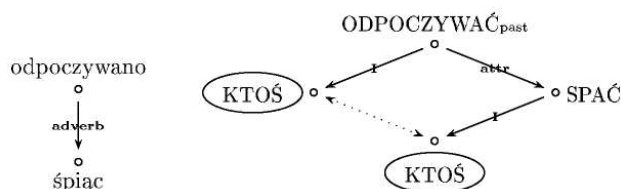
- (5) *Odpoczywaj, śpiąc!* (Rest, sleeping!)

The SSyntS and the DSyntS of (5) are similar to those of (2). The only difference is the syntactic zero, corresponding to the 2nd person.

4. Constructions with impersonal finite verb forms

Furthermore, AdvPartPs can occur also as subordinates of impersonal finite verbal forms which exclude the interpolation of the nominative subject (or vocative). Polish sentences including such forms were interpreted (according to Mel’čuk ‘s proposal) as containing a syntactic zero, expressing a given meaning: *ktos* ‘somebody’ (Saloni, 1976), e.g.:

- (6) *Odpoczywano, śpiąc.* ([They were] resting, sleeping.)



It is possible to use AdvPartP also in sentences whose nominative-subject position is filled by the special element *się*, blocking use of any other subject in the SSyntS, e.g.:

(7) *Odpoczywa się, śpiąc.* ([They are] resting, sleeping.)

Both SSyntS and the DSyntS of (7) are similar to those of (6).

5. Constructions with the infinitive

There is no problem with interpretation of the constructions with the infinitive form used absolutely, in its marginal function — imperative, e.g.:

(8) *Odpoczywać, śpiąc.* (To rest, sleeping.)

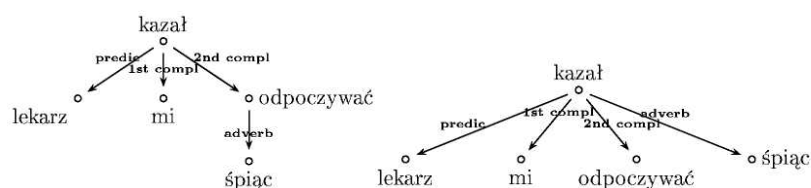
This sentence is interpreted exactly just as the imperative sentence (5). On the other hand, if an infinitive is used typically — as dependent on another verbal form — we have more deep predicates and a higher level of complication, e.g. (9):

(9) *Lekarz kazał mi odpoczywać, śpiąc.*
(The doctor ordered me to rest, sleeping.)

Each of the three verbs has its own GP:

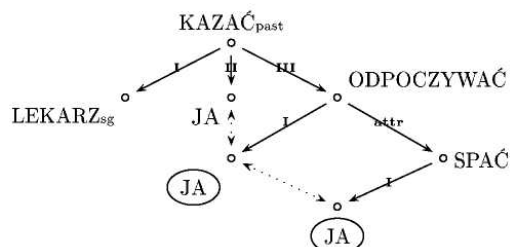
KAZAĆ			ODPOCZYWAĆ		SPAĆ	
X	Y	Z	X		X	
NP _{nom}	NP _{dat}	V _{inf}	NP _{nom}		NP _{nom}	

Even interpretation of the SSyntS causes some difficulties: we can consider AdvPartP as subordinated to the infinitive form *odpoczywać*, as well as to the finite form *kazał*, constituting a sentence⁶:

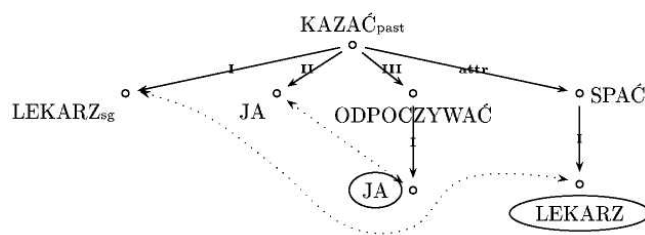


Adequately to two SSynt interpretations we can construct two DSynt interpretations. The first SSynt interpretation corresponds to the following DSyntS, in which the 1st actant of AdvPart is identified with the 1st actant of the infinitive:

⁶ Various possibilities of the surface-syntactic interpretation of the AdvPartP's governor were discussed by Bojałkowska (2005).



This interpretation agrees with the most probable communicative purpose of a speaker but is inconsistent with the natural understanding of a sentence by a recipient, who interprets it spontaneously in the way illustrated by this DSyntS:



This DSynt tree corresponds to the second SSynt tree — the 1st actant of AdvPart is identified with the 1st actant of its governor, the MV. However, such interpretation is unlikely, because it suggests that the doctor was sleeping during consultation.

Some problems of interpretation are caused by so-called impersonal verbs, having no 1st actant slot in their structure at all. However, we omit this point of our consideration because of lack of time.

6. Conclusion

The constructions with AdvPartP are usually easy to analyze on the surface-syntactic level, but the transition into deep-syntactic level enriches the interpretation significantly, because on this level appears the 1st actant of AdvPart, omitted in SSyntS. The key moment is identifying it with some actant of the other verb being a component of sentence structure. This identification is easy within the MTT framework, introducing actants on the deep-syntactic level: 1st actant of AdvPart is identified with a structurally close actant of the governor of AdvPart (which does not have to be the nominative subject in the SSyntS).

S. Kahane (2003) at his paper on the first MTT conference argued: “The deep-syntactic structure is a fundamental representation of a sentence, but it should not be considered as an intermediate structure in the meaning-text correspondence (...) The correspondence from the semantic level to the surface-syntactic level must be direct and the deep-syntactic structure is the witness of this correspondence (i.e. the derivation structure in more technical terms)”.

Our analysis showed that — at least in this case — deep-syntactic structure is very useful as an intermediate level between the semantic and surface-syntactic representations.

References

- Boguslavskij I. (1977). О семантическом описании русских деепричастий: неопределенность или многозначность, *Известия АН СССР. Серия литературы и языка* М., 1977. V. XXXVI. № 3. P. 270—282.
- Bojałkowska K. (2005). Imiesłów przysłówkowy jako składnik konstrukcji zdaniowej, to appear in *Acta Universitas Nicolai Copernici*. 2005.
- Kahane S. (2003). On the status of the deep-syntactic structure, In: *Proceedings from MTT 2003. First International Conference on Meaning-Text Theory*, Paris, 2003. P. 229—238.
- Mel’čuk I. (1974). *Опыт теории лингвистических моделей «Смысл ⇔ Текст»*. Семантика, Синтаксис, Moskva, Nauka, 1974.
- Mel’čuk I. (1988). *Dependency Syntax: Theory and Practice*, State Univ. of New York Press, Albany, 1988.
- Mel’čuk I. (2004). Actants in Semantics and Syntax., *Linguistics* 42—1. 2004. P. 1—66 and 42—2. P. 247—291.
- Mel’čuk I., Pertsov N. (1987). *Surface Syntax of English. A Formal Model within the Meaning-Text Framework*, Amsterdam-Philadelphia, 1987.
- Saloni Z. (1976). *Cechy składniowe polskiego czasownika*, Warszawa, 1976.
- Saloni Z. (2000). *Wstęp do koniugacji polskiej*, Olsztyn, 2000.
- Saloni Z. (2001). *Czasownik polski*, Warszawa, 2000.
- Saloni Z., Świdziński M. (2001). *Składnia współczesnego języka polskiego*, Warszawa, 2001. 5th ed.
- Wróbel H. (1975). *Składnia imiesłowów czynnych we współczesnej polszczyźnie*, Katowice, 1975.

Stable Coordinate Pairs as a Specific Resource of Language *

Igor A. Bolshakov

Center for Computing Research (CIC)
National Polytechnic Institute (IPN), Mexico City, Mexico
igor@cic.ipn.mx

Stable coordinate pairs (SCP) like *trucks and buses* or *safe and sound* are rather frequent in texts, though there are only few thousands of them in language. Statistically, we characterize SCPs by numerical stability index. For their structural description, we propose lexical, morphologic, syntactic, and semantic categories. Some similarities are revealed between SCPs and collocations, i.e. stable word combinations with subordinate dependencies. It is argued that database containing categorized SCPs facilitates several tasks of automatic NLP. The research is based on a set of 4000 Russian coordinate pairs.

Устойчивые сочиненные пары как специфический языковой ресурс. Устойчивые сочиненные пары (УСП) типа *грузовики и автобусы* или *живой и здоровый*, довольно часты в текстах, хотя их в языке всего несколько тысяч. Статистически мы характеризуем УПС числовым показателем устойчивости. Для структурного описания УСП предложены лексические, морфологические, синтаксические и семантические категории. Обнаружено сходство между УСП и словосочетаниями с подчинительными зависимостями. База, составленная из категоризированных УСП, облегчает решение задач автоматической обработки текстов. Исследование основано на наборе из 4000 русских сочиненных пар.

1. Introduction

In all European languages, coordinate constructions are rather frequent in common texts. For example, on average each fifth sentence met on the Russian news site *Gazeta.ru* contains coordination. We name word combination of two content words (or content word compounds) linked

* Work done under partial support of Mexican Government (CONACyT, SNI, and CGPI-IPN).

by a coordinative conjunction Stable Coordinate Pair (SCP), if the occurrence rates of the entire entity and its components satisfy a statistical criterion introduced below. One component in a SCP more or less predict another. In other words, one component restricts the other both lexically and semantically: *coffee and cakes*, *air and space*, *near and far*.

Notwithstanding frequent occurrence of SCPs (e.g., they constitute nearly a third part among coordinate expressions in *Gazeta.ru*), general linguistics gave them scant attention [1, 6]. Within the MTT approach, the early attempt [12] to consider SCP as a lexical function failed, while works [4, 5] seem insufficient.

The objective of this paper is to describe SCPs in some detail. To characterize them statistically, we propose Stable Connection Index similar to Mutual Information Index well known in statistics [7]. To categorize both whole SCPs and their components, we introduce parameters of lexical, morphologic, syntactic, semantic, and pragmatic nature. Then we compare SCPs with collocations, i.e. stable word combinations with subordinate dependencies between components. It is argued that gathering fully characterized SCPs into a database facilitates a variety of NLP applications. The research is based on 4000 Russian coordinate pairs (ca. 3800 after testing).

2. Stability of Coordinate Pairs

SCP as a whole plays the syntactic role of any major part of speech: noun, adjective, verb, or adverb. SCP occurs in a text as a contiguous segment, with its components varying morphologically depending on the context. The surface syntactic structure of SCPs can be of three shapes [8]:

- In the most frequent case $P_1 \rightarrow c \rightarrow P_2$, the components P_1 , P_2 are linked with unique conjunction c that has form u ‘and’, $или$ ‘or’, $да$ ‘and’ or $но$ ‘but’.
- Rarer structures contain disjoint conjunctions $u... u$ ‘both... and’, $или... или$ ‘or... or’, $ни... ни$ ‘neither... nor’, $то... то$ ‘now... now’:

$$\begin{array}{ccc} & \frown & \searrow \\ c_1 \rightarrow P_1 & & c_2 \rightarrow P_2 \end{array}$$

- One can also find a few SCPs without conjunction: $P_1 \rightarrow P_2$ like *без роду без племени* ‘tribeless’.

During the recent 15 years we have gathered a set of ca. 4000 Russian coordinate pairs intuitively considered SCPs. Then the problem arose to formally define and to numerically evaluate stability of the pairs, in order

to clean up the scratch set. We did not take the criterion based only on frequencies of the entire would-be SCPs met in a corpus, since these frequencies depend on the corpus size S while the components P_i taken apart are not considered. A possible solution is to involve Mutual Information known in statistics [7]

$$MI(P_1, P_2) \equiv \log \frac{S \times N(P_1, P_2)}{N(P_1) \times N(P_2)},$$

where $N()$ is frequency of entity in parentheses met through the corpus. Regrettably, only a limited part of our set occurs in the only Russian corpus publicly available [11].

Russian Web search engines like Yandex [13] are incomparably richer, but they deliver statistics on queried words measured in Web-pages. We can re-conceptualize $N()$ as numbers of relevant pages, and S as the page total managed by a search engine. However, now $N()/S$ are not empirical probabilities of occurrences: the same words entering a page are counted only once, while the same page is counted repeatedly for each word included. Thus, MI is not more strictly grounded statistical measure for words. Since MI depends on $N(P_1, P_2)$ and $N(P_1) \times N(P_2)$, we feel free to construe other similar criteria from the same ‘building blocks’, and propose Stable Connection Index

$$SCI(P_1, P_2) \equiv 16 + \log_2 \frac{N(P_1, P_2)}{\sqrt{N(P_1) \times N(P_2)}},$$

where the constant 16 and the logarithmic base 2 are chosen in a way that the majority of the pairs are in the interval 0 to 16. To calculate SCI , we do not need to know the steadily increasing total volume S under the search engine’s control. SCI reaches its maximally possible value 16 when P_1 and P_2 always go together. SCI retains its value when $N(P_1)$, $N(P_2)$, and $N(P_1, P_2)$ change proportionally. This is important since all measured values fluctuate quasi-synchronously in time.

Computing SCI values for the available set by means of Yandex, we have plotted a single-peaked statistical distribution with the mean value $M = 7.90$ and standard deviation $D = 2.92$ (Fig. 1). While dividing the SCP set into three groups, the lower ($SCI < M - D$), the middle ($M - D \leq SCI < M + D$), and the upper one ($SCI \geq M + D$), their relative proportions are 15:70:15. Hereafter a coordinate pair is considered stable if its SCI is nonnegative, though one may take a positive threshold for filtering out pairs with lower SCI values.

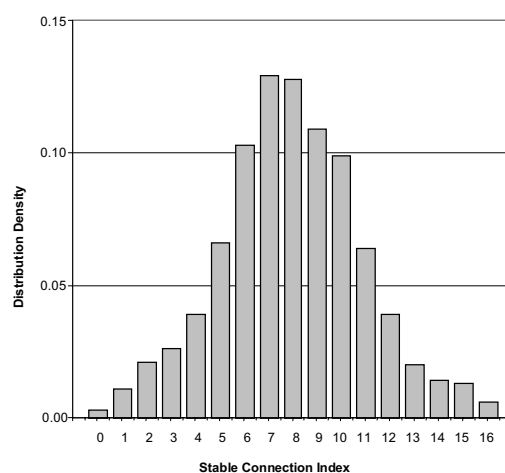
Fig. 1. Distribution of *SCI* values for the SCP set.

Table 1. Several SCPs of the upper group

Russian SCP	<i>SCI</i> (dir/inv)	Translation
<i>ни к селу ни к городу</i>	15.9/4.9	neither here nor there
<i>ни много ни мало</i>	15.8/10.4	neither more nor less
<i>ни рыба ни мясо</i>	15.8/7.0	neither fish nor fowl (of no definite character)
<i>ни днем ни ночью</i>	15.7/11.7	neither day nor night
<i>ни ответа ни привета</i>	15.7/11.3	neither answer nor regard
<i>кнут и пряник</i>	15.7/10.1	carrot and stick
<i>Сан-Томе и Принсипи</i>	15.7/-	San Tome and Principe
<i>денно и ночью</i>	15.6/8.7	day and night
<i>вкривь и вкось</i>	15.4/9.7	indiscriminately
<i>без сучка и задоринки</i>	15.4/-	without a hitch (quite easily)
<i>чрезвычайный и полномочный (посол)</i>	15.4/-	(Ambassador) Extraordinary and Plenipotentiary
<i>ни туда ни сюда</i>	15.1/7.7	neither ping nor pong (no-go)
<i>геодезия и картография</i>	15.0/11.1	geodesy and cartography
<i>Босния и Герцеговина</i>	15.0/-	Bosnia and Herzegovina
<i>ни холодно ни жарко</i>	14.8/14.2	neither cold nor hot
<i>Тринидад и Тобаго</i>	14.7/-	Trinidad and Tobago
<i>холить и лелеять</i>	14.5/10.0	to coddle and protect
<i>приливы и отливы</i>	13.9/9.4	ebbs and flows (ups and downs)
<i>Лейли и Меджнун</i>	13.6/-	Leili and Mejnun
<i>жалобы и предложения</i>	13.5/7.8	complaints and suggestions
<i>дамы и господа</i>	13.2/10.0	ladies and gentlemen

Examples of SCPs with the maximal *SCI* values are given in Table 1. One can see that they are of three types: idioms with disjoint conjunctions (*neither fish nor fowl*); usually inseparable geographic (*San Tome and Principe*) or proper (*Leili and Mejnun*) names; fixed everyday-life expressions, also idiomatic (*carrot and stick*, *ups and downs*). The non-idiomatic pairs (*geodesy and cartography*, *complaints and suggestions*, *etc.*) are rare among the upper SCPs. Except for the names and the fixed formulas like *Extraordinary and Plenipotentiary*, these SCPs can be also used in the inverse order, but with significantly lower *SCI* values (cf. the figures after ‘/’ sign in the middle column).

The most numerous middle group is illustrated by the following SCPs with *SCI* in the interval 7 to 8: *автобусы и автомобили* ‘buses and cars’; *автоматика и робототехника* ‘automation and robotics’; *административный и уголовный* ‘administrative and criminal’; *алгебра и арифметика* ‘algebra and arithmetic’; *американский и британский* ‘American and British’; *анализ и прогноз* ‘analysis and forecasting’; *базы и склады* ‘depots and storehouses’; *безопасность и общественный порядок* ‘security and social order’; *безопасность и оборона* ‘security and defense’; *библиотеки и читальни* ‘libraries and reading rooms’; *бизнес и власть* ‘business and authorities’. Nearly all of them are non-idiomatic with commonly used words as components.

SCPs with the lowest positive *SCI* values can be illustrated as follows: *активизировать и укреплять* ‘to activate and strengthen’; *капитализация и инвестиции* ‘capitalization and investments’; *кризисы и опасности* ‘crises and dangers’; *культура и духовный облик* ‘culture and spiritual background’. These are mostly commonly used non-idiomatic expressions with components occurring apart much more frequently than the components of the pairs in the middle group.

The pairs with negative *SCI* values (ca. 5 %) were removed (after some hesitations) from the initial set, so that the total of actual set is now ca. 3800. Following are examples of pairs that occur in texts but, based on our criterion, do not demonstrate their stability in actual Internet: *выдержка и единство* ‘self-possession and solidarity’; *двурушничество и ренегатство* ‘double-dealing and renegation’; *дислокация и передвижение* ‘dislocation and movement’; *долгая жизнь и благополучие* ‘long life and welfare’; *драпировки и чехлы* ‘drapery and covers’; *инвективы и поношения* ‘invectives and defamation’; *ипотека и рефинансирование* ‘hipothec and refunding’. However, there is always a chance that synonymous paraphrasing of only one component and/or swapping of the components will permit the pair to acquire the SCP status.

3. External Categorization of SCPs

As a whole, SCP can be characterized by the following categories.

Part of speech is determined by its syntactic role of a SCP in a sentence: SCP can be a noun group (NG, 70 % of our set), adjective group (AjG, 18 %), adverb group (AvG, 8 %), or verb group (VG, 4 %). E.g., *coffee and cakes* is NG, *effective and far-reaching* is AjG, *divide and conquer* is VG, *sooner or later* is AvG. Some prepositional groups can play the role of both AjG and AvG, e.g., *in form and in substance* is AjG when modifying *unconstitutionality* (≈‘formal and substantial’) or is AvG when modifying *verify* (≈‘formally and substantially’). We consider these roles as separate SCPs.

Number is relevant only for NGs. Usually it is plural, independently of number of the components P_1 and P_2 . Indeed, both *mom and dad* and *mons and dads* can be substituted by *they*. However, in cases when P_1 and P_2 refer to the same person, the SCP is externally singular (cf. Section 4).

Sphere of usage can be subdivided as follows, without strict borders between the branches:

- Official documentation entries and mass media clichés, including the names of well known organizations: *безопасность и сотрудничество* ‘security and cooperation’; *МВФ и Всемирный Банк* ‘IMF and World Bank’;
- Common business notions including the names of common shops, workshops or store departments: *овощи и фрукты* ‘vegetables and fruits’; *водки и вина* ‘vodka and wines’;
- Everyday life clichés: *холить и лелеять* ‘to coddle and protect,’ *днем и ночью* ‘night and day’;
- Sci-tech terms: *температура и давление* ‘temperature and pressure’; *износ и амортизация* ‘wear and tear’; *нефть и конденсат* ‘oil and condensate’;
- Cultural and religious terms: *Лейли и Меджнун* ‘Leili and Mejnun’; *древние греки и римляне* ‘ancient Greeks and Romans’; *Адам и Ева* ‘Adam and Eva’;
- Official geographic names: *Bosnia and Herzegovina*, *Trinidad and Tobago*.

External semantic correspondences of an entire SCP can be their synonyms in the form of:

- A single word: father and mother = parents;
- A typically Russian comitative expression of the same components: *отец и мать* ‘father and mother’ = *отец с матерью* ‘father with mother’;

- The same SCP given in the reverse order. In the examples *приливы и отливы* (SCI=13.9) = *отливы и приливы* (9.4) ‘ups and downs’ and *жалобы и предложения* (13.5) = *предложения и жалобы* (7.8) ‘complaints and suggestions’ SCI values are comparable, and there is no reasons to prefer any order except of a mere habit. There are also cases when the opposite order changes communicative organization of the expression: *Russia and NATO* (7.4) means approximately ‘Russia and its relations with NATO’, whereas *NATO and Russia* (5.0) means ‘NATO and its relations with Russia’. Such oppositions do not seem fully synonymous.
- A SCP with components synonymous to the corresponding components of the source SCP (one component may be the same). Such SCPs can have comparable SCI values: *раньше и теперь* (5.9) = *раньше и сейчас* (5.4) ‘earlier and today’ (Thus Internet does not prove nowadays the arguments of I. Mel’čuk that only the first option is correct [10].); *ни холодно ни жарко* (14.8) = *ни жарко ни холодно* (14.2) ‘neither cold nor hot’. The options can differ in SCI more significantly: *знакомые и родственники* (9.3) = *знакомые и родня* (4.2) ‘acquaintances and relatives’; *в срок и в полном объеме* (7.9) = *в срок и полностью* (3.1) ‘in time and in full’; *движимое и недвижимое имущество* (13.2) = *движимость и недвижимость* (6.5) ‘movables and immovables’. In such cases it is highly recommendable to replace a SCP by a more stable synonym.

Style is a pragmatic parameter for us: the speaker addresses the given expression to a specific audience. The style can be elevated (very rare: *alpha and omega*), neutral (standard in speech and texts and without any labels in dictionaries), colloquial (used by everybody addressing everyone, very frequent in everyday speech, given in dictionaries with the *colloq* label), and coarse colloquial (commonly used by men addressing men, not so rare in speech but deficiently represented in dictionaries: *пойти и набить морду* ‘to go and to push in one’s mug’).

4. Internal Categorization of SCPs

Internally, SCPs can be characterized as follows.

Inflectionality. A SCP is inflectional if at least one its component changes its morphologic form depending on the syntactic governor of the SCP and/or of its semantically-induced morphologic characteristics (like tense of verb). Noun SCPs that outwardly have both singular and plural forms frequently do not correspond to each other as usual grammatical

numbers. So we consider each number as a separate SCP. Hence declension of a Russian NG is the selection of its case. Russian adjectives change in case, gender, and number (24 combinations, 12 different forms), and these are declension characteristics for AjGs. Note that we compute *SCI* separately for each member of the morphologic paradigm of a SCP, to characterize the whole with the maximal *SCI* value.

Semantic link between components can be of the following types:

- Synonyms, quasi-synonyms, and mere repetitions (11 % in our set): *траур и скорбь* ‘mourning and grief’; *горести и несчастья* ‘misfortunes and troubles’; *лучше и лучше* ‘better and better’.
- Co-hyponyms in an unspecified genus — species hierarchy (58 %): *руки и ноги* ‘hands and feet’; *аксиомы и теоремы* ‘axioms and theorems’; *акушерство и гинекология* ‘obstetrics and gynecology’; *труд и капитал* ‘labor and capital’. The degree of the meaning intersection between quasi-synonyms or co-hyponyms is rather vague.
- Antonyms, quasi-antonyms, conversives, and opposite notions (24 %): *бедные и богатые* ‘the poor and the rich’; *актив и пассив* ‘assets and liabilities’; *Бог и дьявол* ‘God and devil’; *купля и продажа* ‘purchase and sale’; *действие и противодействие* ‘action and counteraction’.
- Co-participants or actions in a situation (7 %): *писатель и читатели* ‘writer and readers’; *закон и порядок* ‘law and order’; *кожа и косметика* ‘skin and cosmetics’.

The latter type is the most complicated semantically. Some subtypes of the situation are as follows:

- In *пойти и не вернуться* ‘to go and not to come back’, there is a time sequence of actions, with the time of P_1 preceding that of P_2 .
- In *война и разруха* ‘war and ruins’, *преступление и наказание* ‘crime and punishment’, *штормы и наводнения* ‘storms and flooding’, there is a material cause-consequence link: a war brings about ruins, crime leads to official punishment, a storm entails flooding, etc.

Idiomatcity. A SCP is idiom if its meaning is not just a sum of its components’ meanings. Idioms whose meaning does not contain meanings of any of their component are complete phrasemes [8], e.g., *ни рыба ни мясо* ‘of no definite character’ (lit. ‘neither fish nor meat’); *ни Богу свечка, ни черту кочерга* ‘absolutely useless’ (lit. ‘neither a candle for the God nor a poker for the devil’). The majority of SCPs are not idiomatic.

Irreversibility of SCP components could be induced by a temporal or causative sequence mentioned above. However many SCPs are reversible, though their *SCI* changes (cf. Sections 2 and 3).

Lexical peculiarity means that at least one component is not used separately. For example, in *и там и сям* ‘here and there’ word *сям* is peculiar (compare the word *fro* in *to and fro*).

Coreferentiality. In very rare cases, P_1 and P_2 co-refer to the same person: *вор и разбойник* ‘thief and robber’, *муж и повелитель* ‘husband and sovereign’. This parameter determines morpho-syntactic agreement in number: co-referent NGs are considered singular.

5. Stable Coordinate Pairs vs. Collocations

We name collocations any stable word combinations of content words connected, directly or through an auxiliary word, by means of subordinate dependency links. Stability of collocations could be measured just as for SCPs—by *SCI* values. In contrast to SCPs, collocations are very numerous in any language. The above definition, even if rather vague, covers the vast specter of word combinations studied by I. Mel’čuk [9], complete phrasemes and semiphrasemes included. However, we also consider collocations free word combinations stable in our sense.

Comparison of SCPs and collocations is possible in the following aspects.

Idiomacity. Just as for collocations, rather few SCPs are complete phrasemes. As to semiphrasemes (containing the main meaning of one component and a lateral meaning of the other), we know only one Russian SCP with this property — *дешево и сердито* ‘cheaply and impressively’ (lit. ‘cheaply and angrily’).

Lexical functions are characteristic for collocations, their components being an argument and the corresponding function value of a co-occurrence type LF. Meantime, components of some SCPs are an argument and the value of a substitution type LF (mostly **Syn** or **Anti**), but the most often semantic link between SCP components — co-hyponymy — has no match among LFs of any type.

Substitutability is possibility to substitute a component of a word combination by its synonyms. As it was shown in Section 3, this property is not characteristic for SCPs, just as for collocations including LFs. However, free word combinations with subordinate dependency links between components can use synonymous paraphrasing at full fledge.

It should be mentioned here that the free word combinability usually means independent transformation of meaning of each component of a combination to the corresponding lexeme. In no way it means full freedom of combining different meanings in speech; it is rather limited for

both a SCP and a subordinate word combination. So we can imply that language imposes rather tight restrictions on both the combinability of SCP components' meanings and lexical representation of these meanings.

Modifiability is possibility to expand a word combination by dependent(s) on any component of the pair without changing its status. For SCPs modifiability is very limited. Our experiments have shown that dependents intruding between the utmost left and the utmost right words of a SCP usually deprive the pair of stability. E.g., *аксиомы и теоремы* 'axioms and theorems' is SCP, whereas *аксиомы и отдельные теоремы* 'axioms and isolated theorems' is not. The cases when a more developed expression remains SCP are extremely rare: *актер и зрители* 'actor and spectators' (6.1) *Vs.* *актер и благодарные зрители* 'actor and grateful spectators' (5.3). In contrast to SCPs, many collocations can be developed with respect to both components: additional words are not taken into account for evaluation of stability.

We can conclude that SCPs have some similarity with collocations, though theoretically SCPs are a quite different resource of language.

6. Use of SCPs in Natural Language Processing

If we supply each SCP of the available set with all parameters introduced above, including the corresponding syntactic subtree, semantic interpretation (for idioms), and *SCI* value, the set of SCPs becomes a database very useful for various applications. Let us give their synopsis.

Referencing in text editing is needed while preparing a new text or editing an already existing one. Indeed, even a native speaker can feel uneasiness while selecting a convenient expression for a given case, SCPs being among such expressions. The appropriate SCP could be found through any its component or a one-word synonym of the whole SCP (if any).

Learning foreign language is greatly facilitated with the SCP database. A student should know that *знакомые и родственники* is more preferable than *знакомые и родня* (both are 'acquaintances and relatives'), and *в срок и в полном объеме* is much more preferable than *в срок и полностью* (both are 'in time and in full').

Word sense disambiguation. Out of context, a component of a SCP can have different meanings. In our set about 20 % of SCPs contain at least one ambiguous word. Nearly all SCPs resolve this ambiguity, selecting only one sense. E.g., in *война и мир* 'war and peace', the noun *мир* has two senses: 'world' and 'peace', and the SCP selects the second one; in *наука и техника* 'science and technology', the noun *техника*

(‘technology’ or ‘technique’) resolves to the first sense. We suppose that all SCP components in the database are labeled by their sense numbers.

Parsing. Since the SCP DB contains their partial parses, the parsing of the embedding sentence is facilitated: the parser finds the sequence of words corresponding to the SCP and copies its dependency subtree to the dependency tree of the sentence under parsing. For Russian this operation includes lemmatization. E.g., the textual expression *овощей и фруктов* ‘of vegetables and fruits’ should be reduced to the standard dictionary form labeled with genitive case: $\{\text{овощи и фрукты}\}_{\text{GEN}}$. Sometimes morphologic or even lexical ambiguity is then resolved, e.g., *багаж и ручную кладь* ‘heavy and hand luggage’ is unified to accusative case taking into account the only morphologically unambiguous wordform *ручную* (the forms *багаж* and *кладь* permit also nominative case interpretation).

Detecting and correcting malapropisms. Malapropisms are semantic errors replacing one content word by another, similar in sound but different in meaning. Syntactic links of malapropos word with its contextual words often remain the same. In [3] semantic text anomalies are detected by noting that malapropisms, as a rule, destroy the collocation(s) that the mutilated word would be in. We can apply the same idea to SCPs. Suppose that the program of malapropism detection and correction finds in a text the syntactically correct coordinate pair *волевой и решательный* ‘strong-willed and computing’ with ultimately negative *SCI* value. By few editing operations on both components, a special subprogram finds the unique similar SCP *волевой и решительный* ‘strong-willed and decisive’, thus indicating both the error and its possible correction.

Linguistic steganography is automatic concealment of digital information in rather long orthographically and semantically correct texts. In [2] a steganographic algorithm replaces words by their synonyms, taking into account the context. However, only few SCPs do have synonyms, while the rest permits synonymous paraphrases of neither the pair nor its components. This knowledge is quite important for steganography.

7. Conclusion

A convenient numerical measure of stability — Stable Connection Index — is proposed for coordinate pairs and on this ground the notion of a stable coordinate pair is introduced. Various lexical, morphologic, syntactic, semantic, and pragmatic features are proposed, for both entire SCPs and their components. So far, as many as 3800 Russian SCPs passed the test on non-negativity of *SCI*. Supplied with all categorial in-

formation proposed, the set of SCPs formed a useful database. Theoretically, this DB can be considered as a small part of Explanatory Combinatorial Dictionary. Taken apart, such DB facilitates several modern applications of NLP.

References

- Bloomfield, L. *Language*. Halt, Rinehart and Winston, 1964.
- Bolshakov, I. A. A Method of Linguistic Steganography Based on Collocation-Verified Synonymy. In: J. Fridrich (Ed.) *Information Hiding*. Proc. 6th Int. Workshop IH 2004, Toronto, Canada, May 2004. Revised Selected Papers. LNCS 3200, Springer, 2004. P. 180—191.
- Bolshakov, I. A. An Experiment in Detection and Correction of Malapropisms through the Web. In: A. Gelbukh (Ed.). *Computational Linguistics and Intelligent Text Processing*. Proc. 6th Intern. Conf. on Computational Linguistics CICLing-2005, Mexico City, Mexico, February 2005. LNCS 3406, Springer, 2005. P. 803—825.
- Bolshakov, I. A., Gaysinski A. N. Slovar' ustojčivyx sočinennyx par v ruskom jazyke. *Nauchnaya i Tekhnicheskaya Informatsiya*. Ser. 2. №. 4. 1993. P. 28—33.
- Bolshakov, I. A., Gelbukh A., Galicia-Haro S. N. Stable Coordinated Pairs in Text Processing. In: V. Matoušek, P. Mautner (Eds.) *Text, Speech and Dialogue*. Proc. 6th International Conference TSD 2003, České Budějovice, Czech Republic, September 2003. LNAI 2807, Springer, 2003. P. 27—34.
- Malkiel, Y. Studies in Irreversible Binomials. *Lingua*. V. 8. 1959. P. 113—160.
- Manning, Ch. D., H. Schütze. *Foundations of Statistical Natural Language Processing*. MIT Press, 1999.
- Mel'čuk, I. A. *An experience on the theory of linguistic models Meaning — Text. Semantics, Syntax* (in Russian). Moscow: Nauka Publ., 1974.
- Mel'čuk, I. Phrasemes in Language and Phraseology in Linguistics. In: M. Everaert et al. (Eds.) *Structural and Psychological Perspectives*. Hillsdale, NJ / Hove, UK: Lawrence Erlbaum Associates Publ. P. 169—252.
- Mel'čuk, I. "Sejčas i teper'" v sovremennom ruskom jazyke. In: I. A. Mel'čuk. *The Russian language in the Meaning—Text perspective*, Wiener Slawistischer Almanach. Sonderband 39. P. 55—76, Moskau; Wien: Jazyki Russkoj Kultury Publ., 1995.
- National Corpus of Russian Language*, <http://www.ruscorpora.ru>.
- Šal'japina, Z. M. O sposobax vyraženija semantičeskix parametrov v anglijskom jazyke. In: *Mašinnyj perevod i prikladnaja lingvistika*. No. 11. Moscow: MGPI Publ., 1969. P. 106—121.
- Yandex*, <http://www.yandex.ru>.

Semantic Roles and Diatheses for Functional Discourse Plans

Alexander Dikovsky, Boris Smilga

LINA — FRE CNRS 2729,
Université de Nantes
alexandre.dikovsky@univ-nantes.fr, boris.smilga@univ-nantes.fr

In this paper, we elaborate the notion of semantic diathesis used in the formal language of discourse plans proposed and described in [Dik03a, Dik03b]. The modern understanding of diathesis builds on the pioneer work of Mel'čuk and Xolodovič [MX70] who formalized it as mapping between the set of semantic and the set of deep-syntactic actants of a given lexeme. Recently, Padučeva [Pad97, Pad02] has linked diathesis to communicative structure. We mainly follow her account, however, with the following important differences: a) instead of unstructured taxonomic characteristics, we use more elaborate semantic typing with inheritance; b) rather than relying on lexical definitions, we establish basic semantic roles based on a small set of cognitively grounded protosituations; c) a different system of communicative ranks is used; d) which is the most important, in our approach, semantic diatheses are a means of semantic and discourse planning rather than a means of lexical derivation.

Семантические роли и диатезы для функциональных дискурсивных планов. В настоящей статье мы уточняем понятие семантической диатезы, как оно используется в формальном языке дискурсивных планов, предлагаемом в работах [Dik03a, Dik03b]. Современное понимание диатезы восходит к новаторской работе Мельчука и Холодовича [MX70], которые формально определили диатезу как соответствие между семантическими и глубинно-синтаксическими актантами данной лексемы. В недавних работах Падучевой [Pad97, Pad02] вводится понятие семантической диатезы, определяемое в терминах коммуникативных рангов. Мы применяем весьма близкий подход, но со следующими принципиальными отличиями: а) слабо структурированные таксономические характеристики заменяются более гибкой и точной системой семантических типов с отношением наследования; б) основные семантические роли определяются не на основе лексикографических толкований, а исходя из небольшого набора когнитивно обусловленных протоситуаций; в) применяется иная система коммуникативных рангов; г) что важнее всего, в нашем подходе семантические диатезы являются средством семантического и дискурсивного планирования, а не средством лексической деривации.

1. Semantic types and compositionality

DISCOURSE PLANS described in [Dik03a, Dik03b] represent the course of event conceptualization by a speaker. They are *functional* in the following sense. Plans are *terms*, i. e. tree-like functional compositional structures. *All* elements (*semantemes*) of these structures, with the only exception of special empty primitives (zero values, empty lists) have arguments and are functions. Ordered naturally top to bottom, left to right, the nodes of a plan form the plan's *sequence of points*.

Semantemes have functional types, defined from primitive types, of which there are four *initial basic types*: **n** (nominators, i. e. 'things' in the most general sense), **s** (sententiators, i. e. 'actions', 'processes', 'events', 'facts', etc.), **q** (qualifiers, 'meanings qualifying nominators'), and **c** (circumscriptors, 'meanings qualifying sententiators and qualifiers'). Basic types are extended by their specific instances. Thus the complete set of basic types is partially ordered under the *specific / generic relation* $\prec$. For instance, in § 3 we will see the following types: $\mathbf{n}_a \prec \mathbf{n}$ (animated nominator, e. g. $\langle\langle * \text{hearer}^{*n_a} \rangle\rangle$, $\langle\langle \text{serpent}^{n_a} \rangle\rangle$), $\mathbf{n}_{\text{mass}} \prec \mathbf{n}$ (mass nominator, e. g. $\langle\langle \text{TV}^{n_{\text{mass}}} \rangle\rangle$, $\langle\langle \text{milk}^{n_{\text{mass}}} \rangle\rangle$), $\mathbf{q}_{\text{qnt-mass}} \prec \mathbf{q}_{\text{qnt}} \prec \mathbf{q}$ (quantifier of mass nominators, e. g. $\langle\langle \text{much}^{\mathbf{q}_{\text{qnt-mass}}} \rangle\rangle$), $\mathbf{s}_{\text{attr}} \prec \mathbf{s}$, $\mathbf{s}_{\text{percep}} \prec \mathbf{s}_{\text{eff}} \prec \mathbf{s}$ (situations of attribution, direct perception and effect, e. g. $\langle\langle \text{be}_{\text{attr}}^{\mathbf{s}_{\text{attr}}} \rangle\rangle$, $\langle\langle \text{kiss}^{\mathbf{s}_{\text{eff}}} \rangle\rangle$, $\langle\langle \text{watch}^{\mathbf{s}_{\text{percep}}} \rangle\rangle$). These basic types are similar to the conventional classes of semantic compatibility (cf. *sočetaemost'* of [Apr74], *T-category* of [Pad97, Pad98], or *semantic class* of [Mel04a]). The difference is that the basic types of discourse plans obey strict compositionality rules (see below).

Every basic type **u** has a corresponding *optional* version $\mathbf{u}^{(0)}$ signifying optional arguments of this type. In addition, circumscriptor and qualifier types have *iterative* versions $\mathbf{u}^{(\omega)}$ (zero or more objects of type **u**). Together, basic, optional and iterative types constitute the set of *primitive* types. Finally, the primitive types serve to form complex *functional types* ($\mathbf{u}_1 \dots \mathbf{u}_k \rightarrow \mathbf{v}$) ($\mathbf{u}_1 \dots \mathbf{u}_k$ and $\mathbf{v}$ being respectively the argument and value types). For instance, the situation $\langle\langle \text{pour out} \rangle\rangle$ has the type $(\mathbf{c}^{(\omega)} \mathbf{n}_a \mathbf{n}_{\text{liqm}}^{(0)} \mathbf{n}_{\text{ctr}} \rightarrow \mathbf{s})$, where $\mathbf{n}_{\text{liqm}}^{(0)} \prec \mathbf{n}_{\text{mass}} \prec \mathbf{n}$ is the optional liquid matter nominator and $\mathbf{n}_{\text{ctr}} \prec \mathbf{n}$ is the container nominator type.

Compositionality is restricted by the condition that a subplan with the value of type $\mathbf{t}_1$ may be substituted in another plan of type $(\mathbf{u}_1 \bar{\mathbf{u}} \rightarrow \mathbf{v})$ in the place of the argument of type $\mathbf{u}_1$, to obtain a composite plan of type $(\bar{\mathbf{u}} \rightarrow \mathbf{v})$, only if $\mathbf{u}_1 = \mathbf{t}_2$ or $\mathbf{u}_1 = \mathbf{t}_2^{(0)}$, and either $\mathbf{t}_1 = \mathbf{t}_2$ or $\mathbf{t}_1$ is a case of $\mathbf{t}_2$ (denoted $\mathbf{t}_1 \preceq \mathbf{t}_2$). In classical lexicographical terms, this means that the class of semantic compatibility of $\mathbf{u}_1$ is a superclass of that of $\mathbf{t}_1$, or that $\mathbf{u}_1$ is optional $\mathbf{t}_1$.

2. Roles of situation actants

We distinguish traditionally two basic classes of semantemes: *situations* and *monads*. A situation is a function, with one or more arguments, called *actants* and identified by their (*semantic*) *roles* (different arguments of the same situation have different roles). Arguments without roles are called *circumstantials* (cf. § 2.3), and semantemes without actants are called monads. Most verbs express situations: *John*_{SBJ} *gave* *the letter*_{OBJ} *to Fred*_{RCP}; Fr. *Il*_{SBJ} *me*_{OBJ} *dérange* *par son chant*_{INS} /He disturbs me by his singing/. By contrast, most names, adjectives and adverbs do not express situations, so that their meanings are represented by monads. The most generic type $\mathbf{t}_C$ of a circumstantial C of a situation S is uniquely determined by the value type $\mathbf{t}_S$ of S . Nominator-value semantemes ($\mathbf{t}_S = \mathbf{n}$) have qualifier-type circumstantials (i. e., $\mathbf{t}_C = \mathbf{q}$). For all other types of semantemes, circumstantials are circumscriptors ($\mathbf{t}_C = \mathbf{c}$). Grouping the circumstantials of a semanteme S into a list, we obtain its generic circumstantial argument of the corresponding iterative type: $\mathbf{q}^{(\omega)}$ for $\mathbf{n}$ -semantemes and $\mathbf{c}^{(\omega)}$ for semantemes of other value types. For instance, the generic type of verbal situations is $(\mathbf{c}^{(\omega)} \bar{\mathbf{u}} \rightarrow \mathbf{s})$, where $\bar{\mathbf{u}}$ represents the sequence of actant types of a situation.

A situation is specified by its *profile*, which includes the situation functor, the argument types, the situation value type and the role of each actant. Thus, the additional feature provided by our account is that actants may be decorated by optional types $\mathbf{u}^{(0)}$ and in this way obtain the status of optional actants. E. g., the situations $\langle\langle \text{give} \rangle\rangle$ and $\langle\langle \text{déranger} \rangle\rangle$ have the profiles: $\langle\langle \text{give}(\text{SBJ}^{\mathbf{n}_a}, \text{OBJ}^{\mathbf{n}}, \text{RCP}^{\mathbf{n}_a})^{\mathbf{s}_{\text{caus-mov}}} \rangle\rangle$ and $\langle\langle \text{déranger}(\text{SBJ}^{\mathbf{n}}, \text{OBJ}^{\mathbf{n}_a}, \text{INS}^{\mathbf{n}^{(0)}})^{\mathbf{s}_{\text{eff}}} \rangle\rangle$. Given such a profile, one can easily reconstruct the complete functional type of the situation. For instance, the complete type of $\langle\langle \text{give} \rangle\rangle$ is $(\mathbf{c}^{(\omega)} \mathbf{n}_a \mathbf{n} \mathbf{n}_a \rightarrow \mathbf{s}_{\text{caus-mov}})$ ($\mathbf{s}_{\text{caus-mov}} \prec \mathbf{s}_{\text{eff}}$ is the type of ‘caused movement’ situations; $\mathbf{c}^{(\omega)}$ stands for a possibly empty list of circumscriptors: circumstantial arguments of $\langle\langle \text{give} \rangle\rangle$).

A formal definition of discourse plans does not require any preliminary definition of roles, types and situation actants, which can be designated as DP signature elements. However, to develop these structures in practice, one must dispose of *linguistic criteria* for assigning particular roles and types to situation arguments and ultimately to decide, for a given situation, what its arguments are, and which among them are actants. Since the 1970s, these criteria are in the center of lively debates (see [Mel04a, Mel04b] for details). Essentially, there are two different approaches, both originating from the pioneer works of Fillmore [Fil68,

Fil77] on Case Grammar, and of Gruber [Gru65]. The one, *functional syntactic* approach tries to deduce actants and their roles from the universals of case/adposition marking of surface-syntactic verb arguments. The other, *lexicographic* approach (cf. [Apr74, Mel04a]) derives the actants and roles from syntactic government patterns and from lexicographic definitions. The main drawback of the first approach is its inherent incompleteness. The second suffers from the absence of a minimal common base of lexicographic definitions (whence the overwhelming abundance and ambiguity of roles proposed by different authors). Below, we outline a different way to approach this problem: one founded on a cognitive hypothesis consistent with both of these approaches.

We suppose that semantic roles originate from the profiles of several cognitively grounded protosituations (see § 2.2). In addition, roles may be assigned as a result of applying diathetic shifts (§ 3). Meanwhile, to define diatheses and diathetic shifts, we need some more notions.

2.1. Communicative ranks

Two main instruments of event conceptualization are (1) choosing semantemes from the dictionary to reference an event and its participants and (2) assigning the speaker's point of view on the relative salience of participants (COMMUNICATIVE STRUCTURE). Concerning communicative structure, we mainly follow the account of [Pad97, Pad02], but with significant alterations due to the data and discussion in [Lam94]. The minimal communicative structure in a plan, which we dub *communicative group*, coincides with a situation, and consists of its functor and *all* arguments (not exclusively actants). Each element of a communicative group is assigned a *communicative rank*. We distinguish two THEMATIC ranks: *topic continuation* and *implied* and three RHEMATIC ranks: *focus*, *background* and *periphery*.

Topic continuation (denoted $\bar{T}$). Assigned to the member of the communicative group to which the new (rhematic) information will be relativized. It corresponds to a referent initialized or previously evoked in the discourse context, which becomes, in the current plan point, the main entity under consideration. E. g.: *Those girls, they $\bar{T}$ giggle when they see me.*

Implied (O). Assigned to a member of a communicative group if its referent is embedded and extremely salient due to deixis, anaphora, etc. (often it is elided in the surface form). E. g.: *Remember Mark?* (= *Do you_O remember Mark?*).

Focus ($\odot$). Assigned to the member of a group which conveys the new (as opposed to given or presupposed) information (cf. [Lam94, § 2.3 and esp. § 5.1]). See example below.

Background ($\oplus$). Assigned to those members of a group which convey other pertinent information (and so cannot be dismissed). E. g.: *I_T touched the stick_{\oplus} against the fence_{\odot}.*

Periphery ($\ominus$). Assigned to the members of a group which become non-salient and should be dismissed. E. g. in the answer to *Do you sell your car?* – *I'm sorry, it's already been bought*, the figure of the buyer obtains the rank $\ominus$.

Rank assignments obey certain constraints:

1. At least one group member should obtain a rhematic rank (= there must be some new information in the utterance);
2. there should be exactly one focused member¹;
3. the semanteme functor may obtain only $\odot$, $\oplus$ or $\ominus$;
4. no more than one member may have the rank $\overline{T}$ (although there may be multiple $\ominus$ s).

2.2. Proto-situations and proto-roles

We believe that there is a small base of PROTO-SITUATIONS stemming from cognitive origins. Every concrete situation obtains its profile through projection onto one proto-situation (or more, in complex cases). Through this projection, the situation inherits the cognitive construal and thus the roles of its proto-situation (this inheritance is denoted the same way as for semantic types, $\prec$). It is not our goal in this short paper to elaborate such a base, or to found its completeness. We just illustrate the principle by a representative sample of proto-situations whose cognitive nature is rather evident.

All proto-situations in the sample have a selected actant which is the cognitive FIGURE. Its semantic proto-role is SBJ. The situation functor and the other actants constitute the GROUND.

Focalized predication, with the *ground* consisting only of the functor: $\langle\langle\text{functor (SBJ}^n)^{\text{sfpd}}\rangle\rangle$. E. g. *She_{SBJ} cheered up*.

Attribution is one in which the only *ground* actant is some attribute of SBJ assigned by the situation. The role of this actant is DFS (*de-*

¹ We don't discuss here the case of presentational constructions classified in [Lam94, § 5.2.4] as communicative groups with neutralized focus

finiens), and may have any type of nominator attribute: $\langle\langle \text{functor}(\text{SBJ}^n, \text{DFS}^{\text{nu}})^{\text{satt}} \rangle\rangle$ ($\mathbf{n}_u$: a nominator attribute type). E. g. *She*_{SBJ} *is busy*_{DFS}; Rus. *On*_{SBJ} *stal vračom*_{DFS} /He.NOM became doctor.INSTR/.

Movement, representing any evolution. It has three optional *ground* actants whose roles are ORIG (origin, initial point), PTH (path), and DST (destination, final point): $\langle\langle \text{functor}(\text{SBJ}^n, \text{ORIG}^{\text{n'0}}, \text{PTH}^{\text{n'0}}, \text{DST}^{\text{n'0}})^{\text{smov}} \rangle\rangle$ (at least one of these optional actants must be present). E. g. Rus. *Lodka*_{SBJ} *otčalila ot pristani*_{ORIG} /Boat.NOM unberthed from pier.GEN/; *This fabric*_{SBJ} *clings to the body*_{DST}.

Effect, representing a dynamic cognitive state, in which an active *force* is applied to a passive *undergoer*, possibly with the help of a *mediator (instrument)*. This proto-situation has two versions:

Nominative: $\langle\langle \text{functor}(\text{SBJ}^n, \text{INS}^{\text{n'0}}, \text{OBJ}^{\text{n'0}})^{\text{snef}} \rangle\rangle$, in which the *force* is represented by the SBJ role actant and the *undergoer* is represented by the OBJ actant. E. g. *Wilhelm Tell*_{SBJ} *shot* *Gessner*_{OBJ} *with an arrow*_{INS}; *The stag*_{SBJ} *sprouted horns*_{OBJ}.

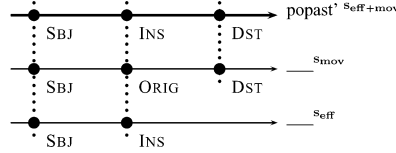
Ergative: $\langle\langle \text{functor}(\text{AGT}^{\text{na}}, \text{INS}^{\text{n'0}}, \text{SBJ}^n)^{\text{seef}} \rangle\rangle$, in which the *force* is represented by the actant with the role AGT of the animated nominator type and the *undergoer* is represented by the SBJ actant. The examples of this kind are at best rare in Russian and English, but abundant in ergative languages, e. g. Alutor *Qutkinjnjaqu-nak*_{AGT} *γ-akmi.l-lin uttā-ʔut*_{SBJ} /Qutkinnjaqu-ERG+SG RES-take-RES+3SG.P club-NOM+SG = Qutkinjnjaqu took a club/ [KKM00, p. 253].

Some simple projections are ‘metaphorical extensions’ of proto-situations. E. g. situations of direct perception are (in many languages) a species of effect; the normal situation of possession in English (*X*_{SBJ} *has* *Y*_{OBJ}) is a species of effect, while in Russian (*u X-a*_{LOC} *estʹ Y*_{SBJ}, /at *X*.GEN is *Y*.NOM/) it is a species of stative location (see below); the English benefactive construction (*give X*_{OBJ} *to Y*_{DST}, etc.) is a species of (caused) movement, while the Japanese one is arguably a species of (caused) stative location (*X*_{LOC} *ni Y*_{OBJ} *o kudasaru*, /*X* LOC/DAT *Y* ACC give.HON/). The above basis of proto-situations is probably incomplete. For two extremely important classes of situations the question remains open whether they are basic proto-situations or just their derivative abstract situations:

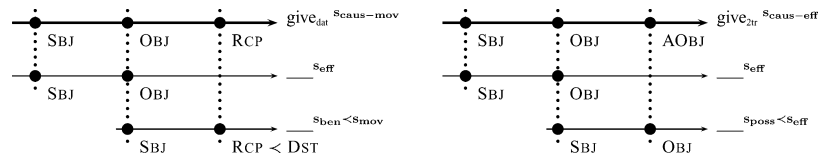
Modal / phasal situations ($\mathbf{s}_{\text{mph}}$, a tentative species of $\mathbf{s}_{\text{nef}}$): $\langle\langle \text{functor}(\text{SBJ}^n, \text{OBV}^{\text{ns}})^{\text{smph}} \rangle\rangle$. $\text{OBV} < \text{OBJ}$ is the *goal*, or *objective* of the situation, and $\mathbf{n}_s$ is a ‘deverbal noun’ type, specialized in English by $\mathbf{n}_{\text{inf}}$ (infinitive). E. g. *I*_{SBJ} *must* {*help you*}_{OBV}; *She*_{SBJ} *began to wonder*_{OBV}; *They*_{SBJ} *managed to* {*get through*}_{OBV}.

Stative location (s_{loc} , a tentative species of s_{attr}): $\langle\langle functor(SBJ^n, LOC^n)^{s_{loc}} \rangle\rangle$. $LOC < DFS$ is the *locus* of the situation and has the type n_1 (nominator of place): *It_{SBJ} happened in Paris_{LOC} / yesterday_{TMP}* $<_{LOC}$; Jap. *{Takusan no o-tera}_{SBJ} ga Nikkō_{LOC} ni arimasu.* /Many GEN temple NOM Nikko LOC is = Nikko has many temples/.

Many concrete situations are projected onto more than one proto-situation. A rather characteristic example of this kind of situations is the Russian verb *popast'* /hit/, which superposes movement onto effect: *Wil'gel'm Tell' popal streloj iz arbaleta priāmo v iābloko* /Wilhelm Tell.NOM hit arrow.INSTR out-of crossbow.GEN directly into apple.ACC²:



The above projection superposes proto-situations while mapping SBJ onto SBJ. This, however, should not always be the case. There is an extremely broad class of *causative* situations, where the undergoer of an effect proto-situation is mapped onto the subject of another proto-situation: modal / phasal (e. g. Rus. *On zastavil meniā uiti* /He.NOM forced I.ACC go.INF/, cf. *Īa dolžen uiti* /I.ACC must go.INF/), attribution (Pol. *IBM mi-anował go dyrektorem* /IBM.NOM appointed he.ACC director.INSTR/, cf. *On je dyrektorem* /He.NOM is director.INSTR/), stative location (*I put the book on the shelf*, cf. *The book is on the shelf*). As to caused movement and caused effect, they are best exemplified by the case of English dative alternation: *Eve gave the apple to Adam* vs. *Eve gave Adam the apple*. The difference between these two sentences is somewhat subtle and can be loosely described as follows (cf. [But04, p. 100] and references cited therein): in the former sentence, Eve causes the apple to change hands and *pass to* Adam, while in the latter she causes Adam to *have* the apple. Therefore, the first case is that of *caused movement*, and the second case is that of *caused effect*:



² The emergence, in Finnish, of a similar, albeit much more systematic pattern of construal superposition is being observed in the article [KL04].

2.3. Circumstantials vs. actants

Whereas the profile of a given situation S entirely defines the logical meaning of every actant X of S through X 's role, the only semantic information in the profile regarding circumstantials is their generic type. It is only in the course of planning that circumstantials receive their logical meaning. The formal markers which identify these meanings and which plans assign to circumstantials are called *attributes*. In addition to this, another fundamental difference is that the logical semantics $sem [X]$ of an actant X is an argument of the function $sem [S]$ representing the logical semantics of the situation S . On the contrary, the logical semantics $sem [C]$ of a circumstantial C of S is a *higher-order function applied to* $sem [S]$. Due to lack of space, we limit ourselves to the following short example. Discourse plans express the noun phrase *your intellect* with the monad $\langle\langle intellect^{n \rightarrow mass} \rangle\rangle$ whose *appurtenance* attribute $Appurt^{(n \rightarrow q)}$ has the value $\langle\langle *hearer^{*na} \rangle\rangle$. The logical meaning of this attribute is the result of applying the meaning of the preposition $\langle\langle of \rangle\rangle$ to $sem [intellect]$ and $sem [*hearer^*]$: $sem [intellect (Appurt^{(n \rightarrow q)} (*hearer^*))] = sem [of] (sem [intellect], sem [*hearer^*])$. More generally:

$$sem [X^{(q^{(o)} \rightarrow n)} (Appurt^{(n \rightarrow q)} (Y^{(q^{(o)} \rightarrow n)}))] = sem [of] (sem [X], sem [Y])$$

More precisely:

$$sem [\lambda X^{(q^{(o)} \rightarrow n)} Y^{(q^{(o)} \rightarrow n)} \bullet X (Appurt^{(n \rightarrow q)} (Y))] = \lambda X^{(q^{(o)} \rightarrow n)} Y^{(q^{(o)} \rightarrow n)} \bullet sem [of]^{e \rightarrow (e \rightarrow t)} (sem [X]^e, sem [Y]^e)$$

3. Diatheses and diathetic shifts

Technically, our notion of semantic diathesis is very close to that of [Pad97, Pad02]³. The difference stems from the role diatheses play in discourse planning: they *specify the intended change in communicative ranks with respect to the prototypical rank assignment in the dictionary* and possibly a new value type. The *prototypical* assignment is as follows: thematic ranks are assigned only to SBJ; (⊙) is assigned to the rightmost argument (in the order of plan points) which is not otherwise assigned a

³ An account compatible with that of Padučeva can be found in [Cro91, pp. 247—260], where changes of voice are described in terms of expanding, shrinking, and moving the ‘verbal segment’ (contained between the two most topical referents, the subject and the object) against the frame of the ‘causal chain’ of an event.

thematic rank (or to the situation functor, if no such argument exists); other arguments are assigned $\ominus$. A SEMANTIC DIATHESIS of a situation S can be defined as a function assigning to the members of the communicative group of S their new ranks. It goes without saying that the function satisfies the constraints 1—4 on p. 3.

Semantic diatheses are implemented by *semantic diathesis shifts*. A diathetic shift licensed by a diathesis D is the minimal transformation of the situation profile that accomplishes all changes specified in D . We address the reader to [Dik03b] for a description of the general form of diathetic shifts. Here, we limit ourselves to several examples.

For every situation, the dictionary defines, in addition to its profile, a list of diathetic shifts which apply to it (and, of course, some other semantic and lexical derivatives, such as lexical functions). In order for a DP to be realizable, the rank assignment for every situation in the plan must match the definition of a diathesis in the dictionary and must be implementable by one of the shifts licensed by this diathesis. A (toy) example of a realizable DP is given in the gerund example below.

Example: English passivization. Passivization occurs when the OBJ actant of an effect situation is promoted to the rank of topic. There are multiple semantic diatheses satisfying this description (it should be noted that they do not directly correspond to voice diatheses of [Mel94, §§II.ii.4.6—7]). In particular, we distinguish the *f-passive* and the *p-passive* which differ in what communicative rank (*focus* or *periphery*) they assign to the SBJ of the situation. Both passive diatheses are implemented by diathetic shifts that make the former OBJ to the new SBJ and the former SBJ to Passive Agent (PAGT), an oblique role, manifested by the preposition *by*. In both cases, the value type of the diathetically shifted situation shall be $\mathbf{s}_{\text{psv}}$ (its place in the type system and the nature of the role PAGT is a complex issue which falls outside the scope of this paper). However, the *p-passive*, unlike the *f-passive*, makes PAGT an optional actant. The profiles of the two diatheses are

$$\begin{aligned} \text{dth}_{p\text{-passive}}(\text{PAGT}^{\mathbf{n}(\mathbf{O})} \curvearrowright \text{SBJ}_{\ominus}, \text{SBJ} \curvearrowright \text{OBJ}_{\bar{\top}})^{\mathbf{s}_{\text{psv}}} \\ \text{dth}_{f\text{-passive}}(\text{PAGT} \curvearrowright \text{SBJ}_{\ominus}, \text{SBJ} \curvearrowright \text{OBJ}_{\bar{\top}})^{\mathbf{s}_{\text{psv}}} \end{aligned}$$

Each of the two passives applies quite predictably and in the same manner to both versions of *give* (p. 5). If we take the situation which, under prototypical rank assignment, has the form $Eve_{\text{SBJ}:\bar{\top}} \text{ gave}_{\ominus} \text{ the apple}_{\text{OBJ}:\ominus} \text{ to Adam}_{\text{RCP}:\ominus}$, the diathetic shift for *f-passive* transforms it to $\text{The apple}_{\text{SBJ}:\bar{\top}} \text{ was given}_{\ominus} \text{ to Adam}_{\text{RCP}:\ominus} \text{ by Eve}_{\text{PAGT}:\ominus}$. The situation with the active form $Eve_{\text{SBJ}:\bar{\top}} \text{ gave}_{\ominus} \text{ Adam}_{\text{OBJ}:\ominus} \text{ the apple}_{\text{AOBJ}:\ominus}$ trans-

forms to $Adam_{\text{SBJ}:\bar{\tau}}$ was given $\ominus$ the apple $_{\text{AOBJ}:\ominus}$ by Eve $_{\text{PAGT}:\ominus}$. *P*-passive optionalizes the PAGT: $The\ apples_{\text{SBJ}:\bar{\tau}}$ was given $\ominus$ [by Eve $_{\text{PAGT}:\ominus}$] to Adam $_{\text{RCP}:\ominus}$; $Adam_{\text{SBJ}:\bar{\tau}}$ was given $\ominus$ [by Eve $_{\text{PAGT}:\ominus}$] the apple $_{\text{AOBJ}:\ominus}$.

Example: English gerund. The gerund is brought about by the need to demote the subject actant and to change the type of the affected situation to n_{ger} (which reflects compositional requirements and sometimes aspect). The plan below shows the gerund diathesis $\langle\langle \text{dth}_{\text{ger}}^{n_{\text{ger}}} (\emptyset \curvearrowright \text{SBJ}_{\ominus})^{n_{\text{ger}}} \rangle\rangle$ being applied to the situation $\langle\langle \text{watch}(\text{SBJ}^{n_a}, \text{OBJ}^{n_{\text{Spercep}}}) \rangle\rangle$.

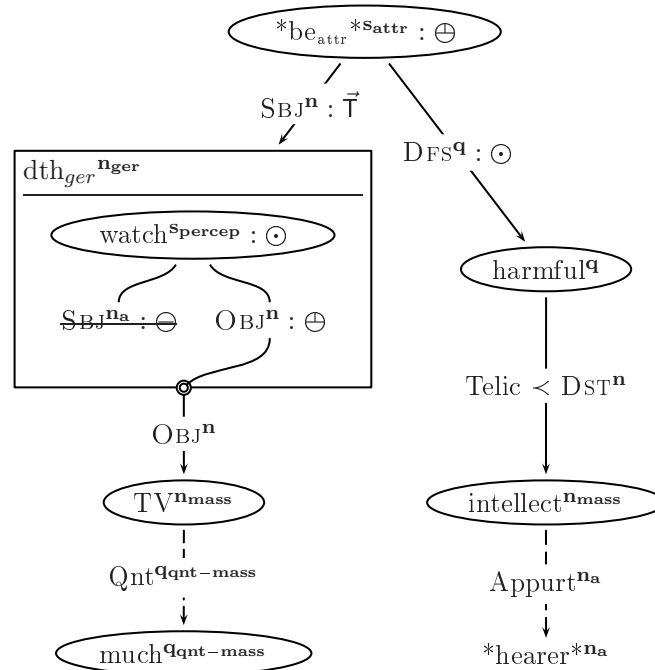


Fig. 1. Discourse plan for the utterance *Watching much TV is harmful for your intellect.*

Ellipses signify functional semantemes.

The box represents a diathesis (communicative rank assignment) and the corresponding diathetic shift (the crossed role signifies a deleted actant, the connections to the outside of the box signify the projection of roles defined by the diathetic shift).

Solid lines link actants to their situations and are labeled with roles.

Dashed lines link circumstantials to their governor semantemes and are labeled with names of attributes.

For discourse plans of this kind, we have implemented a graphical editor, providing the functionality necessary to develop and maintain meaning structure treebanks. It features menu-driven plan construction, realizability testing, graphical and typesetting language output, high-level inheritance structuring. E. g., one can define very general passive patterns and then determine their particular cases in compact and observable form.

4. Conclusion

In this paper, we develop the concept of semantic diathesis introduced by Padučeva in [Pad97, Pad02]. We seek to represent discourse planning guided by lexical choice and by argument salience preferences induced by the intended communicative structure. We extend Padučeva's semantic diatheses by compositional functional semantic types with optionality and inheritance and describe the differences in semantic interpretation of actants and circumstantials of situations. In order to arrive at a practical application of discourse plans, we outline a strategy, consistent with lexicographic definitions and typological casebased analysis, to induce linguistic criteria which would serve to distinguish between actants and circumstantials. The method consists in expanding a small set of basic cognitive proto-situations by means of projection operations. This aspect of our work is in a preliminary phase and requires an analysis of considerable linguistic material.

References

- [Apr74] Ju. D. Apresjan. *Leksičeskaiâ semantika*. M.: Nauka, 1974.
- [But04] M. Butt. Theories of Case. Ms. to be published by Cambridge University Press, in preparation.
- [Cro91] W. Croft. *Syntactic Categories and Grammatical Relations: The cognitive organization of information*. University of Chicago Press, 1991.
- [Dik03a] A. Dikovsky. Discourse Plan $\Rightarrow$ Meaning. In *Proc. of the 1st International Conference "Meaning-Text Theory" (MTT 2003)*, pages 299—308, Paris, France, June 2003. ENS de Paris.
- [Dik03b] A. Dikovsky. Linguistic Meaning from the Language Acquisition Perspective. In *Proc. of the 8th Intern. Conf. "Formal Grammar 2003" (FG 2003)*, pages 63—76, Vienna, Austria, August 2003.
- [Fil68] C. J. Fillmore. The case for case. In E. Bach and R. T. Harms, editors, *Universals in Linguistic Theory*, pages 1—88. Holt, Rinehart and Winston, 1968.

- [Fil77] C. J. Fillmore. The case for case reopened. In P. Cole and J. M. Sadock, editors, *Syntax and Semantics 8: Grammatical Relations*, pages 59—81. New York: Academic Press, 1977.
- [Gru65] J. S. Gruber. *Studies in Lexical Relations*. PhD thesis, MIT, 1965.
- [KKM00] A. E. Kibrik, S. V. Kodzasov, and I. A. Muravjova. *Ĵazyk i fol'klor alĴutorcev*. M.: IMLI RAN / “Nasledie”, 2000.
- [KL04] L. Kolehmainen and M. Larjavaara. The “bizarre” valency behaviour of Finnish verbs: How a specific context gives rise to valency alternation patterns. *Constructions 1/2004*, 2004. <http://www.constructions-online.de>.
- [Lam94] K. Lambrecht. *Information Structure and Sentence Form: Topic, focus, and the mental representations of discourse referents*. Cambridge University Press, 1994.
- [Mel94] I. Mel'čuk. *Cours de morphologie générale. Volume II*. Presses de l'Université de Montréal/CNRS, Montréal, 1994.
- [Mel04a] I. A. Mel'čuk. Actants in semantics and syntax I: Actants in semantics. *Linguistics*, 42(1): 1—66, 2004.
- [Mel04b] I. A. Mel'čuk. Actants in semantics and syntax II: Actants in syntax. *Linguistics*, 42(2): 247—291, 2004.
- [MX70] I. A. Mel'čuk and A. A. Xolodovič. K teorii grammatičeskogo zaloga. *Narody Azii i Afriki*, 4, 1970.
- [Pad97] E. V. Padučeva. Semantičeskie roli i problema soxraneniā invarianta pri leksičeskoj derivacii. *Naučno- texničeskaĴ informaciā, Ser. 2*, (1): 18—30, 1997.
- [Pad98] E. V. Padučeva. Kommunikativnoe vydelenie na urovne sintaksisa i semantiki. *Semiotika i informatika*, 36: 82—107, 1998.
- [Pad02] E. V. Padučeva. Diathesis and diathetic displacement. *Russian Linguistics*, 26(2): 179—215, 2002.

Semantic Motivation of Lexical Functions and its Limits (The Combinatorial Profile of the Russian Degree Modifier *črezvyčajno*)¹

D. O. Dobrovol'skij

Russian Language Institute, Russian Academy of Sciences
dm-dbrv@yandex.ru

The point of departure is the idea that many combinatorial phenomena which were traditionally regarded as arbitrary, in fact, are motivated semantically, at least to a certain extent. The empirical data analysed here is the use of the Russian degree modifier *črezvyčajno* which incarnates the Lexical Function MAGN. The factors that motivate its selection constraints can be either of a semantic or pragmatic nature, or its co-occurrence can be driven by usage.

Границы семантической мотивированности в реализации лексических функций (к сочетаемости наречия степени *чрезвычайно*). В последнее время все более актуальной становится идея наличия семантической мотивации у комбинаторных феноменов, традиционно считавшихся немотивированными. Настоящая работа развивает этот тезис. Материалом исследования послужило наречие степени *чрезвычайно*, реализующее лексическую функцию MAGN. Факторы, мотивирующие особенности сочетаемости, могут быть по своей природе семантическими, прагматическими или узуальными.

1. Preliminary Remarks

The idea that many combinatorial phenomena are motivated to a certain extent is increasingly being accepted. This concerns, among other, words incarnating Lexical Functions (LFs) which were traditionally considered to be lacking of their own lexical meaning, so that their selection was driven only by usage conventions. However, the recent research proves that these phenomena also display a certain degree of motivation. For example, in [Апресян 2004] it is shown that every verb from the

¹ This paper is based on work supported by the RGNF under Grant 05-04-04026a.

group OPER-FUNC-LABOR possesses its own lexical meaning, which duplicates the generic part of the meaning of its argument, and therefore that the selection of LF verbs is semantically motivated, to within the predicate class, though not quite free².

This paper develops this thesis. The empirical data analysed here is the use of the Russian degree modifier *чрезвычайно* (*črezvyčajno*) which roughly means ‘extremely’ (literally “extraordinarily”) and incarnates the Lexical Function MAGN. It differs from the LFverbs analysed in [Апресян 2004] in one significant way. Above all, there has never been any doubt that words like *чрезвычайно* have their own lexical meanings. The problem here is rather that the semantic structures of near-synonymous adverbials (cf. *чрезвычайно*, *необыкновенно*, *необычайно*, *крайне*, *бесконечно*, *безмерно*) do not, at first glance, reveal any relevant distinctive features which could be regarded as factors motivating selection constraints. In fact, how can we explain that certain words prefer to combine with *чрезвычайно*, others take instead *крайне* ≈ ‘utterly’ as an intensifier, some take *неимоверно* ‘incredibly’, and others *бесконечно* ‘infinitely’? Why can we say both *крайне мало* ≈ ‘utterly little’ and *чрезвычайно мало* ≈ ‘extremely little’, but not **крайне много* ≈ ‘utterly much’, whereas *чрезвычайно много* ≈ ‘extremely much’ is again quite acceptable? The simplest way to answer these questions would be an explanation via irregular and unpredictable usage conventions. Another possibility is to look for certain motivating factors, which do not have to possess predicting power and to go back to the differences in the semantic definitions of the words in question. If such factors could be found it would mean that the combinatorial behaviour of degree modifiers mentioned above is not quite arbitrary and can be described in terms of semantic (and maybe also pragmatic) correlations.

My basic assumption is that the combinatorial profile of every word has a core and a periphery. Some lexical co-occurrences are perceived as corresponding fully to the usage norm and stylistically irreproachable, others as being erroneous, and some in the middle being more or less acceptable. The first ones are considered to be central, the second ones are ruled out by the usage norm, whereas those in-between form a broad and vague periphery of the combinatorial profile. The latter part of the com-

² Cf.: “любой глагол семейства OPER-FUNC-LABOR имеет собственное лексическое значение, дублирующее родовую часть значения слова-аргумента, и что поэтому выбор значения ЛФ — с точностью до класса предикатов — семантически мотивирован, хотя и не вполне свободен” [Апресян 2004: 10].

binatorial profile is especially important from the point of view of linguistic change. It is impossible that a given word can abruptly change its contextual behaviour. What seems to be more probable is the hypothesis that peripheral co-occurrences of a given word can change their status in due course. Some of them can change from periphery to centre, whereas others can be excluded from the domain of normative acceptance. In certain cases, combinatorial changes of this kind can lead to the modification of the semantic structure.

The starting point of the investigation into the combinatorial properties of the word *чрезвычайно* was my focus on the use of this word in Dostoevskij's work³.

2. Analysis of co-occurrence

2.1. Usage of the 19th Century

The degree modifier *чрезвычайно* is a very frequent lexical item encountered in Dostoevskij's prose. Some uses of *чрезвычайно* in his works match the present day norm closely (*чрезвычайно важный секрет* 'an extremely important secret', *чрезвычайно простой вопрос / ответ* 'an extremely simple question / answer', *чрезвычайно интересная история* 'an extremely interesting story', *чрезвычайно сильное воображение* 'an extremely strong imagination', *чрезвычайно любопытная болезнь* 'an extremely curious illness'), whereas other occurrences obviously violate them, i.e. they are perceived as being strange by contemporary readers, cf. *чрезвычайно выкатывала глаза* literally "[she] extremely rolled out her eyes", *чрезвычайно раскрыв свой рот* "extremely opening his mouth", *лестница была (...) чрезвычайно грязная* "the stairs were extremely dirty". It can be readily shown that co-occurrences of this kind are not restricted to Dostoevskij, but are also encountered in other written works of the 19th century, in other words, they were typical of that time. Taking the works of the following authors into account (Puškin, Gogol', L. Tolstoj, Turgenev, Leskov, Nekrasov, Grigorovič, Herzen, Bestužev-Marlinskij, Veresaev, Boborykin, Pisemskij, Zagoskin, Čexov, Korolenko), corroborates the assumption that in the 19th century the degree modifier *чрезвычайно* had a broader combinatorial profile than it has today.

³ This is a part of the research project based on a Russian-German parallel corpus developed within the Austrian Academy Corpus (AAC) at the Austrian Academy of Sciences in Vienna.

2.2. Present day usage

In this section, the contemporary usage of *чрезвычайно* is analysed, in order to establish the relevant usage norms. Obsolete contexts which deviate from these norms serve as “negative linguistic material” (in the sense of Ščerba). The analysis of these contexts, as compared to the contexts from corpora of present day texts⁴, reveals the following tendencies that are typical of the contemporary usage of the word *чрезвычайно*.

1. The degree modifier *чрезвычайно* combines with adjectives and adverbs rather than with verbs. Co-occurrences such as *чрезвычайно интересный / заинтересован* ‘extremely interesting / interested’ sound better than, for example, *чрезвычайно интересуюсь* literally “I extremely interest myself”.

2. Words in the scope of *чрезвычайно*, more often denote “non-physical” properties (*чрезвычайно тяжелое положение* ‘an extremely difficult — literally “heavy” — situation’ is better than *чрезвычайно тяжелый чемодан* ‘an extremely heavy suitcase’).

3. Words in the scope of *чрезвычайно* tend to denote features which are biased towards the positive pole of various scales. For example, co-occurrences like *чрезвычайно широкий* ⟨*одаренный, энергичный*⟩ ‘extremely broad ⟨gifted, energetic⟩’ are encountered more often than co-occurrences such as *чрезвычайно узкий* ⟨*бездарный, вялый*⟩ ‘extremely narrow ⟨ungifted, inert⟩’.

Obviously, these tendencies should not be interpreted as strict rules.

Before discussing every of these tendencies in some detail, I would like to point out one more specific property of present day contexts with *чрезвычайно* as contrasting with contexts from the 19th century. Currently, the adverbial *чрезвычайно* is perceived as a non-ordinary word. This makes it bind to certain thematic domains and to a discourse type which can be labelled “not ordinary discourse” (in contrast to everyday language).

2.2.1. Let us consider the first tendency mentioned above. It is obvious that the co-occurrence of *чрезвычайно* with verbs of certain thematic classes (in the sense of Padučeva, cf. e.g. [Падучева 2003]) is in princi-

⁴ The following text corpora were used: (1) National Corpus of Russian (<http://www.ruscorpora.ru>), (2) Russian corpora at the University of Tübingen (Germany) — Tübinger Russische Korpora: <http://www.sfb441.uni-tuebingen.de/b1/korpora.html>, (3) the textual database of the Digital Fund of the Russian Language (Машинный фонд русского языка), and the corpora of the Department of Experimental Lexicography at the Russian Language Institute of the Russian Academy of Sciences.

ple impossible, for example, with verbs denoting space position, parameter or existence, such as *stand*, *sit*, *contain*, *number*, *be*. This is trivial from the point of view of the issues discussed here, because such verbs cannot occur in the scope of any degree modifier. However, these natural constraints do not explain the relevant combinatorial peculiarities of *чрезвычайно* seen in text corpora. The combinatorial profile of this word in the domain of verbs differs from that of other degree modifiers, such as *очень*; compare *я очень опоздал* <*спешил, просил*> literally “I was very late <in a hurry>, I very asked”, but “*я чрезвычайно опоздал* <*спешил, просил*> “I was extremely late <in a hurry>, I extremely asked”. It is easier and more economical to list the verbs which co-occur with *чрезвычайно* than those which do not. In the corpora of present day texts I repeatedly came across verbs such as *(по)нравиться* ≈ ‘to like’, *(за)интересовать* ≈ ‘to interest’, *возрасти* ‘to grow’, *ценить* ‘to appreciate’, *(вз)волновать* ≈ ‘to agitate’, *обрадовать* ‘to make happy’, *расширять(ся)* ‘to widen’, *способствовать чему-л.* ‘to promote sth.’ in the scope of *чрезвычайно*.

It cannot be claimed that the restrictions in the co-occurrence of *чрезвычайно* with verbs that allow for graduating in principle are motivated by the thematic class of a given verb. However, this does not mean that such restrictions are unmotivated. Certain semantic features turn out to be relevant for the combinatorial behaviour of the word *чрезвычайно*. In most cases, the features promoting the co-occurrence of verbs with this degree modifier operate only within certain thematic classes.

For reasons of space, I restrict myself to pointing out only one of these semantic features, namely the bias of the adverbial *чрезвычайно* towards co-occurrence with words that lack the feature ‘action’. Rare exceptions are some verbs denoting abstract actions, such as *упростить* ‘to simplify’. The more prominent the feature ‘action’ is in the semantic structure of a given verb, the stronger then is this verb biased against the co-occurrence with *чрезвычайно*; cf. **чрезвычайно выкатывать глаза* literally “to extremely roll out the eyes” or **чрезвычайно раскрыть рот* “to extremely open the mouth”.

Following this line of reasoning, we can assume that this semantic factor determines in general the tendency towards avoiding verbs in the scope of this modifier. The verbs are prototypical designations of actions. Even verbs belonging to non-agentive taxonomic classes are perceived as names of certain “quasi-actions”, at least to a certain degree. This is especially obvious while contrasting finite verbs with the corresponding participles or adjectives. Compare, for example, *чрезвычайно распространить* <*распространиться*> ‘to extremely spread (sth.)’ vs. *чрезвычайно*

распространен ⟨*распространенный*⟩ ‘to be extremely spread’. In the text corpora used here I did not encounter any instances of the co-occurrence of the former type whereas the co-occurrence of the latter type is richly presented. Hence, the “rule 1” gets a certain semantic explanation.

2.2.2. Let us briefly consider “rule 2” and “rule 3” that describe the current combinatorial norms of the adverbial *чрезвычайно*. Here we deal with tendencies according to which *чрезвычайно* combines with (a) words denoting “non-physical” entities, and (b) with words denoting features which are biased towards the positive pole of various scales, for example, with words whose semantic structure contains the feature of positive evaluation or desirability.

The relevance of feature (a) becomes particularly obvious while comparing two meanings of the same word, one of which denotes a phenomenon of the “visible world” and the second one a phenomenon of the “invisible world”. For instance, the word *простой* ‘simple’ forms with *чрезвычайно* a word combination which is perceived as not quite usage conformable from the perspective of current norms, if the object being described is a physical one: a building, a piece of furniture or a dress (cf. ⟨...⟩ *сказала она Анне, которая в третьем, опять в чрезвычайно простом, платье вышла к ней.* [Л. Н. Толстой. Анна Каренина]). But at the same time, the lexical co-occurrence *чрезвычайно простой* is considered to be completely normative in contexts dealing with tasks, principles or situational circumstances. This is not a single case. The co-occurrence *чрезвычайно высокий / высоко* ‘extremely high ⟨tall⟩ / highly’ is perceived as corresponding to the current usage norms in contexts such as *чрезвычайно высоко ценить кого-л. / что-л.* ⟨*отзывать-ся о ком-л. / чем-л.*⟩ ‘to estimate sb. / sth. extremely highly, to give high praise to sb. / sth.’, *чрезвычайно высокий уровень безработицы* ‘an extremely high level of unemployment’, *чрезвычайно высокая стоимость* ‘an extremely high cost’, *чрезвычайно высокий рейтинг / статус* ‘an extremely high rating / status’, but not in the contexts such as *чрезвычайно высокий человек* ‘an extremely tall person’.

As far as the bias of the combinatorial profile of *чрезвычайно* towards the “positive pole” is concerned, it has to be stressed that here we are dealing not only with the opposition of “good” or “desirable”, on the one hand, and “bad” or “undesirable”, on the other⁵, but also with all

⁵ The sensitivity of the word *чрезвычайно* to evaluative properties of its contextual partners was pointed out in the previous research; cf., for example, [Червенкова 1975: 17].

other semantic scales, whose poles can be interpreted as positive vs. negative; compare the contrasting of “big” and “small”, of “long / tall” and “short”, and the like. This tendency should not be taken as a strict rule. It expresses itself in statistical preferences rather than in strict constraints. Co-occurrences such as *чрезвычайно высоко* *⟨активно, ярко⟩* ‘extremely highly ⟨actively, brightly⟩’ are encountered in present day texts much more often than co-occurrences such as *чрезвычайно низко* *⟨пассивно, бледно / тускло⟩* ‘extremely low ⟨passively, dimly / wanly⟩’.

2.2.3. Interpretation of the results must attempt to find some common properties that are shared by the tendencies discussed in the previous sections. Though these tendencies appear to be independent from each other, at a deeper level, they reveal common features that can lead back to more general phenomena. Usually, the question is whether in the semantics and / or pragmatics of a given word properties can be found, which could explain the observed combinatorial constraints and preferences.

2.2.3.1. In order to answer this question we need at least a tentative definition of the degree modifier *чрезвычайно*:

чрезвычайно P ≈ ‘P in a very high degree; the speaker claims that the degree of P is so high that it exceeds the bounds of the domain of ordinariness’

Under “the bounds of the domain of ordinariness” I understand the usual bounds of variability of the feature P, i.e. such bounds of its variability within which the degree of P is perceived as not violating habitual expectations. This domain includes also a high degree of P, i.e. the meaning ‘very P’.

Let us compare this definition with definitions of near-synonymous degree modifiers, such as *бесконечно*, *безгранично*, *безмерно* ≈ ‘infinitely’ — in [Григорьева 2004]⁶ — or *крайне*, which can be approximately defined in the following way:

крайне P ≈ ‘P in a very high degree; the speaker claims that the degree of P is so high that it is near the bound possible for P’

⁶ Cf.: “*бесконечно* *⟨безгранично, безмерно⟩ P* = ‘P в очень высокой степени; говорящий считает, что степень P настолько высока, что это трудно представить’” [Григорьева 2004: 15] — *бесконечно* *⟨безгранично, безмерно⟩ P* = ‘P in a very high degree; the speaker claims that the degree of P is so high that it is difficult to imagine’.

The relevant semantic differences between these degree modifiers evidently influence the meaning of the utterance containing them. These semantic differences can be traced back to the specifics of the inner form or “etymological memory”, which is potentially able to express itself in the actual meaning of lexical items (for more detail see [Добровольский 1996; Баранов, Добровольский 1998; Зализняк 1998]). Semantic differences of this kind turn out to be relevant, first of all, in contexts profiling just those meaning components that are inherited from the inner form. Nevertheless, since the traces of the source concept are present, in the latent form, in the content plane of a given lexical item, it can be assumed that they can have an impact on the combinatorial profile.

The adjective *чрезвычайный* came into Russian between the late 17th and the beginning of the 18th century, and first was used only in the expression *чрезвычайный посол* ‘Ambassador Extraordinary’. The derivation *чрезвычайно* was fixed for the first time in the first half of the 18th century. *Чрезвычайный* is a hybrid loan-translation from the Polish *nadzwyczajny* with a by the morpheme substitution: the prefix *чрез-* from the Church Slavonic instead of the Polish *nad-* + *звычайный* (cf. the Polish *zwyczajny*) = *чрез-звычайный*, developed to *чрез-вычайный* (see for more detail [Gardiner 1983: 512—517]).

This principle assumption lays the foundation for a hypothesis about certain combinatorial properties resulting from the inner form of the adverbial *чрезвычайно*. From the definition of this word the assumption is that its preferred co-occurrence demands lexical items denoting properties for which the DOMAIN OF ORDINARINESS can be presumed in a natural way. The idea of the DOMAIN OF ORDINARINESS is related to the notion of NORM. Here I prefer not to use the word *norm* as an element of metalanguage because of its polysemy (compare similar ideas in [Урысон 2000]). Therefore, with regard to the DOMAIN OF ORDINARINESS, it should be explicitly stressed that here we are dealing with the “standard norm” (i.e. with the idea of the characteristics typical of a given situation or a given object) rather than with the “deontic” or “teleological” norm.

2.2.3.2. It can be assumed that the feature ‘domain of ordinarieness’ is, to a certain extent, responsible for the combinatorial properties of the word *чрезвычайно*. By virtue of the relatively transparent morphological structure of this word, its “etymological memory” addresses the idea of bounds of habitual expectations, rather than the NORM as a point on a certain scale. Connections of the adverbial *чрезвычайно* with the idea of exceeding the bounds of ordinarieness are also maintained by the seman-

tics of the adjective *чрезвычайный*, first of all, taken even not so much in the corresponding meaning (cf. *чрезвычайно удивлен* — *чрезвычайное удивление* ‘extremely astonished — extreme astonishment’), as in the meaning that is realised in co-occurrences, such as *чрезвычайный съезд* ‘extraordinary congress’, *чрезвычайные меры* ‘extraordinary measures’.

Following this line of reasoning, we can put forward a hypothesis that the feature ‘domain of ordinariness’ forms a semantic bridge between some of those seemingly isolated combinatorial factors that were discussed above. It is extremely difficult to postulate the DOMAIN OF ORDINARINESS for actions, such as “staring” or “laughter”. Obviously, this is the reason why the co-occurrences *чрезвычайно выкатывала глаза* literally “[she] extremely rolled out her eyes” and *чрезвычайно расхохотался* “extremely bursted out laughing” used by Dostoevskij are perceived as norm violating. Similar contexts could not be found in texts from other authors of the 19th century. Presumably, co-occurrences of this kind were perceived even then as being marginal. Therefore, it can be assumed that the bias of the adverbial *чрезвычайно* towards co-occurrence with words that lack the feature ‘action’ relates to the idea of the domain of ordinariness. As a whole, ‘states’ and ‘properties’ associate with the domain of ordinariness in a more natural way than ‘actions’. Also, the bias towards the “positive pole” correlates, to some extent, with the idea of the domain of ordinariness. Admittedly, such an interpretation of the concept of ORDINARINESS relating back to the idea of the “standard norm” receives an additional “deontic touch”.

2.2.3.3. An important group of factors that govern the combinatorial properties of the adverbial *чрезвычайно* must be described in terms of pragmatics. The most significant parameter is here the type of discourse, i.e. the set of thematic domains and stylistic registers that form a natural environment for the word *чрезвычайно*. Probably, this is exactly the parameter that is common for the lexical items which, in spite of their semantic heterogeneity, can potentially occur in the scope of this modifier. Preference of contexts belonging to “not ordinary discourse” in the vicinity of the modifier *чрезвычайно* becomes particularly evident when comparing co-occurrences such as *чрезвычайно короткий <длинный> доклад* ‘an extremely short <long> presentation’, on the one hand, and *чрезвычайно короткие <длинные> волосы* ‘extremely short <long> hair’, on the other. The difference in the degree of acceptability of these contexts is due not so much to the contrast of physical and non-physical entities (though the temporal meanings of the words *короткий* ‘short’ and *длинный* ‘long’ are evidently less “physically profiled” than their spatial

meanings) as to the thematic domain of the discourse in question. Obviously, there is a certain correlation between this pragmatic factor and the bias towards co-occurrence with names of “non-physical” entities: discussing abstract issues rather than outward, physical properties of things and people is more typical of discourse types which can be labelled “not ordinary”. Moreover, it can be assumed that the “non-physical tendency” is also governed by the relevant pragmatic features of the word *чрезвычайно*.

3. Discussion

Factors motivating the combinatorial properties of a lexical item can be either of a **semantic** or **pragmatic** nature, or its co-occurrence can be driven by **usage**. With regard to the degree modifier *чрезвычайно*, it is evident that its bias against co-occurrence with verbs is, to a certain extent, motivated semantically. The explanation for this tendency is that the adverbial *чрезвычайно* generally prefers to combine with words whose meanings lack the feature ‘action’. This semantic factor, together with the “tendency to the positive pole”, can at least hypothetically, go back to certain components in the content plane of the word *чрезвычайно*: namely, to that part of its semantic definition which refers to the inner form of this word. Its plane of content contains the idea of exceeding the bounds of habitually expected variability (compare also its nearest synonyms *необыкновенно* and *необычайно* $\approx$ ‘extraordinarily, exceptionally’). This semantic component has a motivating function, on the one hand, because the DOMAIN OF ORDINARINESS can be associated with ‘properties’ and ‘states’ in a more natural way than with ‘actions’, and on the other, because the negative values of various parameters are perceived, in the naive model of the world, as norm violation, and therefore, do not easily correlate with the idea of habitual expectations.

Pragmatic factors, namely the markedness of the word *чрезвычайно* as “not ordinary”, manifest themselves, above all, in that this word is perceived as more appropriate in contexts of “well educated” discourse rather than in thematic domains typical of everyday language. Furthermore, this factor influences the stylistic parameters of contextual partners of the word *чрезвычайно*. Compare, for example, *чрезвычайно алчный* $\approx$ ‘extremely acquisitive’ vs. the stylistically not quite irreproachable co-occurrence *чрезвычайно жадный* ‘extremely greedy’. The same pragmatic factors explain the bias towards co-occurrences with words denoting “non-physical” entities and, to a certain extent, the “tendency to the

positive pole” because, in the “well educated” discourse, the discussed issues are mostly abstract, non-physical entities, and, in the normal case, lexical items with negative evaluative components are avoided. It is these pragmatic properties that distinguish the modifier *чрезвычайно* from its nearest synonyms *необычайно* and *необыкновенно*. Being not bound to some special type of discourse, these words do not display any selection constraints concerning lexical items denoting physical properties. For example, *необыкновенно хорош собой* ‘exceptionally handsome’ sounds better than *чрезвычайно хорош собой* ‘extremely handsome’, and *необычайно большие глаза* ‘exceptionally big eyes’ better than *чрезвычайно большие глаза* ‘extremely big eyes’. Thus with regard to the combinatorial profile, pragmatics is not less important than semantics.

The combinatorial profile of the word *чрезвычайно* also shows those properties which can be explained neither in terms of semantics nor in terms of pragmatics. This is, so to say, pure usage. The existence of such cases is naturally determined because it is a well-known fact that in Russian the co-occurrence of degree modifiers (as well as words of many other semantic classes) is lexicalised to a large extent. Nevertheless, in general the combinatorial behaviour of the word *чрезвычайно* turns out to be motivated to a certain degree. In this sense, it is reasonable to postulate a sphere of co-occurrence which takes an intermediate position between semantically constrained and purely lexically constrained co-occurrence (in the sense of Apresjan [Апресян 1974]). The selection of potential contextual partners is neither completely arbitrary (as it is the case with the lexical co-occurrence) nor strictly predictable (as it is the case with the semantic co-occurrence). This “intermediate” type of co-occurrence has to be described in terms of tendencies which mainly operate within certain lexical classes.

* * *

I am most grateful to Ju. D. Apresjan, E. V. Padučeva and E. V. Uryson for sharing with me their ideas about the issues discussed in this paper.

References

- Апресян 1974 — Ю. Д. Апресян. Лексическая семантика: Синонимические средства языка. М., 1974.
Апресян 2004 — Ю. Д. Апресян. О семантической непустоте и мотивированности глагольных лексических функций // *Вопр. языкознания*. 2004. № 4. С. 3—18.

- Баранов, Добровольский 1998 — *А. Н. Баранов, Д. О. Добровольский*. Внутренняя форма идиом и проблема толкования // Изв. АН. Сер. лит-ры и языка. Т. 57. 1998. № 1. С. 36—44.
- Григорьева 2004 — *С. А. Григорьева*. Словарная статья БЕСКОНЕЧНО... // Новый объяснительный словарь синонимов русского языка. 2-е изд., испр. и доп. / Под рук. Ю. Д. Апресяна. М.; Вена, 2004. С. 15—17.
- Добровольский 1996 — *Д. О. Добровольский*. Образная составляющая в семантике идиом // Вопр. языкознания. 1996. № 1. С. 71—93.
- Зализняк 1998 — *Анна А. Зализняк*. О месте внутренней формы слова в семантическом моделировании // Труды Международного семинара Диалог '98 по компьютерной лингвистике и ее приложениям. Т. 1. Казань, 1998. С. 61—68.
- Падучева 2003 — *Е. В. Падучева*. Таксономическая категория как параметр лексического значения глагола // Русский язык в научном освещении. № 2. 2003. С. 192—216.
- Урысон 2000 — *Е. В. Урысон*. Понятие нормы в метаязыке современной семантики (параметры человеческого тела с точки зрения русского языка) // Слово в тексте и в словаре: Сб. ст. к 70-летию академика Ю. Д. Апресяна. М., 2000. С. 243—252.
- Червенкова 1975 — *И. В. Червенкова*. Общие адвербиальные показатели меры признака (в современном русском литературном языке). Автореферат ... канд. филол. наук. М., 1975.
- Gardiner 1983 — *S. C. Gardiner*. Loan adaptation and the discovery of the genetic relationships of language // SEER. V. 61. №. 4. London, 1983. P. 512—517.

On the Aspectual Derivation of “Dual Simplexes”

Ronald F. Feldstein

Indiana University
feldstei@indiana.edu

Paradigmatic relations of word-formation are considered for both single and dual simplex (non-prefixed) verbal stems. Dual simplexes are usually linked to motion verbs, but the paper cites non-motion verbs where dual simplex stems derive two different perfective stems, e.g. *хватит-* / *хватит-* and *захватит-* / *захватит-*. The regular syncretism of the derived imperfective (*захватывай-*) is noted, along with other rules for such verbs.

О видовом словообразовании при наличии двух бесприставочных основ. Рассматривается парадигматика русского глагольного словообразования в случаях одной / двух бесприставочных (симплексных) основ. Наличие двух симплексных основ обычно связано с глаголами движения; доклад приводит ряд глаголов не движения, где двойные бесприставочные основы порождают две основы совершенного вида, типа *хватит-* / *хватит-* и *захватит-* / *захватит-*. Отмечается синкретизм в приставочной форме несовершенного вида (*захватывай-*) и ряд других правил.

0. Introduction

This paper is a comparative study of the different possible aspectual paradigms of Russian verbal stems on three different derivational levels. The first such level is represented by the **simplex stem**, which is defined as a **non-prefixed** stem (Timberlake 2004: 94). The stem *работай-* fits this definition. The second level of derivation represents the addition of a prefix to the simplex stem, which has the invariant property of changing the aspect to perfective, e.g. the prefixes *пере-* or *но-* can be added to *работай-*, resulting in the perfective stems *переработай-*, *ноработай-*. The third level is that of secondary imperfective suffixation; this level is only activated when the prefix of the second level has added a lexical meaning that differs from that of the simplex. When the second level prefixation has added only a sublexical difference of meaning (called **Aktionart** or **способ действия**), there may be no imperfective suffixation

at level three. Native speakers and dictionaries often disagree about whether a given prefixed perfective can indeed form a secondary suffixed imperfective. In our example, the stem *переработай-* permits the formation of the suffixed imperfective *перерабатывай-*, but the stem *нопаботай-* differs only sublexically from the simplex stem *работай-* and, therefore, *нопаботай-* does not form a secondary imperfective. Note that this verb has only one simplex stem (*работай-*), and represents the most common Russian verb type. However, certain small classes of Russian verbs can have two simplex stems, e.g. the so-called motion verbs (e.g. *кату-* / *камай-*), as well as several non-motion verbs (e.g. *хвату-* / *хвамай-*). I will refer to such instances of two simplex stems as **dual simplexes**. While the existence of dual simplexes is frequently mentioned in reference to the motion verbs of Russian, it has rarely, if ever, been discussed in relation to the non-motion verb category. This paper will attempt to examine the paradigmatic differences between the three derivational levels of verbs with a single simplex, as well as both motion and non-motion verbs with two simplex stems.

1. The single simplex type

As mentioned above, this is the simplest type of aspectual derivation. There is one simplex, which is imperfective, and the addition of a prefix to the simplex stem creates a perfective, which may or may not form a derived imperfective. This paradigmatic relation has been shown in table 1. Part 1A shows the non-Aktionsart type of paradigm with a derived imperfective in the third cell, while 1B illustrates the Aktionsart situation, with a defective third cell.

Table 1. One simplex and one prefixed perfective with a given prefix

1A. With a derived imperfective		1B. Without derived imperfective (Aktionsart)	
Simplex:	<i>работай-</i>	Simplex	<i>работай-</i>
Prefixed perfective	<i>переработай-</i>	Prefixed perfective	<i>нопаботай-</i>
Derived imperfective	<i>перерабатывай-</i>	Derived imperfective	-----

2. Verbs of motion

By definition, a **verb of motion** possesses two unprefixated imperfective stems (i.e. **dual simplexes**), which are grammatically opposed by the

feature determinate / indeterminate (also known as monodirectional / polydirectional). In the motion verb paradigm (tables 2 and 3), two simplexes are shown on the first level of the derivational paradigm. The second level contains two prefixed perfectives, one derived from the determinate, and the other from the indeterminate. Verbs of motion can be divided into two groups on the basis of the formal relation of the two simplex stems. One group of dual simplex stems shares the same root and only differs in the suffixal part of the stem (e.g. *кати-* / *катай-*, *тащи-* / *таскай-*, *лете-* / *летай-*, *полз-* / *ползай-*), shown in table 2. The other group has simplexes with suppletive roots or various irregular root alternations (e.g. *ид-* / *ходи-*, *нес-* / *носи-*, *лез-* / *лази-*), as seen in table 3.

Table 2. Non-Suppletive Verbs of Motion

A. Full paradigm

Simplex stems	<i>кати-</i>	<i>катай-</i>
Prefixed perfective	<i>скати-</i>	<i>скатай-</i>
Derived imperfective	<i>скатывай-</i>	

B. Partially defective (Aktionsart) Paradigm

Simplex stems	<i>кати-</i>	<i>катай-</i>
Prefixed perfective	<i>покати-</i>	<i>покатай-</i>
Derived imperfective	-----	<i>покатывай-</i>

Table 3. Suppletive Verbs of Motion

A. Full paradigm

Simplex stems	<i>ид-</i>	<i>ходи-</i>
Prefixed perfective	<i>зайд-</i>	<i>заходи-</i>
Derived imperfective	<i>заходи-</i>	<i>захаживай-</i>

B. Defective (Aktionsart) Paradigm

Simplex stems	<i>ид-</i>	<i>ходи-</i>
Prefixed perfective	<i>пойд-</i>	<i>походи-</i>
Derived imperfective	-----	-----

The concept of word-formational suppletion has been clearly defined by Mel'čuk (1995: 463), who states, "Аналогично тому, как обстоит дело в системе словоизменения, супплетивные словообразовательные формы занимают клетки, предусмотренные системой, но не занятые." Note that all six potential cells are never filled with six different

forms; the maximum number of different cells is five, as seen both in tables 2A and 3A. This suggests that both suppletive and non-suppletive verbs of motion have a **regular syncretism of two cells**. In the case of *кату-* / *катай-* (table 2A), the syncretism is seen in the shared derived imperfective form for both stems (*скатывай-*). In the case of the suppletive simplexes *уѡ-* / *хѡду-*, the syncretism occurs between the prefixed perfective derived from the indeterminate and the derived imperfective which is ultimately derived from the determinate simplex (*захѡду-*).

In 2A and 3A, one sees that each simplex can derive both a prefixed perfective and a derived imperfective. Although the simplexes themselves are grammatically opposed as determinate vs. indeterminate, this feature does not oppose the prefixed forms on levels two and three of the paradigms. In our specific example of table 3A, the stems *уѡ-* and *хѡду-* clearly represent the determinate / indeterminate opposition, but does the opposition of the two prefixed perfective stems *зауѡ-* and *захѡду-* embody this opposition in any way? Does the grammatical opposition of determinacy convert to a lexical opposition of an Aktionsart vs. non-Aktionsart type after the derivational process of prefixation? While a definitive answer cannot be given at this point, it does look as if the feature of determinacy (at level one) is in complementary opposition with whatever lexical features oppose the two stems at levels two and three.

4. Dual simplexes other than verbs of motion

While verbs of motion have often been treated in terms of their dual simplexes, representing the feature of determinacy, I know of no specific discussion of the role of dual simplexes in verbs outside the motion category, i.e. among verbs whose simplex (unprefixed) stems do not observe the opposition of determinate / indeterminate. This section will introduce the major types of dual simplex verbs in the non-motion category.

In one group of non-motion dual simplex verbs, (e.g. *пеуу-* / *пеуай-*, *смьну-* / *смьнай-*, *нрѡсму-* / *нрѡцать*), only four of the six potential cells are filled. There is no opposition of two different prefixed perfectives. Any minimal opposition of these two stems is always accompanied by the aspectual opposition of perfective vs. imperfective (e.g. both *пеуу-* vs. *пеуай-* and *ѡмпеуу-* vs. *ѡмпеуай-*), as shown in table 4 (see Garde 1980: 368 and Isačenko 1960: 141 for a list of such verbs).

Table 4. Non-motion dual simplex stems
with constant aspectual opposition and one prefixed perfective

Simplex stems	<i>реши-</i> (perf.)	<i>решай-</i> (imperf.)
Prefixed perfective	<i>отреши-</i>	-----
Derived imperfective	-----	<i>отрешай-</i>

In contrast to the *реши-* / *решай-* type, which always has an aspectual opposition between the two stems, there is another type, in which the aspectual opposition only occurs at the simplex level, and which can have two different perfective forms on the second level of prefixed perfectives, which are then lexically opposed to each other. This type includes such stems as *броси-* / *бросай-* and *хвати-* / *хватай-*, as shown in table 5.

Table 5. Non-motion dual simplex stems with aspectual opposition
only in the simplex form and two prefixed perfectives

Simplex stems	<i>хвати-</i> (perf.)	<i>хватай-</i> (imperf.)
Prefixed perfective	<i>захвати-</i>	<i>захватай-</i>
Derived imperfective	<i>захватывай-</i>	

Certain other verbs are very similar to the pattern of table 5, except for the fact that both simplex stems are imperfective. This type includes such simplex pairs as *вали-* / *валяй-*, *веси-* / *вешай-*, *ломи-* / *ломай-*, *меси-* / *мешай-*, *сади-* / *сажай-*, and is shown in table 6. Thus, we can say that the dual simplex level has either a purely aspectual opposition (as in *хвати-* / *хватай-*) or a lexical opposition (as in *вали-* / *валяй-*).

Table 6. Non-motion dual simplex stems with no aspectual opposition
in the simplex form and two prefixed perfectives.

Simplex stems	<i>вали-</i> (imperf.)	<i>валяй-</i> (imperf.)
Prefixed perfective	<i>отвали-</i>	<i>отваляй-</i>
Derived imperfective	<i>отваливай-</i>	

There is yet another category of this type, in which one of the dual simplexes may be either substandard, dialectal, or non-existent, but which still contains two prefixed perfectives, one of which is derived from the virtual second simplex which does not surface as such in the standard language. Examples of this type include *-скочи-* / *скака-*, *-куси-* / *кусай-*, *-мени-* / *меняй-*, *-стрели-* / *стреляй-*. (The virtual property of one of the simplexes in each pair is indicated by a hyphen which precedes the stem, since it can only occur when preceded by a prefix.) This type is illustrated in table 7.

Table 7. Non-motion virtual dual simplex stems, in which one simplex does not occur as such, but derives a prefixed perfective.

Simplex stems	-стрели-	стреляй- (imperf.)
Prefixed perfective	отстрели-	отстреляй-
Derived imperfective	отстреливай-	

In the type illustrated in table 7, there is no aspectual opposition on level two, where there are two prefixed perfectives. As is the case whenever there are two opposing stems at the level of the prefixed perfective (cf. tables 2, 3, 5, 6, 7), there is a lexical opposition. Among these prefixed perfectives, motion verbs derived from indeterminates seem to share a property with non-motion verbs which are derived from simplexes which use the *-aj-* suffix. They tend to have meanings which refer to actions which are restricted in time, repeated, or quantified, similar to the usual Aktionsart meanings. For example, the perfective *заходи-* (derived from indeterminate *ходи-*) refers to the beginning of the action, and *отстреляй-* refers to a cessation of the action of shooting. The opposed stems (*зайд-* and *отстрели-*) are not so easy to characterize in terms of the typical Aktionsart meanings, defined as ‘drop by’ and ‘shoot an object off of something else’. Since one class of dual simplex stems even lacks an overt realization of one of the simplexes, it would seem that one of the most critical defining features of this class is the binary opposition of two different prefixed perfectives.

Another significant property of dual simplexes with an opposition of two prefixed perfectives is the syncretism of the suffixed imperfective stem which is derived from these two prefixed perfectives. Isačenko thought it exceptional (“Тип *забросить* // *забрасывать*: ожидаемое чередование *с / ш* отсутствует” 1960: 189) that such a perfective stem as *заброси-* does not undergo a consonant mutation of *s > š* in the formation of its derived imperfective stem, *забрасывай-*. I would suggest that this is not an exception, but a rule of Russian, related to the fact that *заброси-* is derived from a dual simplex type, and coexists with the second prefixed perfective *забросай-*. Whenever prefixed perfectives are derived by the prefixation from dual simplex stems, one of the two simplexes functions as the derivational base form for both, for the purposes of forming the derived imperfective, as follows:

1. *-aj-* (or *-a-*) is primary and there is no consonant mutation in either the simplex or the derived imperfective, in the following stems: *бросай-*, *ломай-*, *хватай-*, *кусай-*, *скакай-*, like the motion verbs *катай-* and *таскай-*.

2. *-i-* is primary and there is a mutation both in the *-aj-* simplex and the derived imperfective: *валяй-, меняй-, мешай-, стреляй-, вешай-, сажай-*.

The possible reasons for the invariable syncretism of the derived imperfective and its derivation from one of the two simplexes are intriguing questions and suggest that dual simplexes, which share everything in the stem except the *-i-* or *-aj-* suffixes, stand on the borderline between being a single lexeme and two lexemes, and are thus subject to various types of cross-contamination. In this connection, it might be useful to consider Mel'čuk's distinction between the terms "лексема... ~ вокабула" (1995: 500).

5. Some conclusions and practical considerations.

One can say that all paired instances of dual simplexes and dual prefixed perfectives have a choice of grammatical or lexical opposition. The grammatical opposition of determinate / indeterminate occurs in simplex stems of verbs of motion (*кати-* / *катай-*); the grammatical opposition of aspect occurs in some of the dual simplex pairs of non-motion verbs (*броси-* / *бросай-*, *пеши-* / *пешай-*). However, when two stems (either simplex or prefixed) differ only in their suffix and are **not** opposed by either the grammatical features of determinacy or aspect, we are dealing with a lexical opposition which recalls the difference between non-Aktionsart and Aktionsart meanings (*вали-* / *валяй-*, *ломи-* / *ломай-*, *захвати-* / *захватай-*, *отстрели-* / *отстреляй-*, etc.). Such verbs with dual simplexes and dual prefixed perfectives can lead to a rather complex situation, since a given root with two simplexes can present a different situation for each prefix. A given prefix may form a semantically different prefixed perfective with each of the dual simplexes (e.g. if the prefix *за-* is used with perfective simplex *хвати-*, we obtain *захватить* 'grab', but when used with the imperfective simplex *хватай-*, we obtain *захватать* 'soil as result of grabbing'). Of course, both share the syncretic imperfective *захватывать*. However, another prefix might be used only with either the *-i-* simplex, or the *-aj-* simplex, instead of forming prefixed perfectives with both simplexes. Correspondingly, one or both of the prefixed perfectives may or may not admit the formation of a derived imperfective. Table 8 gives examples of the dual simplex use and shared imperfectives for the 11 verbal roots identified.

Table 8. Examples of dual *simplexes* with two prefixed perfectives and a shared imperfective.

Simplex Pair	Prefixed Perfective Pair	Imperfective
<i>броси- / бросай-</i>	<i>заброси- / забросай-</i> ‘throw far’ vs. ‘cover by throwing’	<i>забрасывай-</i>
<i>вали- / валяй-</i>	<i>развали- / разваляй-</i> ‘ruin, destroy’ vs. ‘spread on a field’	<i>разваливай-</i>
<i>веси- / вешай-</i>	<i>завеси- / завешай-</i> ‘cover with curtains’ vs. ‘hang all over’	<i>завешивай-</i>
<i>ломи- / ломай-</i>	<i>переломи- / переломай-</i> ‘break in two’ vs. ‘break many things’	<i>переламявать</i>
<i>меси- / мешай-</i>	<i>перемеси- / перемешай-</i> ‘knead’ vs. ‘mix’	<i>перемешивать</i>
<i>хвати- / хватай-</i>	<i>захвати- / захватай-</i> ‘grab’ vs. ‘make dirty by grabbing’	<i>захватывать</i>
<i>-скочи- / скака-</i>	<i>заскочи- / заскака-</i> ‘jump across’ vs. ‘gallop far away’	<i>заскакивать</i>
<i>-стрели- / стреляй-</i>	<i>пристрели- / пристреляй-</i> ‘kill by shooting’ vs. ‘test a gun’s aim’	<i>пристреливай-</i>
<i>-куси- / кусай-</i>	<i>перекуси- / перекусай-</i> ‘bite through’ vs. ‘bite many people’	<i>перекусывай-</i>
<i>-мени- / меняй-</i>	<i>обмени- / обменяй-</i> ‘accidentally exchange’ vs. ‘exchange’	<i>обменивай-</i>
<i>-сади- / сажай-</i>	<i>просади- / просажай-</i> ‘spend’ vs. ‘spend time planting’	<i>просаживай-</i>

Table 8 provides a single dual simplex example for each of the eleven non-motion verbs with lexical oppositions in this category. In order to accommodate all of the verbs in table 8, only a single prefix example was selected for each verb. Table 9 is an example of a more comprehensive chart, with examples of dual perfective stems formed with 10 different prefixes, but for a single one of the eleven dual simplex stems, *броси- / бросай-*.

Table 9. БРОС

		Stem-1	Stem-2	Approximate meaning opposition perfective vs. imperfective
Simplexes		<i>броси-</i>	<i>бросай-</i>	
Prefixed perfectives	1	<i>вброси-</i>	<i>вбросай-</i>	'throw in' vs. 'throw several times'
Derived imperfective		<i>вбрасывай-</i>		

Prefix perfectives	2	<i>зброси-</i>	<i>збросай-</i>	‘throw (cast) a long distance’ vs. ‘throw and cover with (mud)’
Derived imperfective		<i>збрасывай</i>		
Prefix perfectives	3	<i>доброси-</i>	<i>добросай-</i>	‘throw up to a specific place’ vs. ‘finish throwing’
Derived imperfective		<i>добрасывай-</i>		
Prefix perfectives	4	<i>наброси-</i>	<i>набросай-</i>	‘throw on (e.g. clothing on one’s shoulders)’ vs. ‘throw many times or in quantity; sketch hastily’
Derived imperfective		<i>набрасывай-</i>		
Prefix perfectives	5	<i>отброси-</i>	<i>отбросай-</i>	‘throw several times (e.g. stones)’ vs. ‘throw off to the side; round off a number’
Derived imperfective		<i>отбрасывай-</i>		
Prefix perfectives	6	<i>переброси-</i>	<i>перебросай-</i>	‘throw everything, one after the other’ vs. ‘throw across’
Derived imperfective		<i>перебрасывай-</i>		
Prefix perfectives	7	<i>приброси-</i>	<i>прибросай-</i>	‘throw several times’ vs. ‘throw something additional, extra’
Derived imperfective		<i>прибрасывай-</i>		
Prefix perfectives	8	<i>проброси-</i>	<i>пробросай-</i>	‘exhaust (e.g. deck of cards) by throwing’ vs. ‘throw across; make a mistake while throwing’
Derived imperfective		<i>пробрасывай-</i>		
Prefix perfectives	9	<i>разброси-</i>	<i>разбросай-</i>	‘carelessly throw several times in a scattered direction’ vs. ‘scatter by throwing’
Derived imperfective		<i>разбрасывай-</i>		
Prefix perfectives	10	<i>сброси-</i>	<i>сбросай-</i>	‘throw several times’ vs. ‘throw off of; throw downwards’
Derived imperfective		<i>сбрасывай-</i>		

This has been a brief survey of the paradigmatic differences between a variety of Russian verbal stems which share the property of having dual simplexes. It has been seen that this stem category shares a number of specific features of morphophonemics, such as the property of stem-final consonant mutation, as well as specific properties of syncretism, particularly of the derived imperfective form. Future work might work towards a

comprehensive listing of such verbs, as well as a pedagogical description of the complexities of the dual simplex class.

References

- Garde, Paul. 1980. *Grammaire russe. Tome 1: Phonologie, Morphologie*. Paris: Institut d'Etudes Slaves.
- Isačenko, A. V. 1960. *Грамматический строй русского языка*. Volume II. Bratislava: Vydavateľstvo slovenskej akadémie vied.
- Mel'čuk, I. A. 1995. *The Russian Language in the Meaning / Text Perspective*. Moscow, Vienna: Школа «Языки русской культуры».
- Timberlake, Alan. 2004. *A Reference Grammar of Russian*. Cambridge: Cambridge University Press.

Взаимодействие категорий переходности и залога в русском языке (Производящие и инициирующие переходные глаголы и их возвратные корреляты)

В. Гаврилова

МГПУ (Московский городской педагогический университет)
v_gavrilova@pochta.ru

Вводятся две взаимообусловленные оппозиции «агенса-исполнителя : агенса-инициатора» и «преобразующая переходность : инициирующая переходность». Показано, что существует соответствие между этими двумя типами переходности, с одной стороны, и пассивным или квазипассивным характером возвратного глагола, с другой стороны. Также заявлено, что русский пассивный залог включает как возвратные формы с пассивным, так и возвратные формы с квазипассивным значением.

Interaction of transitivity and voice categories in Russian (Transformative and initiating transitive verbs and their reflexive correlates). Two mutually conditioned oppositions: “agent-executor : agent-initiator” and “transformative transitivity : initiating transitivity” are introduced. It is shown that there is a correlation between the two types of transitivity on the one hand and the passive and quasi-passive types of reflexive verbs on the other hand. It is also declared that the Russian passive voice includes both reflexive verbs with passive meaning and reflexive verbs with quasi-passive meaning.

Основу исследования составляют поиски адекватной интерпретации особенностей трансформационного поведения абсолютно совпадающих по внешней форме русских переходных конструкций:

(1) Прачка высушила белье	— (1a) Белье высушилось
(2) Повар сварил картошку	— (2a) Картошка сварилась
(3) Кухарка заквасила капусту	— (3a) Капуста заквасилась
(4) Кухарка запекла овощи	— (4a) Овощи запеклись
(5) Прачка погладила белье	— (5a) *Белье погладилось
(6) Повар почистил картошку	— (6a) *Картошка почистилась
(7) Кухарка нарезала капусту	— (7a) *Капуста нарезалась
(8) Кухарка помыла овощи	— (8a) *Овощи помылись

Двучленные конструкции с возвратным глаголом совершенного вида, которые абсолютно безупречны в случае (1a)–(4a), в случае (5a)–(8a) вызывают сомнение и должны быть охарактеризованы как грамматически неправильные. Например, спеллер Word 97 подчеркивает красной чертой глагол в конструкции (5a) и никак не подчеркивает глагол в конструкции (1a).

Ответ на вопрос о том, в чем глубинное смысловое отличие переходных конструкций (1)–(4) от конструкций (5)–(8), совпадающих на поверхностном уровне, лежит в осознании некоторой смысловой общности глаголов-сказуемых в каждой из этих двух противопоставленных по трансформационным характеристикам групп конструкций.

Существуют такие два вида процессов изменения неодушевленных объектов по воле одушевленных субъектов, что между подлежащим, сказуемым и дополнением переходных конструкций, описывающих каждый из этих двух типов процессов, возникают отличающиеся друг от друга субъектно-объектные отношения. А именно: в окружающей нас действительности имеют место два следующих онтологических типа ситуаций, во время развертывания которых осуществляются процессы, изменения неодушевленных предметов и субстанций по воле человека:

Ситуация I. Процессы, которые на всех этапах своего осуществления обусловлены волей человека и нуждаются в его непосредственном участии. Во время осуществления этих процессов между одушевленным агенсом и неодушевленным пациенсом, т. е. **между субъектом и объектом существует пространственная смежность, контакт.**

Изменения некоторого неодушевленного предмета всецело определяются непосредственным антропогенным воздействием. Денотативные ситуации такого рода складываются, например, во время осуществления процессов механической обработки пищевых продуктов, в частности, во время чистки картофеля. Процесс чистки картофеля может происходить только при непосредственном воздействии человека на картофель, поэтому во время осуществления этого процесса картофель и чистящий его человек находятся в контакте. Если изменение производится с помощью некоторого инструмента, то этот контакт осуществляется через посредство этого инструмента. В этой связи напомним, что А. Вежбицкая считает, что универсальную модель «инструментальности» («орудийности») демонстрируют «предложения, обозначающие сознательное действие, которое осуществляется с помощью “механического приспособления”, входящего в контакт и с агенсом, и с пациенсом» [Вежбицкая 1985: 317].

Ситуация II. Процессы, обусловленные волей человека, лишь в том смысле, что благодаря его волеизъявлению они имеют место в жизни. Но специфика этих процессов такова, что человек именно вызывает их к жизни, а не производит, так как во время осуществления этих процессов не требуется непосредственного участия человека. Участие человека в этих процессах является опосредованным и сводится к предварительному продумыванию и последующему осуществлению технологического процесса, обеспечивающего условия, необходимые и достаточные для осуществления задуманных изменений в некоторых телах или субстанциях внешнего видимого мира. Человек инициирует эти технологические процессы, а в ряде случаев предварительно создает необходимую техническую базу для их осуществления, поддерживает их опосредованное энергетическое обеспечение (обеспечивая поступление тепловой, электрической или других видов энергии), а также варьирует параметры осуществляемого технологического процесса. Во время осуществления этих процессов не требуется непосредственного участия человека, **между** одушевленным агенсом и неодушевленным пациенсом, **субъектом и объектом отсутствует пространственная смежность, контакт.**

Изменения некоторого неодушевленного предмета всецело определяются некоторыми не антропогенными внешними факторами (некоторым смещением тел неживой природы, некоторым стечением обстоятельств). Другими словами, для осуществления изменений, соответствующих данной ситуации, существуют необходимые и достаточные условия, всецело состоящие из природных факторов. Произвольная деятельность человека может быть внешним фактором, помогающим сформировать эти необходимые и достаточные условия, но само изменение, как скоро эти условия созданы, происходит самопроизвольно без участия человека. Денотативные ситуации такого рода складываются во время осуществления процессов горячей обработки пищевых продуктов. Так, например, картофель варится или жарится, в результате того, что он оказывается в определенной среде (водной или масляной) и при определенной, достаточно высокой температуре. В эти условия картофель помещает человек, но в самом процессе варки или жарки человек участвует не как исполнитель, а всего лишь как наблюдатель, способный лишь варьировать условия, в которых находится приготавливаемый продукт.

В группе примеров (5)—(8) мы сталкиваемся с описанием ситуации I, а в группе примеров (1)—(4) — с описанием ситуации II. Русские переходные глаголы, описывающие ситуации первого типа предлагаются называть **переходными глаголами преобразующего типа**, или,

коротко, **преобразующими переходными глаголами**, а переходные глаголы, описывающие ситуации второго типа, — **переходными глаголами иницирующего типа**, или, коротко, **иницирующими переходными глаголами**. Мы будем говорить также о двух типах переходных конструкций: о **преобразующих переходных конструкциях** и об **иницирующих переходных конструкциях**, называя так конструкции, в которых в качестве сказуемого выступает, соответственно, переходный глагол преобразующего и иницирующего типов.

Подлежащее конструкции с переходным глаголом производящего типа в качестве сказуемого можно охарактеризовать как **агенса-исполнителя**, а подлежащее конструкции с переходным глаголом иницирующего типа — как **агенса-инициатора**. Проведение между двумя разновидностями агенса (*агенса-исполнителя* и *агенса-инициатора*), с одной стороны, и между двумя типами конструкций с переходными глаголами (*преобразующие переходные конструкции* и *иницирующие переходные конструкции*), — это, по сути дела, две стороны одной и той же медали. Ведь разграничение как двух подвидов агенса, так и двух типов переходности представляет собой расширение языка грамматического описания, основанное на оценке степени энергетического участия лица, обозначенного подлежащим, в процессе воздействия на объект, обозначенный прямым дополнением.

Рассмотрение примеров (1a)–(8a) показывает, что принадлежность собственно переходного глагола к иницирующему или преобразующему типу является **скрытой категорией русской грамматики**, существование которой обнаруживается при попытке построить двучленную конверсивную конструкцию с парным возвратным глаголом совершенного вида. А именно, эта попытка успешно завершается в случае переходных конструкций иницирующего типа и терпит крах в случае переходных конструкций преобразующего типа.

Некоторые исследователи возвратные корреляты переходных глаголов иницирующего типа, в частности, возвратные корреляты переходных глаголов, обозначающих процессы горячего приготовления пищи, трактуют как декаузативные глаголы с удаленным агенсом. Нам такая точка зрения представляется ошибочной.

Необходимо отметить, что процессы горячего приготовления пищи и плавления металлов немислимы без участия в них человека. Собственно эти процессы вошли в жизнь человечества после двух великих открытий — после обретения умения разводить огонь и овладения умением плавления металлов, поскольку, как мы знаем, в истории человечества был и каменный век, во время которого технологией получения металла человечество не владело.

Поэтому в случае рассматриваемых нами глаголов, по нашему мнению, при переходе к возвратным коррелятам происходит не потеря агентивной валентности, а перевод ее в слабую факультативную форму валентности, как и в случае агентивной валентности у возвратных глаголов, являющихся коррелятами переходных глаголов преобразующего типа. Эти агентивные валентности в конструкциях конверсных по отношению к переходным конструкциям преобразующего и иницирующего типов (9а) и (10а) имеют разные семантические интерпретации и разное морфологическое оформление, а в двучленных конструкциях (9б) и (10б) ощущается разница в степени автономности происходящих изменений от агенса соответствующей денотативной ситуации. Но это полностью согласуется с существованием различия в смысловом наполнении роли агенса уже в переходных конструкциях преобразующего и иницирующего типов (9) и (10). При этом в случае описания процесса механической обработки пищевой субстанции, агентс следует охарактеризовать как **агентс-исполнитель**, а в случае горячего приготовления пищи — как **агентс-инициатор**.

- (9) Повар **чистит** картошку
- (10) Повар **варит** картошку
- (9а) Картошка **чистится** поваром
- (10а) У повара **варится** картошка
- (9б) Картошка **чистится**
- (10б) Картошка **варится**

Трехчленная конструкция (9а) с творительным производителем действия хотя и правильна грамматически, но неудовлетворительна с точки зрения узуса, и очень маловероятна в реальных русских текстах. Переходная иницирующая конструкция (10) и трехчленная конструкция (10а) с парным возвратным глаголом, в которой существительное, употребляющееся в конструкции с переходным глаголом в качестве подлежащего, употреблено в форме родительного падежа с предлогом «у», описывают одну и ту же ситуацию действительности¹. Конструкция (10а) абсолютно безупречна не только с

¹ Конструкция *У меня чистится картошка* становится возможной, если имеет место такая денотативная ситуация, когда картошка чистится с помощью, например кухонного комбайна, т. е. в случае, когда агентс приобретает характеристику агенса-инициатора, а конструкция *Повар чистит картошку в кухонном комбайне* является переходной иницирующей конструкцией.

точки зрения грамматической правильности, но и с точки зрения узуальных норм, поскольку, например, русская разговорная речь просто пестрит конструкциями, построенными по этой модели.

Так, в ответ на вопрос о том, почему имярек не осуществляет или недостаточно эффективно осуществляет ту или иную деятельность, связанную с каким-либо предметом, можно в ответ услышать, что причина в том, что эта деятельность не осуществляется в предмете самопроизвольно, например:

- (11) — Почему ты не **читаешь** книгу?
— Она у меня что-то никак **не читается**.
- (12) — Почему ты **не пишешь** письмо?
— Оно никак **не пишется**.
- (13) — Почему ты так долго **жаришь** котлеты?
— Они никак **не прожарятся**.

Список подобных примеров можно расширять сколь угодно долго.

Человечество имеет обширный опыт по осуществлению разного рода воздействий на неодушевленные объекты. Опыт этот по-разному преломляется при отражении его средствами разных языков. Особый интерес представляет в этой связи роль категории залога, в частности, такого частного значения этой категории как «страдательный залог», использование которой обеспечивает словоформе, обозначающей неодушевленную субстанцию, подвергающуюся изменению, возможность занять в предложении центральное место подлежащего.

Это категориальное залоговое значение оказалось очень неподатливым для строгого описания именно на материале русского языка, поскольку русский глагол не имеет специальных морфологических показателей для оформления страдательного залогового значения и потому требует от исследователя работы в области пересечения морфологии, синтаксиса и лексикологии. Несмотря на это, разделы академических русских грамматик, содержащие описание противопоставления действительных и страдательных конструкций, на протяжении всего прошлого века оставались неоправданно короткими, практически не развиваясь за счет привлечения нового материала. В разделах, посвященных залогам, не нашли отражения результаты фундаментальных исследований возвратных русских глаголов Н. А. Янко-Триницкой, а также результаты работ многих школ и отдельных исследователей по вычленению многообразных лексико-семантических групп русских глаголов.

Одной из заслуг Н. А. Янко-Триницкой является разделение возвратных глаголов, имеющих парные переходные глаголы, на глаголы отобъектного и отсубъектного значения. Возвратные глаголы отобъектного значения это по существу глаголы, которые могут быть получены в результате осуществления объектной конверсивной трансформации, которая может быть представлена трансформационной формулой (14). Результирующую конструкцию этой трансформации мы будем называть объектной конверсивной конструкцией:

$$(14) N_{1\text{им}} V N_{2\text{вин}} \leftrightarrow N_{2\text{им}} V\text{-ся}$$

Те сопоставительные ряды действительных и страдательных конструкций, на материале которых строятся теоретические выкладки в разделах, посвященных в академических грамматиках страдательному залогу, укладываются в эту формулу. Однако приводимые примеры весьма немногочисленны и монотонны с точки зрения лексико-семантической принадлежности их глаголов-сказуемых. Поэтому для достижения хоть сколько-нибудь достаточной полноты описания необходимо уметь отвечать на вопрос о принадлежности к страдательным конструкциям в случае объектных конверсивных конструкций с таким лексико-семантическим заполнением глагола-сказуемого, которое отсутствует в иллюстративном материале грамматик.

При этом необходимо отметить, что в иллюстративном материале разделов залог в академических грамматиках подлежащее при переходном глаголе всегда является **агентом-производителем**, а переходные конструкции, в которых лексико-семантическая принадлежность такова, что подлежащее должно быть охарактеризовано как **агента-инициатор**, в грамматиках не представлены.

Общепризнанно, что парадигматические связи словоформ русского глагола характеризуются тесной взаимосвязью и переплетением грамматических категорий переходности, возвратности и залога. В настоящее время описание связи категорий переходности и залога в грамматиках современного русского языка сводится по существу лишь к констатации того факта, что страдательное значение имеют только переходные глаголы, которые теряют свою переходность, принимая значение страдательного залога.

Представление о связи переходности и залога можно существенным образом углубить, поскольку установлено, что принадлежность сказуемого переходной действительной конструкции к преобразующему или инициирующему типам переходности предопределяет морфологические и синтаксические свойства парной объектной конверсивной конструкции следующим образом:

- переходность производящего типа обуславливает пассивный характер возвратного глагола в парной объектной конверсивной конструкции;
- переходность иницирующего типа, обуславливает квазипассивный характер возвратного глагола в парной объектной конверсивной конструкции.

Так, например, переходность иницирующего типа и квазипассивный характер коррелирующих конверсивных объектных конструкций свойственны глаголам, которые С. Д. Кацнельсон называет *фактивными глаголами*. К фактивным глаголам С. Д. Кацнельсон причисляет, например:

- глаголы изменения величины (*наращивать, растягивать, сжимать, уменьшать, увеличивать*);
- глаголы изменения физического состояния (*замораживать, размораживать, согревать, разжигать, сушить*);
- глаголы вызывания процессов (*зажигать, варить, кипятить*).

По нашему мнению, возвратные корреляты переходных глаголов иницирующего типа следует считать формами страдательного залога от этих глаголов действительного залога. Тогда различия между ситуациями I и II можно проинтерпретировать как различия, лежащие в области пересечения категорий переходности и залога.

Необходимо отметить, что в значении глаголов, выражающих преобразующую непосредственную переходность, а также в значении их возвратных коррелятов можно выделить сему волеобусловленности, произвольности. В то время как в значении переходных глаголов, выражающих иницирующую опосредованную переходность, сема произвольности (или свободы волеизлияния) входит в значение лишь переходного глагола и отсутствует в значении его возвратного коррелята. При этом произвольность в значении переходного глагола — это возможность решать, следует ли создавать условия, необходимые для инициации процесса, события или состояния, выражаемого значением основы этого глагола. Само течение иницированного процесса можно контролировать в смысле мониторинга процесса, но контроль как управление процессом возможен далеко не всегда, поскольку далеко не всегда иницированный человеком процесс может быть прерван по его воле. Техногенные катастрофы — это как раз те случаи, когда процесс развертывания преобразований материальных субстанций, неизбежно осуществляющихся по законам природы, вызывает нежелательные последствия, но инициатор этого процесса бессилен остановить дальнейшее развертывание этого процесса.

Отношения семантической деривации в парах «переходный глагол ↔ возвратный глагол» имеют разную направленность в зависимости от того, является ли переходный глагол в этой паре переходным глаголом преобразующего типа или переходным глаголом иницирующего типа.

У переходных глаголов преобразующего типа семантически более простым является переходный глагол, а возвратный глагол означает «подвергаться действию, обозначенному переходным глаголом», например:

значение («чиститься») = подвергаться действию «чистить»;
значение («строиться») = подвергаться действию «строить».

В то время как у переходных глаголов иницирующего типа семантически более простым является возвратный глагол, а переходный глагол означает «создавать условия для того, чтобы осуществился процесс, обозначенный возвратным глаголом», например:

значение («варить») = создавать условия для того, чтобы «варилось»;
значение («греть») = создавать условия для того, чтобы «грелось»;
значение («разбивать») = создавать условия для того, чтобы «разбивалось».

Например, с помощью переходного глагола иницирующего типа деструктивной семантики **разбить** могут быть описаны денотативные ситуации, которые сильно разнятся по своему содержанию, но их объединяет то, что каждая ситуация обеспечивает возникновение удара, приводящего к тому, что хрупкий по природе предмет, например, стакан, рассыпается на части. При этом обеспечение возникновения такого удара может быть как спонтанным нецеленаправленным (случайно осуществившийся удар), так и преднамеренным целенаправленным (намеренно предпринятый удар), а способы нанесения удара могут быть самыми различными. Важно лишь одно: имевшее место действие обеспечило удар, вследствие которого стакан разбился. Поэтому независимо от способа нанесения удара и степени его преднамеренности, если искомый удар имел место, мы скажем:

(15) *Х разбил стакан.*

Аналогичным образом глагол **греть** в словосочетании **греть воду** может обозначать разнообразные денотативные ситуации: 1) сосуд с водой может быть помещен над костром; 2) поставлен на солнцепек;

3) помещен на плиту, в печь; 4) включен в электросеть (если вмещающим воду сосудом является электрочайник). Во всех этих многообразных ситуациях мы скажем:

(16) *X греет воду.*

И, наоборот, одно и то же действие, состоящее, например, в том, что человек сел на стул, не заметив лежавшего на этом стуле предмета, будет описано совершенно разными глаголами в зависимости от того, к каким изменениям это действие привело оказавшийся на стуле предмет. Качества же этих изменений зависят от свойств того предмета, на который человек случайно сел. Поэтому это действие может быть описано в зависимости от свойств предмета, например, следующими предложениями:

(17) *X сплющил коробку;*

(18) *X смял платье;*

(19) *X раздавил елочную игрушку.*

В случае с переходными глаголами иницирующего типа и парными возвратными глаголами мы сталкиваемся примером обратного словообразования, когда исходным (более простым) с семантической точки зрения глаголом является один глагол пары, а морфологически исходным (более простым) — другой глагол этой же пары. Этот словообразовательный тип для русского языка был предложен И. А. Мельчуком для описания глаголов *радовать-радоваться* [Мельчук 1967]. О рецессивных конструкциях с «обратным смысловым соотношением», когда форма глагола без *-ся* толкуется чаще всего на основе фактивной деривации из формы на *-ся*, говорит и Ж. Веренк [Веренк 1985: 296—297]. В типологическом исследовании М. Хаспельмата рассматриваются направления семантической деривации глаголов с каузативным и декаузативным значением, представленные в различных подгруппах таких глаголов в языках различных типов [Haspelmath 1993: 103—106].

Как мы уже говорили, выбор между пассивным и квазипассивным конверсивом определяется тем, какой — преобразующей или иницирующей, соответственно, — является переходность глагола в исходной конструкции. Точнее, тип переходности глагола-сказуемого в исходной конструкции определяет прототипическое значение формы глагола в конверсивной конструкции. Прототипическим значением мы будем называть значение глагола-сказуемого, проявляющееся в двухчленной объектной конверсивной конструкции, т.е. значение, которое свойственно возвратному глаголу-сказу-

емому при отсутствии влияния какого-либо контекста. Преобразующий тип переходности глагола-сказуемого в исходной конструкции определяет прототипическое пассивное значение возвратной формы глагола в конверсивной конструкции, но в качестве периферийного, обусловленного контекстом, значения этот возвратный глагол может иметь и квазипассивное значение. И наоборот, иницирующий тип переходности глагола-сказуемого в исходной конструкции определяет прототипическое квазипассивное значение формы глагола, но в качестве периферийного, обусловленного контекстом, значения этот возвратный глагол может иметь и пассивное значение ².

В паре глаголов *строить* / *строиться* первый глагол следует охарактеризовать как переходный глагол преобразующего типа, а второй — как возвратный глагол с прототипическим пассивным значением. Поэтому переходный глагол считается исходным. Этот исходный глагол в толковых словарях снабжают самостоятельным толкованием, и это толкование с соблюдением ряда условий переносится на парный возвратный глагол. Так, например, на равенство значений переходного глагола преобразующего типа *строить* и прототипического пассивного значения глагола *строиться* указывает Ю. Д. Апресян [Апресян 1995: 263].

(20) *В строится X-ом = X строит В*

Составители толковых словарей в случае глаголов с пассивным прототипическим конверсивным значением, таких как глаголы *строить* / *строиться*, дают самостоятельное толкование лишь переходному глаголу, а при возвратном глаголе дают помету «*страд.* от парного переходного глагола».

В то же время возвратные глаголы, для которых прототипическим конверсивным является квазипассивное значение, как правило, имеют в толковых словарях самостоятельную словарную статью и толкования значений квазипассивного возвратного глагола и парного переходного глагола иницирующего типа. Примерами возвратных глаголов с прототипическим квазипассивным значением могут служить глаголы следующих лексико-семантических групп (ЛСГ):

² Так в газетном заголовке у глагола *разрушаться* с прототипическим квазипассивным значением под влиянием контекста, а именно, вопросительного местоимения *где*, развивается периферическое пассивное значение: *Если браки заключаются на небесах, то где они разрушаются?* А в примере И. Г. Милославского *Сама собою жизнь ведь не построятся, Вода под камушек не потечет* присутствует периферийное квазипассивное значение глагола *построиться*. См. также сноску 1.

«глаголы со значением горячего приготовления пищи», «глаголы значением деструктивных изменений» и «глаголы со значением нейтральных физико-химических изменений». Мы проанализировали формы представления значений возвратных глаголов этих ЛСГ и парных переходных глаголов в словарях [Ожегов 1972; Ушаков 1935—1940; Тихонов 2001]. Анализ показал, что значения такого рода возвратного и парного переходного глаголов совпадают в основной своей части, но словарные статьи переходного и возвратного глаголов никак между собой не сопряжены. При этом сдвиг, происходящий при переходе от возвратного глагола к парному переходному глаголу, можно описать без обращения к конкретной семантике отдельных глаголов. Представляется, что если описать этот сдвиг в грамматиках, то определение в толковых словарях можно будет давать только возвратным глаголам, а значение переходных глаголов можно давать в виде ссылки «*действ. залог* от соответствующего возвратного глагола»

Обобщая все вышесказанное, отметим, что существует бинарный дифференциальный признак, по которому русские действительные переходные конструкции могут быть разделены на такие два типа: **«преобразующая (непосредственная) переходность»** и **«иницирующая (опосредованная) переходность»**. Дифференциальный признак, по которому эти типы переходности выделяются, имеет онтологический характер и может быть охарактеризован как «степень участия агенса ситуации в процессе воздействия на пациенс ситуации».

В настоящее время, по крайней мере, в русистике субъектно-объектные отношения, возникающие между подлежащими и прямыми дополнениями собственно переходных глаголов, признаются одинаковыми на всем множестве собственно переходных глаголов. Другими словами, множество переходных глаголов, описывающих целенаправленное воздействие одушевленного агенса на неодушевленный пациенс, воспринимается как некое единое монолитное с точки зрения выражаемых на нем субъектно-объектных отношений пространство. Однако бытийные ситуации, которые описывают глаголы с преобразующей и инициирующей переходностью, отличаются друг от друга разным распределением агентивности между субъектом и объектом ситуации.

В бытийных ситуациях, описываемых преобразующими переходными глаголами, неодушевленный объект является лишь местом приложения сил одушевленного агенса. Объекту онтологических ситуаций такой специфики чужда даже малейшая доля агентивности. Субъект в полной мере наделен агентивностью.

В то время как в бытийных ситуациях, описываемых иницирующими переходными глаголами, неодушевленный объект является не местом приложения сил одушевленного агенса, а ареной, на которой при некоторых условиях способны проявиться некоторые имманентно присущие неодушевленному объекту свойства, например его физические или химические свойства. Объекту онтологических ситуаций такой специфики присуща спонтанная агентивность, агентивность, которая не может не проявиться, как скоро объект попадает в некоторые условия, определяемые семантикой соответствующего иницирующего переходного глагола. Субъект не в полной мере наделен агентивностью, ему присуща агентивность, направленная на создание необходимых и достаточных условий для проявления спонтанной агентивности объекта.

Учитывая вышесказанное, мы вправе заявить, что в онтологических ситуациях, описываемых глаголами с преобразующим и иницирующим типами переходности, мы сталкиваемся с двумя разными логиками субъектно-объектных отношений. Другими словами, мы вправе говорить о субъектно-объектной логике преобразующего типа и о субъектно-объектной логике иницирующего типа.

Нам представляется, что субъектно-объектная логика преобразующего типа определяет устройства глагольной системы в номинативных языках, а субъектно-объектная логика иницирующего типа в языках номинативного строя оказывается подавленной. И, наоборот, субъектно-объектная логика иницирующего типа определяет устройства глагольной системы в эргативных языках, а субъектно-объектная логика преобразующего типа в языках номинативного строя оказывается подавленной.

Как нам представляется, совокупность преобразующей и иницирующей логик субъектно-объектных отношений является базисной матрицей, базисной структурой мышления. Поэтому проявления как преобразующей, так и иницирующей субъектно-объектной логики не могут быть полностью искоренены ни в одном естественном языке. Именно этим объясняется наличие номинативных конструкций в эргативных языках и наличие элементов эргативности в номинативных языках.

На обсуждение выносятся следующие гипотезы и проблемы:

1. Дают ли различия субъектно-объектных отношений в случае переходности производящего и иницирующего типов возможность говорить о субъектно-объектных отношениях производящего и иницирующего типов?

2. Носит ли понятийная категория, стоящая за разграничением производящего и иницирующего типов переходности, универсаль-

ный характер и каковы способы выражения этой понятийной категории в конкретных языках?

3. Существует ли содержательная связь между иницирующей переходностью, рассматриваемой как онтологическая категория и феноменом лабильности глаголов?

Литература

- Апресян 1995 — *Апресян Ю. Д.* Лексические конверсивы // *Апресян Ю. Д.* Лексическая семантика. Синонимические средства языка. М., 1995. С. 256—283.
- Вежбицкая 1985 — *Вежбицкая А.* Дело о поверхностном падеже // Новое в зарубежной лингвистике. Вып. XV. Современная зарубежная русистика. М., 1985. С. 303—341.
- Веренк 1985 — *Веренк Ж.* Диатеза и конструкция с глаголами на -ся // Новое в зарубежной лингвистике. Вып. XV. Современная зарубежная русистика. М., 1985. С. 286—302.
- Мельчук 1967 — *Мельчук И. А.* К понятию словообразования // Изв. АН СССР. Сер. лит-ры и языка. Вып. 26. 1967. № 4. С. 352—362.
- Янко-Триницкая 1962 — *Янко-Триницкая Н. А.* Возвратные глаголы в современном русском языке. М.: Изд-во АН СССР, 1962.
- Haspelmath 1993 — *Haspelmath M.* More on typology of inchoative/causative verb alternations // *Causatives and Transitivity*. B. Comrie, V. Polinski (eds.). Amsterdam / Philadelphia: Benjamins, 1993. P. 87—120.

On the Descriptive Adequacy of Topology

Kim Gerdes Sylvain Kahane Hi-Yon Yoo

ERSS, Modyco, ARP,
Université Bordeaux 3 Université Paris X Université Paris 7

This article attempts to sum up the history of the topological structure in MTT and to analyze its prospects in the future development of the theory. We discuss the *raison d'être* for considering the existence of the topological constituent tree as well as the constraints on its form. We discuss in particular the necessary information for the computation of the topology-phonology interface.

1. Linearization with and without topology

The history of topology can be traced back to the 19th century descriptions of German sentence structure, where this notion is particularly useful for the modeling of the German(ic) V2 constraint (i.e. exactly one constituent of any function before the finite verb) (Erdmann 1886, Drach 1937). The basic quite trivial idea is to suppose the existence of templates (domains) of places (often called “fields”), which exist independently of a particular syntactic function and can be filled based both on syntactic and communicative rules. This allows a straightforward description of the V2 phenomenon and any learner of German hears somehow about the existence of these fields. In a sense, the topological description can be seen as a first step in the separation of syntax and word order constitutive to dependency approaches ever since Tesnière 1959.

1.1. Alternatives to topology

It is clear that GB based descriptions have great difficulties controlling the unique element before the verb (cf. the discussion on IP vs. CP occupation of the finite verb). Similarly, dependency linearization rules based exclusively on precedence rules among governors and its dependents (or among dependents of the same governor), as it was commonly proposed and implemented (Mel'čuk & Pertsov 1987, Nasr 1995, Kahane 2001),

have the exact same expressive power as context free grammars and can thus never handle any long distance phenomena in a satisfying manner (Gerdes & Kahane to appear).

Mel'čuk's model (1988) has already a hierarchical constituent structure on the morphological level, the prosodic markup, very similar to the resulting topological structure¹. However no details concerning this prosodic markup are put forward, let alone precise criteria of what it should contain and how it is computed from the syntactic representation.

In order to be understood beyond the borders of MTT, "topology" is a more adequate name than "morphology" for the (linearized) level following the dependency representation because the structure and the rules handle interword relations and not only intraword structures. (With the habitual meaning of "morphology", it is odd to consider prosodic word groups on a "morphological" level.)

classical MTT (Mel'čuk 1988)	model (following Gerdes & Kahane 2001 and Yoo 2003)
Surface syntax	Surface syntax
Deep morphology (with prosodic grouping and contours)	Morpho-topology (constituent tree and morphological structure: leaves are morphemes)
Surface morphology	Phonology (prosodic grouping and contours)
Deep phonology	Phonetics
Surface phonology (text)	Sound

1.2. Why topology is useful for linearization

It has been shown on numerous occasions that topological linearization grammars can handle very difficult word order constraints in an elegant and simple manner:

- V2, verb clusters and verbal pied piping of German (Debusmann & Duchier 2001, Gerdes & Kahane 2001)
- final verb cluster of Korean (Gerdes & Yoo 2004)
- extraction constraints from phrasal complements in Modern Greek (Yoo 2003b)
- long distance subject inversion and clitic climbing in French (Gerdes & Kahane 2004)

¹ More precisely, in Mel'čuk 1988, the prosodic markup consists of a set of delimiters in various levels. This leveling induces a hierarchy of constituents (constituents with level 1 delimiters are included in constituents with level 2 delimiters and so on).

- dislocation from nominal phrases of Arabic (El Kassad & Kahane 2004).

The topological approach has also proven to be a useful simplification for the order rules of other formal frameworks as HPSG (Kathol 1995) and LFG (Clément *et al.* 2002, Frank 2003)

We consider that the existence of a hierarchical constituent structure in the linearization process of a dependency tree is incontestable. The question that remains to be discussed is the following: Is a topological tree just the technical result of the linearization process or is it a linguistically motivated independent structure? If it is an independent structure, what should it represent?

We will proceed by developing two points: The explanatory power of the topological structure for certain word order constraints. This will be done by showing how different order constraints combine to form the topological structure. Secondly, we will show that the topological structure is necessary for the computation of the phonological structure (with its prosodic groupings).

2. The topological structure

When describing word order topologically, we state the linear order of the fields composing a domain (see Mel'čuk 1967 for a similar idea). As the same domains can occur recursively in embedded sentences, it is somehow natural, though rarely explicitly stated, that we obtain a hierarchy of domains. The resulting structure, called a topological structure (Gerdes & Kahane 2001), is an ordered constituent structure, a fact that scares away hard core dependency advocates. It is important to understand though that the topological structure is not a syntactic structure for it does not express any syntactic function of any of its arguments, contrarily to X-bar approaches where the same structure jumbles functional information and constituent structure.

The original formalization by Gerdes & Kahane 2001 does not include communicative information in the linearization grammar, resulting in a quite high number of possible word orders for a given dependency tree. For even “freer” word order language like Russian and Modern Greek (“free” in the sense that order depends only little on the category or the syntactic function of the words), a grammar that does not take into account communicative information overgenerates massively. Yoo 2003b has extended the formalization to include into the linearization grammar a simple cutting up of the dependency tree in communicative zones, a sim-

plification of the communicative zones of Mel'cuk 2001. Such a grammar is called a "communicative topological grammar". This grammar results in domains (sequences of fields) that put together into one structure these two quite independent constraints. The status of the resulting topological structure is not clear cut. In Greek, for example, you can have fields for complementizing elements (morpho-syntactic constraint) next to fields reserved for prominent theme elements (communicative constraint).

2.1. German

A communicative linearization grammar for German (see Gerdes 2002) has to include the usual topic-focus-tail division (see Buring 1997, Vallduvi 1992) in addition to the syntactic constraints (e.g. V2: finite verb of declarative sentences in second position). Instead of the famous unique *main domain* (*Vorfeld*, *left bracket*, *Mittelfeld*, *right bracket*, *Nachfeld*)², the separation between the communicative units can fall anywhere in the main domain, in particular into the *Mittelfeld*, which brings upon a list of different main domains having the *Mittelfeld* being cut off differently between the communicative parts. In such a description, it might seem as if we lost the intuitive (and classical) character of the topological description and we ended up in a two step process of explanation: We have different order constraints (the five fields from the syntactic constraints and three independent fields from the communicative constraints) and these constraints are subsequently overlaid to produce a certain number of combined domains. Only the latter are used in the actual topological grammar.

Simple communicative constraints for German:

[Theme (prominent, or "topic" (T)) < Rheme (or "Focus" (F)) < Theme (non prominent, or "tail" (t))]

We can add the common operators from regular expressions to the fields in order to express the fact that we need exactly one focus element and an arbitrary number of theme and tail elements: [T* F! t*].

Syntactic constraints for German:

[Vorfeld (vf)! < left bracket (lb)! < Mittelfeld (mf)* < right bracket (rb)? < Nachfeld (nf)*]

The left bracket accepts the finite verb or the complementizer, the right bracket hosts the verb cluster.

² *Vorfeld* ('prefield'), *Mittelfeld* (middlefield), and *Nachfeld* (postfield) are the classical German terms in the topological description.

When superposing the communicative and the syntactic constraints, we end up with different combined main domains that allow various orders depending on the communicative division of the sentence. In the context of (Q1) for example, the focal part will have to end up completely in the Mittelfeld of the combined resulting main domain shown below (which will be excluded if the finite verb is part of the focus, because the finite verb has to rejoin the left bracket — see second example below).

syntactic constraints:	vf	lb	mf			rb	nf
communicative constraints:	T			F	t		
one possible resulting domain:	vf-T	lb-T	mf-T	mf-F	mf-t	rb-t	nf-t

(Q1) Wann hat Peter die Erdbeeren den Goldfischen gefüttert?

When did Peter feed the goldfish with strawberries?

	vf-T	lb-T	mf-T	mf-F	mf-t	rb-t	nf-t
(A1)	Peter	hat	die Erdbeeren	schon gestern	den Goldfischen	gefüttert.	

Peter (nom.) has the strawberries (acc.) already yesterday the goldfish (dat.) fed?

Peter fed the goldfish with strawberries yesterday already.

Another combination for a different communicative division without topical part (possible since the topic is not obligatory):

syntactic constraints:		vf	lb	mf	rb	nf
communicative constraints:	T	F				t
another possible resulting domain:		vf-F	lb-F	mf-F	rb-F	nf-t

(Q2) Und essen Goldfische denn eigentlich Erdbeeren?

And do goldfish really eat strawberries?

	vf-F	lb-F	mf-F	rb-F	nf-t		
(A2)	Ich	war	lange davon	überzeugt,	dass Goldfische Erdbeeren essen.		

I (nom.) was for_long of_that convinced that goldfish (nom.) strawberries (acc.) eat.

I was convinced for long that goldfish eat strawberries.

All syntactic fields can superpose any communicative field, the only constraint being the existence of a focal part. This implies that tail and Vorfeld are incompatible.

We see how the two easy order constraints (syntactic and communicative) combine to form the quite complex main domains for German.

2.2. French

Let us turn to a similar phenomenon in French. Contrarily to German, it seems that the communicative structure has only a little role to play in French. The verbal domain is rich and contains numerous fields in order to insure the correct placement of clitics (fields having the *X-cl* tag) as well as the possibility of left or right dislocation (the *prefix* and *postfix* fields). The *Vorfeld* is, as in German, open to a unique element, usually the subject, whereas the *Mittelfeld* receives any number of complements. The *comp* field receives the possible complementizer. The finite verb is placed in the *verb* field whereas any bare predicative element (participle, infinitive, predicative noun or adjective) possibly separated by an adverb from its head go to the *pred* field.

Thus, the syntax of the finite verb in French consists of the following template of fields which accommodate the verb's syntactic dependents:

verbal domain: *comp* < *prefix* < *Vorfeld* < *neg-cl* < *me-cl* < *acc-cl* < *dat-cl* < *y-cl* < *en-cl* < *verb* < *subj-encl* < *adv* < *pred* < *Mittelfeld* < *postfix*

Just like in German, this “flat” structure results from the overlaying of a second set of linear constraints whose linguistic motivation is more easily accessible and which shows that French is actually quite similar to German. This second set of linear constraints organizes the verbal domain in four templates embedded one into the other: macro-domain, kernel, cluster, and chunk.

The (verbal) **chunk** presents a very rigid order and does not allow any inset clause. It receives the verb and the clitics and must form a prosodic word. The **cluster** contains the chunk: It can carry other light elements (we use the term introduced by Abeillé & Godard 2000) such as bare predicative elements, some specific adverbs and “floating” quantifiers. The **kernel** receives the other verbal dependants, especially all the elements that saturate the verbal valency (including the verbal chunk). Finally, we obtain the complete structure with the macro-domain which can be compared to the macrosyntactic level of Blanche-benveniste 1990: In addition to the kernel which represents the “proper” syntactic level, the macro-domain contains the *prefix* and the *postfix* fields for detached elements with specific communicative values, which do not have a direct but proleptic syntactic link with the verb. The complete verbal constraint set is as follows:

macro-domain: *comp* < *prefix* < *kernel* < *postfix*
 kernel: *Vorfeld* < *cluster* < *Mittelfeld*

cluster: *chunk* < *adv* < *pred*
 chunk: *neg-cl* < *me-cl* < *acc-cl* < *dat-cl* < *y-cl* < *en-cl* <
verb < *subj-encl*

We could say that French resembles somehow to a V2 language since its *Vorfeld* is, as in German, open to a unique element. The element that goes in this field is usually the subject but other complements can occupy this place too as shows the case of (e.g. locative) inversion:

Sur la place | se dresse | une cathédrale. ('On the square stands a cathedral.')

Au terrorisme | ne pouvait pas ne pas répondre | l'activisme israélien
 (Le Monde, 1967).

However, when the subject is a clitic pronoun, it has to occupy the *Vorfeld*, but this field is then "aspirated" into the verbal chunk, the obligatory template for clitics. When, in this case, a locative complement is fronted it is necessarily dislocated:

Sur la place, elle se dresse. ('On the square it stands.')

A comparable phenomenon of "aspiration" appears when the complementizer is also a verbal dependent: The left dislocation is possible in French completives and ungrammatical in relatives:

Je crois que, Marie, Jean ne lui donne plus d'argent. ('I think that, Mary, John doesn't give her money')

**l'argent que, Marie, Jean lui donne* ('the money that, Mary, John gives her')

l'argent que Jean donne à Marie ('the money that John gives to Mary').

Our double system of constraints accounts perfectly for this ungrammaticality: The relative pronoun saturates the valence of the verb and must go into the kernel (as any other verbal argument would). But since it occupies the *comp* field of the macro-domain, there can no longer be a *prefix* field (because the *prefix* field disappears when the *comp* field is aspirated in the kernel)!

Moreover, the aspiration can even cause a fusion with the *Vorfeld*, the first field of the kernel, which explains the optional subject inversion in relatives:

le livre que Pierre lit; le livre que lit Pierre. ('the book that Peter reads').

This fusion is obligatory for the placement of the subject relative pronoun *qui*:

**la cathédrale qui sur la place se dresse.* ('the cathedral that on the square stands').

The same process goes even further with the interrogative pronoun *que* which is a clitic: In this case, the *comp* field is aspirated all the way into the verbal chunk causing on its way the disappearance even of the *Vorfeld*:

Que veut-il faire ? ('What does he want to do ?')

**Que Pierre veut-il faire ?* ('What does Peter want to do ?')

A qui Pierre veut-il parler ? ('To whom does Peter want to talk ?').

Of course, chunk, cluster, and kernel also play other roles in the description. We find the same type of constraints in German where non-stressed pronouns of the *Mittelfeld* must be close to the verb while adverbs precede the other complements.

It is imaginable to apply the topological constraints without preserving the combined topological structure: Either we just keep the word string or we keep parallel topological structures for each type of order constraints (syntactic, communicative, rhythmic... constraints). (One could suppose that syntactic and communicative constraints act independently, constructing two different families of constituent trees. It could be argued that the resulting acceptable linearizations are those appearing in both tree families.) In the remaining of this paper we show that these alternatives make the computation of the next representation in the generation process more complex.

3. From topology to phonology

It holds for any representation level in the MTT framework that the justification for the independence of the level stems from the ease in describing the correspondences with the neighboring representations. Or, to put it differently, MTT presupposes the existence of 7 levels in its language model, because the breaking up of the meaning-sound correspondence makes the description of language easier. Thus, we have to show that the topological tree is not a mere byproduct of the linearization process: We have to show the usefulness of the topological tree structure in the following correspondence, in the topology-phonology interface.

We discuss three examples indicative of the necessity of the topological structure and the insufficiency of pure word strings for the computation of the prosodic structure (and thus of the sound output):

3.1. German

The verbal domain for the so-called “Verlaufsform” (partial nominalization of an infinitive) obliges a contraction of the preposition and the determiner whereas a nominal domain allows also not to contract (and communicative constraints apply to choose among the two possibilities):

*Maria ist am (*an dem) Schlafen.* (‘Maria is sleeping.’)

Maria sitzt [am / an dem] Tisch. (‘Maria is sitting at the/that table’).

Another possibility to handle this problem is to consider the two cases of “dem” as different words, as in English where the definite and demonstrative articles differ. For German this approach results in an ambiguity for every single article. Moreover, we would have to keep morpho-syntactic information in the linearized representation. In a topological approach we can abstract away from this and even the distinction between article and relative pronoun is not needed for the computation of the sound output: The topological domain structure allows us to recognize that the “an dem” string in the following sentence cannot be contracted as the words belong to different domains: *Maria gab das Script an [dem die Daten zu übergeben sind]*. (‘Maria specifies the script [to which the data has to be passed on]’)

3.2. French

Similarly, the French flat topological structure is sufficient for the computation of the correspondence with the phonological representation. Even though the verb cluster (“amas verbal”) is a more constraint domain concerning for example extraction and insertion of parenthetical clauses, the border of this domain is neither essential for the computation of prosodic groups nor for the placement of parentheticals (e.g. parentheticals can be placed at different places, among them before the verbal dependant inside the verb clusters: *Il a voulu vraiment, c’est Jean qui me l’a dit, acheter cet appartement* ‘He really wanted, John told me that, to buy this flat.’).

The phenomenon of liaison (and elision) in French can illustrate the fact that the internal structure of topological constituents are necessary for the computation of the next level. The following example illustrates that some positions are obligatory [z], optional ([z]), or forbidden # for liaison:

Les petits [z] enfants ([z]) intelligents # étaient malades. ‘The little intelligent children were sick.’

We stipulate that liaison can only take place inside a constituent (and not via constituent borders). The optional character of the liaison between

enfants and *intelligents* stems from the fact that the flat nominal domain has been obtained from a simplification of an embedded structure where determiner, pre-posed adjective, and noun formed a nominal chunk, thus excluding *intelligents*. This explains the topology—phonology rule that allows us not to realize a liaison between the noun field and the post-posed adjective field. Without the topological structure, we would need to stipulate ad-hoc rules of the existence of liaison between certain categories of words.

Moreover, the topological structure is necessary to explain different levels of prosodic breaks as in the following example (more important breaks are indicated by higher number of bars):

Le départ | de Paris || de Marie ||| est repoussé. ('Mary's departure from Paris is put off.').

These borders follow naturally from the topological constituent structure:

[Le départ [de Paris] [de Marie]] est repoussé.

It is easy to see that the first border (*depart — de*) is topologically less important than the second border (*Paris — de*). The higher importance of the last border (*Marie — est*) stems from the fact that we are leaving all nominal domains to continue in the (verbal) main domains³.

3.3. Greek

Greek clitics do not bear lexical stress. In some special positions, however, we can assign them secondary stress if they are in an enclitic position. The example below is ambiguous two groupings depending on whether the clitic *mou* 'my/me' is part of the nominal domain (*[i Despina mou] to edose*) or of the verbal domain (*[i Despina] mou to edose*). If it is part of the nominal domain, then it will obtain a secondary stress⁴, oth-

³ This latter border is even more enhanced by the fact that in the original constraint sets that constructed the (flattened out) domain, we had the *chunk* and *cluster* opening at this position:

[[Le départ] [de Paris] [de Marie]] [[est]_{chunk} repoussé]_{cluster}

This shows again how the original constraints explain some topology—phonology correspondence rules.

⁴ Following the trissyllable rule, we can assign a secondary stress on unstressed syllables when the primary stress is followed by three unstressed syllables. In our example, the noun *Despina* is stressed on the antepenultimate syllable.

erwise it will not. On a suprasegmental level, the decision of putting the corresponding contours will depend on the position of the clitic in the topological structures.

i Despina mou to edose
 det Despina my/me it gave
 ‘Despina gave it to me’ or ‘My Despina gave it’

We see that the topological structure, and not just the word order, is crucial for the correspondence with the phonological level. Without the topological structure, it would only be possible to obtain the correct grouping under two conditions:

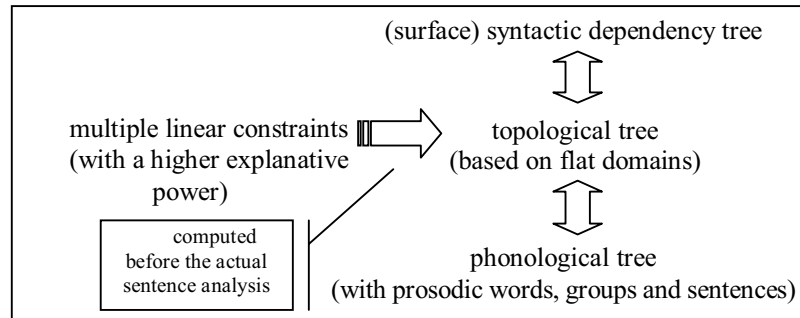
1. The prosodic grouping is assigned before the words are constructed (which has been shown to be difficult in Yoo 2003a);
2. We need quite complex *ad hoc* rules for assigning the frontiers that take into account categorical and syntactic information.

Conclusion

Looking at these results for various languages, we can conclude that topology is more than just a computational tool among others to compute word order, but it is a self-sufficient linguistic structure. It is clear that we have to distinguish the explanatory part from the computational part of topology. When explaining word order or phonological constraints, we sometimes refer to simpler linear constraints. However, these domains do not appear in the topological structure as they are, but they are used to construct the actual domains of the topological structure. The absence of the “explanatory” constraints from the topological structure can be justified by the fact that they have no direct relevance for the computation neither of the syntax-topology interface nor of the topology-phonology interface.

The (explanatory) constraints of different domains are overlaid to form combined flat domains when constructing the grammar. The following diagram may recall the base generated deep structure and the corresponding surface structures of Government and Binding Theories, but our language model is completely different in that the actual computation of the Meaning-Text correspondence for a given sentence does not include the multiple constraint domains but only the combined flat structures.

ble: if the clitic is part of the noun domain, it can get a secondary stress while when it is part of the verbal domain, the trissyllable rule does not apply.



The topological structure is necessary for a complete description of word order phenomena independently of the language to describe and independently of the model chosen for the formalisation. A topological grammar takes into account syntactic and communicative order constraints and allows in this way a typological classification of languages. Moreover, we have shown that topology is not only needed for the linearization process of the dependency tree but also for the next step, the computation of prosodic groups. However the actual domains (and thus the syntax-topology and topology-phonology correspondence rules) are often the result of the combination of various order constraints with a higher explanatory power than the resulting domains themselves. Ongoing work on various languages and phenomena tends to verify the descriptive power of the topological formalization.

References

- Abeillé A. & Godard D., 2000, "French word order and lexical weight", in R. Borsley (ed.), *The nature and function of Syntactic Categories*, *Syntax and Semantics* 32, Academic Press, New York, 2000. P. 325—358.
- Blanche-Benveniste C., 1990, *Le français parlé: études grammaticales*, avec M. Bilger, C. Rouget et K. van den Eynde, Editions du CNRS, Paris, 1990.
- Büring, D., 1997, *The 59th street bridge accent: the meaning of topic and focus*, London, Routledge, 1997.
- Clément L. et al., 2002, "An LFG-type Grammar for German Based on the topological Model", in *Proceedings of LFG'02 Conference*, CSLI Publications, 2002.
- Debusmann R. & D. Duchier, 2001, "Topological Dependency Trees: A Constraint-Based Account of Linear Precedence", *39th Annual Meeting of the Association for Computational Linguistics ACL*, Toulouse, 2001.
- Drach E., 1937, *Grundgedanken der deutschen Satzlehre*, Diesterweg, Frankfurt, 1937.

- El Kassas D. & Kahane, 2004, "Modélisation de l'ordre des mots en arabe standard", *Atelier sur le Traitement Automatique de la Langue Arabe Ecrite et Parlée, JEP-TALN*, Fès, 2004. P. 259—264.
- Erdmann O., 1886, *Grundzüge der deutschen Syntax nach ihrer geschichtlichen Entwicklung dargestellt*, Cotta, Stuttgart, 1886.
- Frank A., 2003, "Projecting LFG F-structures from Chunks", in *Proceedings of the LFG'03 Conference*, CSLI Publications, 2003.
- Gerdes K., 2002, *Topologie et grammaires formelles de l'allemand*, PhD. thesis, Université Paris 7, 2002.
- Gerdes K. & S. Kahane, 2001, "Word Order in German: A Formal Dependency Grammar Using a Topological Hierarchy", *Proceedings of ACL*, Toulouse, 2001.
- Gerdes K. & S. Kahane, 2004, "L'amas verbal au cœur d'une modélisation topologique du français", in *Preproceedings of the Workshop 'Ordre des mots dans la phrase française, positions et topologie, Journées Syntaxe'*, Bordeaux, 2004.
- Gerdes K. & S. Kahane, to appear, "Phrasing It Differently", in L. Wanner (ed.), *Papers in Meaning-Text Theory in Honour of Igor Mel'čuk*, Benjamins, 34 p.
- Gerdes K. & H. Yoo, 2004, "A dependency account of Korean Word Order", in *Proceedings of the International Conference of the Linguistic Society of Korea*, Seoul, 2004.
- Kahane S., 2001, "What is a natural language and how to describe it? Meaning-Text approaches in contrast with generative approaches", *Computational Linguistics and Intelligent Text Processing*, Springer, 2001. P. 1—17.
- Kathol A., 1995, *Linearization-Based German Syntax*. Ph.D. thesis, Stanford University, 1995.
- Mel'čuk Igor, 1967, "Ordre des mots en synthèse automatique des textes russes", *T. A. Informations*, 8:2, 1967. P. 65—84.
- Mel'čuk I., 1988, *Dependency Syntax: Theory and Practice*, State Univ. of New York Press, Albany, 1988.
- Mel'čuk I., 2001, *Communicative Organization in Natural Language (The Semantic-Communicative Structure of Sentences)*. Amsterdam, Benjamins, State Univ. of New York Press, Albany, 2001.
- Mel'čuk I. & N. Pertsov, 1987, *Surface Syntax of English. A Formal Model within the Meaning-Text Framework*, Benjamins, Amsterdam, 1987.
- Nasr A., 1995, "A formalism and a parser for lexicalised dependency grammars", *4th Int. Workshop on Parsing Technologies*, State Univ. of NY Press, 1995.
- Vallduvi, E., 1992, *The Informational Component*. Garland, New York, 1992.
- Yoo H., 2003a, "De la topologie à la chaîne parlée : les niveaux de représentation pour une Théorie Sens-Parole", in *Proceeding of the First International Conference on Meaning-Text Theory*, Paris, 2003.
- Yoo H., 2003b, *Ordre des mots et prosodie : Essai de description et de formalisation pour le grec moderne*, PhD. Thesis, Université Paris 7, 2003.

Размышления о взаимодействии лингвистики и математики

А. В. Гладкий

Московский институт открытого образования
gladjkij@rol.ru

Проникновение в лингвистику математических методов и «математического духа» способствовало развитию лингвистики в сторону точности и объективности. Однако на пути ее дальнейшего развития в этом направлении стоят серьезные препятствия. Автор размышляет о причинах сближения лингвистики и математики, о границах применимости в лингвистике математических методов и о природе факторов, препятствующих взаимопониманию математиков и лингвистов.

Thinking about relations between linguistics and mathematics. Pollination with the “mathematical spirit” and methods turned linguistics toward more precise and objective lines. Yet there are serious drawbacks on this way. The author is thinking why the two sciences are coming closer to each other, how far is mathematics applicable to linguistics, and which factors oppose understanding between mathematicians and linguists.

Когда во второй половине 50-х гг. некоторые молодые лингвисты задумались о применении математических методов для исследования структуры языка и начали сотрудничать с математиками, это вызвало у очень многих их коллег удивление и даже шок — ведь они с детства были убеждены, что гуманитарные науки, одной из которых является лингвистика, с математикой и другими «точными» науками не имеют и не могут иметь ничего общего. Между тем наличие тесной связи между естественным языком и математикой во все не было в то время новым открытием. Л. С. Выготский писал в опубликованной в 1934 г. книге «Мышление и речь»: «Первым, кто увидел в математике мышление, происходящее из языка, но преодолевающее его, был, по-видимому, Декарт» и продолжал: «Наш обычный разговорный язык из-за присущих ему колебаний и несоответствий грамматического и психологического находится в со-

стоянии подвижного равновесия между идеалами математической и фантастической гармонии и в непрестанном движении, которое мы называем эволюцией». Возникшее в Древней Греции учение о грамматических категориях уже представляло собой описание ряда важнейших аспектов строения языка с помощью абстрактных моделей, близких по стилю к тем моделям, которые были созданы древнегреческими математиками для описания пространственных форм; только привычность таких понятий, как падеж, род и т. п., ставших, как писал Х. Штейнталь, «нашей второй натурой», мешает нам понять, какого высокого уровня абстрактного мышления потребовало их создание. Так что удивляться следовало бы скорее тому, что первые попытки использовать для описания языкового «идеала математической гармонии» настоящие математические средства были предприняты лишь в середине XX столетия.

Можно указать две причины такого «запоздания». Во-первых, наука о языке после значительных шагов, сделанных в античную эпоху, снова начала по-настоящему развиваться только в XIX столетии, но в течение всего этого столетия главное внимание лингвистов было обращено на историю языка, и лишь в следующем веке, который вообще был для гуманитарных наук веком структурализма, лингвистика впервые после античного периода обратилась к изучению языковых структур, но уже на новом уровне. Когда лингвисты осознали, что язык представляет собой, говоря словами Ф. де Соссюра, «систему чистых отношений», т. е. систему знаков, физическая природа которых несущественна, а существенны только отношения между ними, стала совершенно очевидна параллель между языком и математическими конструкциями, которые тоже являются «системами чистых отношений», и уже в начале XX столетия тот же де Соссюр мечтал об исследовании языка математическими средствами. Во-вторых, в математике в начале Нового времени вышли на первый план количественные методы, и только в XIX веке математики снова начали строить неколичественные абстрактные модели, отличавшиеся от античных более высоким уровнем абстракции, а также — что для нашей темы особенно важно — тем, что они могли использоваться для описания значительно более широкого круга явлений, чем пространственные формы; нередко такие модели оказывались удобным и даже необходимым средством для изучения явлений, о которых строившие их математики вовсе не думали и даже не знали об их существовании. Среди этих моделей были и те, которые впоследствии получили применение в лингвистике; особенно интенсивное развитие математических дисциплин, содержанием которых

было их построение, пришлось на первую половину XX столетия. Поэтому встреча математики и лингвистики в середине этого столетия была вполне закономерна.

Одним из результатов этой встречи было возникновение новой математической дисциплины — математической лингвистики, предметом которой является разработка математического аппарата для лингвистических исследований. Центральное место в математической лингвистике занимает теория формальных грамматик, по характеру используемого в ней аппарата родственная математической логике и в особенности теории алгоритмов. Она доставляет формальные методы описания правильных языковых единиц различных уровней, а также, что особенно важно, формальные методы описания преобразований языковых единиц — как на одном уровне, так и межуровневых. К теории формальных грамматик примыкает теория синтаксических структур, значительно более простая в отношении аппарата, но не менее важная для лингвистических приложений. В математической лингвистике разрабатываются также аналитические модели языка, в которых на основе тех или иных — считающихся известными — данных о «правильных текстах» производятся формальные построения, результатом которых является описание каких-то «составных частей» механизма языка. На этом пути можно получить формальное описание некоторых традиционных грамматических понятий. Сюда же следует отнести описание смысла предложения с помощью аппарата интенциональной логики («семантику Монтегю»).

Разумеется, с помощью математического аппарата можно описать только один из двух идеалов языка, о которых говорил Выготский; поэтому часто раздающиеся возражения против использования той или иной математической модели (или математических моделей вообще) на том основании, что такие-то и такие-то частные случаи она не охватывает, не имеют смысла: для описания присущих языку «колебаний и несоответствий» нужны совсем другие, не математические средства, и как раз четкое описание «математического идеала» могло бы помочь их находить, поскольку оно позволило бы ясно отграничивать в языке «фантастическое» от «математического». Но это пока что дело будущего.

Не меньшее, а может быть и большее значение, чем возникновение математической лингвистики, имело непосредственное проникновение в лингвистику фундаментальных математических идей и понятий — таких, как множество, функция, изоморфизм. В совре-

менной лингвистической семантике важную роль играют пришедшие из математической логики понятия предиката и квантора. (Первое из них возникло в логике еще тогда, когда она не отграничивалась от лингвистики, и теперь вернулось в лингвистику в обобщенном и математически обработанном виде.) И, наконец, очень большое значение имеет уточнение языка лингвистических исследований, происходящее благодаря проникновению в лингвистику «математического духа» не только в тех ее областях, где возможно использование математических идей и методов. Все это можно коротко резюмировать так: лингвистика становится все более точной и более объективной наукой — не переставая, само собой, быть наукой гуманитарной.

Однако на этом естественном пути развития лингвистики стоят серьезные препятствия, которые могут его надолго затормозить. Главное из них — возникшее в начале Нового времени «разделение факультетов»: естествоиспытатели и математики с одной стороны и гуманитарные ученые с другой не интересуются работой коллег «на другом факультете» и, более того, — в глубине души, а нередко и открыто презирают их. Математики и естествоиспытатели (и еще больше «технари») склонны видеть в гуманитарных исследованиях всего лишь некое «украшение» или даже «пустую болтовню», а «гуманитарии» готовы терпеть математику и естественные науки лишь ради практической пользы и убеждены, что они ничем не могут помочь постижению природы человеческого духа. Только в середине XIX столетия в этой, говоря словами великого биолога и великого мыслителя Конрада Лоренца, «зловредной стене между естественными и гуманитарными науками (die böse Mauer zwischen Natur- und Geisteswissenschaften)» была пробита первая брешь в самом тонком месте, отделявшем логику от математики. В XX столетии появились и другие бреши — среди них и та, которую пробили с двух сторон математики и лингвисты, — но их все еще мало, стена крепка до сих пор, и нет недостатка в усилиях с обеих сторон укреплять ее дальше и латать пробоины. Нередко эти усилия бывают довольно успешны; последнее «достижение» в этом направлении — «профильное образование» в средней школе, уже в детстве разделяющее способных и интересующихся людей на «факультеты» и приучающее их гордиться невежеством в «чужих» науках — может очень сильно воспрепятствовать дальнейшему сближению естественных и гуманитарных наук, действительно необходимому для нормального развития тех и других. Одно из последствий воздвижения стены состоит в том, что «гуманитарии», включая подавляющее большинство лингвистов,

ничего не знают даже об азах как раз тех разделов математики, которые имеют наибольшее значение для гуманитарных наук (и представляют себе математика как человека, занятого исключительно вычислениями).

Другое препятствие — характерная для нынешнего состояния науки бешеная гонка, безостановочная погоня за все новыми и новыми «результатами», сужающая кругозор и не оставляющая времени задуматься над более глубокими проблемами или заняться серьезным изучением смежной и тем более не совсем смежной научной дисциплины. Это относится в равной степени к лингвистам и к математикам — как, впрочем, и ко всем, кто профессионально занимается наукой. И третье — инертность, или, проще говоря, лень. На первый взгляд лень и бешеная гонка несовместимы, но в действительности они прекрасно уживаются между собой и, более того, поддерживают и стимулируют друг друга. Когда человеку лень взяться за трудное дело, он хватается за более легкое и более «надежное», успехи в котором оправдывают и поощряют его инертность. Высокомерное отношение к «меньшим братьям», копошащимся по другую сторону стены, также поощряет лень и поощряется ею. Когда, например, математик предлагает пересмотреть все представления о древней истории, не дав себе труда хоть немного познакомиться с древними языками, за это в весьма значительной степени ответственна та же лень-матушка.

Опасность для развития науки, создаваемая этими препятствиями, гораздо серьезнее, чем может показаться на первый взгляд. Когда невежество в «чужих» науках становится предметом гордости, это закономерно ведет к поверхностности и невежеству также и в «своих». «Факультетов» давно уже много больше двух, число их растет из года в год, и каждый отгораживается стеной от других; появляются стены и внутри факультетов. Кругозор исследователей постепенно сужается; правда, аппарат исследования становится все более тонким и изысканным, но в поле его зрения попадают почти исключительно мелкие предметы, и укрепляется представление, будто только они и заслуживают изучения. Есть все основания говорить о кризисе в науке, и лингвистика не является исключением. Сейчас, как мне представляется, самое время оглянуться и задуматься.

Здесь собрались лингвисты того направления, которое связано с моделью «Смысл — Текст». Эта модель, созданная в 60-е годы теперь уже прошлого столетия, была одним из первых и лучших результатов встречи лингвистики и математики, после которой выросли уже два поколения лингвистов, со студенческих лет приучивших-

ся к точному мышлению. Но и они, к сожалению, не свободны от инертности, мешающей им осознать наличие кризиса и задуматься о путях его преодоления. Между тем среди всех лингвистов — и, может быть, даже среди всех, кто занимается гуманитарными науками — у них больше всего объективных возможностей для такого осознания, и хотелось бы надеяться, что они этими возможностями воспользуются.

A Hypothesis of Two Syntactic Starts¹

Leonid Iomdin

Institute for Information Transmission Problems, Russian Academy of Sciences
iomdin@iitp.ru

The paper deals with a special type of Russian constructions: sentences with embedded (semi-)phraseological expressions like *Он занимается **чёрт знает** чем* 'He does **the Devil knows** what'. It is very difficult to build adequate surface syntactic representations for such sentences. An unexpected solution is proposed for this problem, admitting that sentences of this type have two syntactic starts. As such, they seem to violate the fundamental requirement of the surface syntactic component of the Meaning $\Leftrightarrow$ Text theory that the syntactic structure of any sentence should be a tree. Apart from this, such constructions have other interesting syntactic and semantic features, which are also discussed in the paper.

Гипотеза о двух синтаксических началах. Работа посвящена особому типу конструкций русского языка: предложениям с вложенными (полу-)фразеологическими выражениями, ср. *Он занимается чёрт знает чем*. Построить адекватную поверхностно-синтаксическую структуру для таких предложений оказывается весьма сложно. В работе предлагается неортодоксальное решение этой проблемы, исходящее из того, что предложения такого типа имеют два синтаксических начала. Тем самым эти конструкции, похоже, нарушают фундаментальное правило модели Смысл $\Leftrightarrow$ Текст, согласно которому синтаксическая структура любого предложения должна представлять собой дерево. Помимо этого, обсуждаемым конструкциям присущи другие интересные синтаксические и семантические особенности, также рассматриваемые в работе.

Problem

In long experience of handling syntactic structures of sentences in the Meaning $\Leftrightarrow$ Text² formalism, both for theoretical purposes³ and appli-

¹ This study was partially supported by a research grant No. 04-04-00263a from the Russian Humanitarian Sciences Foundation, whose assistance is gratefully acknowledged.

² See e.g. Мельчук 1974, Mel'čuk 1988.

cation parsers⁴, I found one type of constructions especially challenging. In Russian, these are sentences with embedded (semi-) phraseological expressions (given in bold) like

- (1) *Он занимается **чёрт знает** чем* ‘He does **the Devil knows** what’,
 (2) *Мне было — так лестно / Лезть за тобою — **Бог / Знает** куда!* ≈ ‘I felt so flattered to climb after you **God knows** where’
 (Marina Tsvetayeva)

and many others. In other Slavic languages as well as in English, German, or French, these constructions have close matches, so the phenomenon is rather general. For English, such constructions have been dealt with by George Lakoff, who referred to them as syntactic amalgams⁵ (Lakoff 1974), Martin Haspelmath (1997), who gave a typological picture of related phenomena, and many others. I will however restrict myself to Russian for definiteness⁶.

The challenge is that it is extremely difficult to build adequate surface syntactic representations for such sentences. Namely, it is unclear what the syntactic role of the verb *знает* ‘knows’ in the boldface part of (1) and (2) can be.

On the one hand, it definitely cannot be the topmost head of the surface syntactic tree, for the very simple reason that sentence

- (1') *Чёрт знает, чем он занимается* ‘The devil knows what he does’

where *знает* is the top of the tree, is neither syntactically nor semantically equivalent to (1); and that, for sentence (2), no similar permutation is possible at all.

On the other hand, there is no reasonable syntactic governor for *знает* within the sentences considered. Indeed, should we wish to subordinate it to the main verb of the sentence (*занимается* ‘does’, ‘is engaged in’ in (1) or *лезть* ‘climb’ in (2)), we shall immediately face the problem of what the syntactic relation between the verbs is — and this problem has no plausible answer.

³ See e.g. Иомдин 1990 and especially Иомдин 2003, where similar types of syntactic idioms are discussed.

⁴ See e.g. Apresjan *et al.* 2003, Iomdin *et al.* 2002

⁵ Similar (but not identical!) phenomena were described in Апресян–Иомдин 1989 with the help of a close term “syntactic agglomerate”.

⁶ Recently, these constructions have been given a thorough consideration in a comprehensive paper by Yakov Testelefs and Elizaveta Bylinina (Testelefs–Bylinina 2005). Regrettably, this paper came to my attention after I had completed the present article, so I could not take adequate account of their work.

There may also be a temptation to view the syntactic governor of *знает* in the pronouns *куда* ‘whither, where to’ or *где* ‘where’. As was mentioned, the expressions like *чёрт знает* seem to be phraseological and may be suspected of having transformed into merged lexical units (amalgams or agglomerates) more or less equivalent to indefinite particles like *-нибудь* or *-либо*, which are commonly used in Russian to form indefinite pronouns out of interrogative ones (*куда-нибудь* ≈ ‘somewhere’, ‘anywhere’ etc.).

However, such a solution does not really hold water, since the embedded constructions of this type are not confined to phraseological expressions cited and may include rather free clauses formed with different verbs — as in

- (3) *Когда я был подростком, сильное впечатление на меня произвела вычитанная **не помню уже** в какой книге история панамской авантюры.* ‘When I was a youth I was deeply impressed by the story of the Panama adventure that I read in **I don’t remember** which book’ (Novoye Vremya, No 10, 2005);
- (4) *«Вдогонку» мне в «Вестнике ЛГУ» была напечатана статья И. Лапицкого, в которой я обвинялся во всех смертных грехах: я и монархист, и эсер, и троцкист, и **ещё не упомяну** кто.* ‘Following this, Leningrad University Bulletin published a paper of I. Lapitsky, where I was accused of all mortal sins: I am a monarchist, a socialist-revolutionary, a Trotskyist and I can’t remember who else’ (Dmitry Likhachev, Reminiscences, Logos, St. Petersburg, 1997);
- (5) *В Америке, **не припомню**, в каком городе, два торговца зашли в трактир и сели обедать.* ‘In America, **I can’t remember** in which town, two salesmen dropped in in a pub and sat down to dinner (Sodeistvie Newspaper, 1868)
- (6) *На Унжу мы забрались **сами не знаем** зачем* lit. ‘We came to the Unzha River **we do not know ourselves** what for’ (Denis Osokin, Angels of Revolution, 1923).

Moreover, even the second parts of the constructions in question are not necessarily interrogative pronominal words — they may be represented by the conjunction *или* ‘or’ or the particle *ли* ‘whether’, cf. perfectly acceptable examples like

- (7) *Его судят за преступление, которое он неизвестно совершил **или нет*** lit. ‘He is being tried for a crime which it is not clear if he has committed or not’

- (8) *Кроме того, есть ещё такие сдерживающие факторы, как наличие Северной Кореи с непонятно имеющимся ли у неё ядерным оружием* 'Besides, there are such deterrent factors as the presence of North Korea with nuclear weapons that it might or might not have' lit. '... the presence of North Korea with it-is-unclear-whether-available-to-it nuclear weapons' (Irina Kobrinskaya, an analytical review on the Polit.Ru website, April 25, 2005).

It is hardly possible to consider all these expressions as equivalents of particles.

Whilst there is no evident syntactic governor for the second verbs of the sentences considered, the pronominal words (or noun phrases including the word *какой* 'which' in (3) and (5)) have as many as two plausible candidates for governor.

To demonstrate this, we will focus, for the sake of definiteness, on sentence (1) and try to find the syntactic governor for the pronominal word *чем*. Obviously enough, both verbs of the sentence — *занимается* in the first part and *знает* in the embedded construction — have good reasons to be classified as the pronominal word's governor.

Indeed, on the one hand, one may suggest that *чем* instantiates the 1st completive valency of *заниматься*, being the only word of sentence (1) that stands in the instrumental case — exactly the one that is required by *заниматься*. On the other hand, the same pronominal word may be viewed as instantiating the 1st completive valency of the verb *знать*, the way it does in isolated (elliptic) sentences like *Я знаю чем* 'I know what'. It should be added that both verbs in (1) appear to have their first complements filled, as the sentence does not look elliptic at all.

So, the syntactic structure of (1) has two oddities at a time: one word in need of a governor (*знает*) has no good candidate while another word (*чем*) has two.

Solution

To solve the puzzle of syntactic representation for (1), I propose to take a somewhat broader look at the organization of this sentence.

First of all, note that the duality of syntactic dominance for *чем*, if it is really present in (1) as just demonstrated, is far from trivial and requires further reasoning. Namely, one has to take account of the fact that in sim-

ple single-clause sentences pronominal words like *чем* **cannot** depend on verbs that, unlike *знать*, do not take propositional complements: one cannot say for example **Я занимаюсь чем* ‘I do what’ unless the sentence is interrogative⁷. Accordingly, since (1) is not interrogative, we must either discard the idea of *чем* depending on *заниматься* or conclude that it is **not** a single-clause sentence.

The second option is, to my mind, more reasonable. Assuming that (1) is not a single-clause sentence, let us try to define what clauses it may consist of. The most natural assumption is that (1) consists of two clauses, one constituted by verb *занимается* and the other constituted by verb *знает*. In this case, where are the boundaries of the two clauses? The left-hand boundaries of both clauses are evident: for the first clause it is the beginning of the whole sentence and for the second clause it is the word *чёрт* which is the subject of the verb *знает*. What about the right hand boundaries?

My hypothesis is that the right-hand boundaries of both clauses **are the same and coincide** with the end of the sentence, so that the pronominal word *чем* belongs to both clauses.

As a matter of fact, this suggestion alone has nothing especially revolutionary about it. After all, in many types of (non-coordinative) two-clause sentences the whole subordinate clause may be viewed as fully contained in the main clause, with the topmost node of the subordinate clause dominated by a node in the main clause. Cf. for instance sentences like

(9) *Я знаю, чем он занимается* ‘I know what he does’,

where the head *занимается* of the subordinate clause *чем он занимается* is subordinated to the verb *знаю* with the help of the 1st completive syntactic relation.

It is exactly **the lack of such subordination** that distinguishes the second clause of (1) from the subordinate clause of (9). The head of the second clause of (1) remains without a governor at all, and this is, to our mind, the most crucial characteristic of this type of sentences.

If we are allowed to use a metaphor to describe the process of sentence generation, we can say that sentences (9) and (1) are unfolding differently: while (9) is smoothly produced by the speaker, (1) has a sort of

⁷ To be more precise, such pronouns may either form a special question like *Чем вы занимаетесь?* ‘What do you do?’ – in which case the pronoun is interrogative too — or a highly colloquial general question like *Вы занимаетесь↑чем?* ‘Do you do anything?’ where *чем* is an indefinite pronoun and really means ‘anything’.

leap amidst generation: before the first clause is finished, the second clause starts to evolve⁸, and, after some time, the two proceed together until the end of the whole sentence. Continuing with the metaphor, we can further say that the second clause in (1) behaves like **a tributary to a river, which contributes to its course**.

In a way, such evolution of sentence (1) resembles the correlation between the main and the parenthetical clauses if the latter is situated in the middle of the sentence, as in

- (10) *В этот момент какой-то молодой человек (это и был Иван) поднялся с места:* ‘At this moment a young man (**this was Ivan**) rose from his place’

or, to take a more colloquial example,

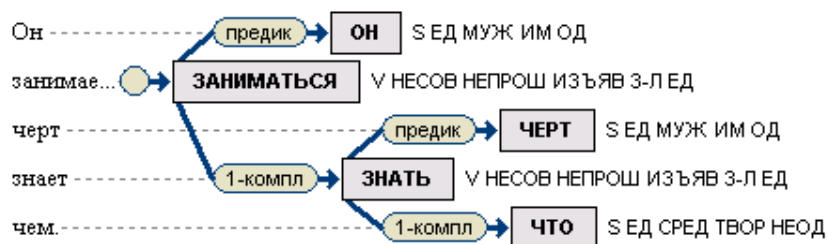
- (11) *Её отец, пусть земля ему будет пухом, много лет работал учителем* ‘Her father, **may he rest in peace**, worked as a teacher for many years’.

Indeed, in (10) and (11) parenthetical clauses given in bold start before main clauses are finished. The drastic difference between (10) and (11) on the one hand and (1) on the other hand is that parenthetical clauses are finished sooner than the main clauses while in (1) the “tributary” clause ends together with the first clause. The author knows of no other constructions in Russian (or other languages mentioned earlier) that share this property with sentences like (1)–(8).

Of course, if this stand is taken, we will have to admit that sentences of this type have **two syntactic starts** — and as such they violate the fundamental requirement of the surface syntactic component of the Meaning $\Leftrightarrow$ Text theory that the syntactic structure of any sentence should be a tree. I believe, however, that in this case such a decision is fully justified.

On the other hand, once we agree theoretically that the underlying syntactic structures of sentences like (1) are graphs with two top nodes, we may **offer technical solutions** to retain tree representations even for such sentences (in particular, this would be useful for NLP applications). A reasonable idea could be that in (1) the verb *знает* that forms the second clause is subordinated to the first verb of the sentence, instantiating its 1st completive valency, as shown in the following screenshot:

⁸ In oral speech, this leap is clearly supported prosodically and intonationally. For reasons of space we will however not go into any details of this support.



With the theoretical realization that the structure looks differently, such a representation will no more be misleading. Mutatis mutandis, similar syntactic representations could be built for all sentences (1)–(8).

Discussion

Apart from the general syntactic peculiarities of sentences discussed above, these constructions have other features that have to be accounted for.

One more **syntactic** particularity is that expressions like *чёрт знает что* may include a personal pronoun (given in bold) whose syntactic status is not very clear; e.g.

- (12) *Ему давно уже пора дом покупать, снимает чёрт **его** знает что!* lit. ‘It’s high time he buys a house, he rents the Devil knows **him** what’ (Alexander Torin, Gelikon Plus, St. Petersburg, 2000);
- (13) *Деньги уходят чёрт **их** знает куда* lit. ‘Money goes the devil knows **it** where’ (Vladimir Lenin, in a letter to his mother, 1895).

In most (but, importantly, not all!) cases, these personal pronouns are coreferent with the subject of the utterance (*he* in (12) and *money* in (13)) agreeing with it in gender and number. It is however difficult to determine what syntactic relation holds between the verbal head of the construction and this pronoun: on the one hand, such a pronoun stands in the accusative and should be treated as the 1st (direct) complement of the verb *знать*; on the other hand, the 1st completive slot of the verb is already instantiated by the pronominal word like *что* or *куда* while, as is well known, the 1st completive syntactic relations is non-repeatable. It should be added that in this case the verb *знать* seems to controversially combine in one occurrence two mutually exclusive senses: the stative sense ‘know a fact’ (which corresponds to the German *wissen* or French *savoir*) and the “acquaintanceship” sense ‘know a person or thing’ (correspond-

ing to German *kennen* and French *connaître*)⁹. At the moment, the author is unable to offer a fully satisfactory solution to this controversy: more study and thinking is needed. A purely technical solution could be to subordinate the personal pronoun to the verb using an auxiliary syntactic relation, to avoid the non-repeatability condition violation.

Even though the constructions discussed are not confined to phraseological expressions, they are subject to rather tight **lexical restrictions**.

Within the phraseological subset, the constructions are formed with the verbs *знать* and, occasionally, *ведать* ‘know’, almost always in the present tense, whose subjects can be either

1) nouns *чёрт*, *дьявол* ‘devil’, *леший* ‘wood goblin’, *бес* and *бис*¹⁰ ‘demon’, *шут* ‘jester’ and *нёс* ‘dog’ (the last two are probably euphemisms for *чёрт*), practically always in the singular¹¹, or

2) derogatory nouns like *физ* or *хрен* that are in fact euphemisms for an obscene word, or this obscene word itself, as in *В стране скоро физ знает что начнется* ‘Soon, goodness knows what will start in this country’, or

3) nouns *Бог* ‘God’, *Господь* ‘Lord’, *Аллах* ‘Allah’, *Всевышний* ‘Almighty’, as in *Мне не нравится, что на юбилей города приглашают Бог знает кого* ‘I don’t like it that they invite God knows whom to attend the city anniversary’. With the noun *Бог*, an additional obsolete verbal form *весть* ‘knows’ can be used: *Он отправился Бог весть куда* ‘He went God know where’¹².

Interestingly, even close synonyms or cognates of these words (like *сатана* ‘Satan’, *шайтан* ‘Shaitan’, *демон* ‘demon’, *чертовка* ‘she-devil’, *чертёнок* ‘little devil’, *ведьма* ‘vixen’; *Богиня* ‘Goddess’ etc.) are virtually impossible here, although expert speakers can say something

⁹ As a matter of fact, such superposition of senses in Russian *знать* sometimes occurs even outside the limits of these expressions; cf. the famous quotation from Gogol’s *Dead Souls*: *Знаем мы вас, как вы плохо играете!* ‘I know you and your poor play’, lit. ‘We know you how poorly you play’. Such examples also defy regular surface syntactic representation.

¹⁰ *Бис* (*бис*) is in fact the Ukrainian for *бес* but it does not appear to be felt as a Ukrainism by an average Russian speaker.

¹¹ In jocular usage, one can also meet something like *По обе стороны трассы размещаются таверны, рестораны, кафе и ещё черти знают что* — lit. ‘On either side of the highway there are pubs, restaurants, cafes and the devils know what else’ (Web posting).

¹² A syntactically similar expression with the negation — *не бог весть* forms a different idiom with the meaning close to ‘not really impressive’, as in *не бог весть что* ‘no great thing’.

like *Первая корректура ушла из издательства Будда знает сколько времени назад* — lit. ‘The first proof-sheet left the publisher Buddha knows how long ago’ (from a posting about the publication of a manuscript on East Asia).

Within the non-phraseological subset, the construction can be formed by a few verbs with the meaning of ‘know’ (*знать*), ‘understand’ (*понимать*, *разбирать*) or ‘remember’ (*помнить*, *припоминать*, *упоминать*). These verbs stand in the present or, remarkably often, in the perfective future with a meaning of the so-called present tense of vain expectation, while their subjects are mostly personal pronouns (which are often elided), first of all *я* and *ты*, as in *Это уже тогда не бой, а я не знаю что* ‘It is not a battle in this case but I don’t know what’ (Arkady and Boris Strugatskys); *Им ещё, сверх того, показывали длинное представление с декорациями, на которых было нарисовано не поймёшь что, с актёрами, которые изображали не разберёшь кого*. lit. ‘Apart from that they were shown a long performance with scenery that presented **you won’t understand** what and actors who played **you won’t make out** whom’ (Nikolay Nosov); cf. also examples (3)—(5) above.

In addition to these verbs, the construction can be formed by several predicative adjectives and adverbs, including *непонятно* ‘incomprehensible’, *неизвестно*, *неведомо*, *невесть* and *незнамо* ‘not known’, as in *Живёте вы непонятно зачем* ‘You live one does not understand why’ (Arkady and Boris Strugatskys), *Ходит неизвестно где* ‘He goes one does not know where’ The last four words are close synonyms, and the last two of them are used solely within the construction discussed; cf. *Говорили слово по-нашему, слово по-татарски, а то промеж себя невесть по-каковски* ‘They spoke one word in our language, one word in Tartar and between themselves heaven knows in what language’ (Nikolay Leskov); *За это незнамо кого благодарить надо* ‘One needs to thank goodness knows whom for this’ (Konstantin Paustovsky).

This list can be supplemented by an expression with an obscure etymology *чёрт-те*, as in *Он был одет классно, а я чёрт-те в чем* ‘He was smartly dressed and I was wearing the devil knows what’ (Galina Shcherbakova), a highly colloquial expression *не пойми* ‘do not understand’ where the verb *понимать*, somewhat enigmatically, stands in the imperative perfective (cf. *В Ираке не пойми что происходит* ‘In Iraq, goodness knows what happens’, and a recent slang expression, spreading amazingly fast in youth subculture, *X3* — evidently an abbreviation of the substandard *хрен знает* or even its stronger variant; cf. *Мы забрели X3 куда* ‘We wandered the devil knows where’).

As a matter of fact, most of the expressions listed in the last two paragraphs cease to act as collocations and turn into single words, approximating, at individual paces, to indefinite particles like *-либо*, *-то* etc.

The **semantics** of the *Чёрт знает что* type of construction is very interesting and deserves special attention and careful study. Due to space limitations, I will restrict myself to several remarks only.

1. Despite lexical richness of expression, the meanings of collocations that represent the construction are remarkably close to each other. All of these collocations have a strong evaluative component that expresses the speaker's negative attitude toward the participant or circumstance of the situation conveyed by the collocations. There is, however, a noticeable difference of meaning between the variety of collocations based on *бог* and its synonyms, on the one hand, and the remaining collocations, on the other hand. In the former, the speaker's negative attitude becomes milder and is substituted by regret and, possibly, compassion. To my mind, the speaker's negative attitude belongs to the assertive part of the meaning rather than the presupposition. In particular, this may account for the ungrammaticality of sentences like **Он предал чёрт знает кого* 'He betrayed the devil knows whom': in all probability, the semantics of the verb *предать* ≈ 'be disloyal to' requires that its object deserve loyalty and the collocation *чёрт знает кто* introduces an unknown and/or bad person who does not deserve loyalty.

2. The construction considered here has clear **negative polarity**. Indeed, almost all of its lexical realizations have an overt or incorporated negation — but even the variants without the negation (*чёрт знает что*, *бог знает куда* etc.) introduce unknown entities. *He went God knows where* really means hardly anything very different from *Nobody knows where*.

3. At least some of the collocations that represent the construction lack compositionality. A good example is the expression containing *сколько* 'how much': for reasons that I cannot explain at the moment, sentences like *Он получил чёрт знает сколько денег* 'He got the devil knows how much money' refer to situations that involve an indefinitely **large** amount of money but never to situations that involve an indefinitely **small** amount of money.

Concluding the discussion, I would like to emphasize once again that the constructions considered here are unique and have no close cognates in the language. In particular, the constructions like *Иди куда хочешь* <*куда тебе угодно*> 'Go wherever you please', *Он танцует с кем попало* 'He would dance with the first person he comes across', *Ребёнок*

ест что ни попадя ‘The child eats whatever comes to hand’ that share with our constructions the presence of interrogative pronouns and the meaning of indefiniteness are nonetheless drastically different from them (in English terminology, they are often referred to as sluicing constructions). Most importantly, they do not have an additional syntactic start.

References

- Apresian, Jurij, Igor Boguslavsky, Leonid Iomdin, Alexander Lazursky, Vladimir Sannikov, Victor Sizov, Leonid Tsinman. ETAP-3 Linguistic Processor: a Full-Fledged NLP Implementation of the MTT // MTT 2003, First International Conference on Meaning $\Leftrightarrow$ Text Theory. Paris, École Normale Supérieure, Paris, June 16—18 2003, pp. 279—288.
- Iomdin, L. L., V. G. Sizov, L. L. Tsinman. Utilisation des poids empiriques dans l’analyse syntaxique: une application en Traduction Automatique // META, vol. 47. 2002. N. 3. pp. 351—358.
- Haspelmath, Martin. Indefinite pronouns. Oxford Studies in Typology and Linguistic Theory. Oxford: Oxford University Press, 1997.
- Lakoff, George. Syntactic Amalgams // Papers from the 10th Meeting of the Chicago Linguistic society, 1974, pp. 321—344.
- Mel’čuk, Igor A. Dependency syntax: Theory and practice. Albany, NY: SUNY, 1988.
- Testeleys Y., E. Bylinina. Sluicing-Based Indefinites in Russian // Formal Approaches to Slavic Linguistics 13: The South Carolina Meeting. Ann Arbor, MI: Michigan Slavic Publications. 2005, 355—364.
- Апресян, Ю. Д., Иомдин Л. Л. Конструкции типа НЕГДЕ СПАТЬ в русском языке: синтаксис и семантика. (Constructions of the NEGDE SPAT' type in Russian: Syntax and semantics.) // Semiotika i informatika, No. 29. Moscow, 1990, pp. 3—89.
- Иомдин, Л. Л. Автоматическая обработка текста на естественном языке: модель согласования. (Automatic natural language processing: A model of grammatical agreement). М., Наука, 1990.
- Иомдин, Л. Л. Большие проблемы малого синтаксиса. (Big problems of minor syntax) // Труды международной конференции по компьютерной лингвистике и интеллектуальным технологиям Диалог’2003. Протвино 2003, С. 216—222.
- Мельчук, И. А. Опыт теории лингвистических моделей «Смысл $\Leftrightarrow$ Текст» (The Theory of the Linguistic “Sense $\Leftrightarrow$ Text” models.), М., Наука, 1974.

Hooking Up Syntagmatic Lexical Functions to Lexicographic Definitions¹

Lidija Iordanskaja, Alain Polguère

OLST — Département de linguistique et de traduction
Université de Montréal

iordansl@magellan.umontreal.ca / alain.polguere@umontreal.ca

This paper presents a research in progress on the modeling of syntagmatic lexical relations in explanatory combinatorial dictionaries. More precisely, we examine how meanings expressed by collocatives and encoded by means of lexical functions connect to specific semantic components in the definition of bases in collocations. This research is based on the analysis of French data taken from the DiCo and BDéf databases, with a focus on lexical functions of the **Magn** family.

Связь синтагматических лексических функций с толкованием лексем. Статья посвящена представлению синтагматических лексико-функциональных отношений (типа Magn) между элементами коллокаций в толково-комбинаторных словарях. Предлагается задавать компонент толкования ключевого слова коллокации, с которым связан смысл лексической функции, опираясь на структуру толкования, используемую в базе данных BDéf. Установлены типы связи между этим компонентом и смыслом лексической функции — для рассмотренных французских коллокаций, содержащихся в базах данных DiCo и BDéf.

1. Introduction

This paper presents a research in progress conducted within the theoretical and descriptive framework of explanatory combinatorial lexicology (Mel'čuk, 1995; Mel'čuk *et al.*, 1995). Its general topic is the modeling of syntagmatic lexical relations in formal dictionaries. A syntagmatic lexical link between two lexical units L_1 and L_2 corresponds to a collocation — in Hausmann (1979) and Mel'čuk (1998) sense — of

¹ Thanks to Sylvain Kahane, Igor Mel'čuk and two MTT-05 reviewers for their comments on a first version of this paper.

which L_1 is the base (the controlling element) and L_2 the collocative (the controlled element). The question we are trying to answer is how meanings expressed by collocatives connect to specific semantic components in the definition of bases in collocations. This is a particularly interesting issue as a better understanding of the relationship between the semantic properties of lexical units and their lexical combinatorics could eventually lead to a better, more precise encoding of syntagmatic lexical function relations; on lexical functions — hereafter, *LF* — see Mel'čuk, 1996; Mel'čuk, 1998. This preliminary research is based on the analysis of French data taken from the DiCo and BDéf lexical databases (Polguère, 2000; Altman & Polguère, 2003). It mainly focuses on lexical functions of the **Magn** family: **(Anti)Magn**, **(Anti)Bon** and **(Anti)Ver**. For lack of space, we will presuppose familiarity with descriptive principles of explanatory combinatorial lexicology and, more specifically, with the system of standard lexical functions.

It has been common practice for lexicographers working within the framework of explanatory combinatorial lexicology to introduce in lexicographic definitions components whose sole purpose was to ensure direct link between LFs and lexicographic definitions. However, we believe that this approach may be “unnatural” if used without caution. For instance, we believe it would be artificial (and potentially wrong) to introduce in the definition of Fr. CELIBATAIRE_N (‘bachelor’) a semantic component pertaining to a certain “way of life,” to convictions, etc. in order to account for the collocation *célibataire endurci* (‘confirmed bachelor’, lit. *hardened bachelor*). The French lexical unit CELIBATAIRE_N strictly denotes a matrimonial status and can be used, in a very neutral way, without any reference to personal preferences (for instance, in official documents).²

In order to correctly hook up syntagmatic lexical functions to definitions, a decision has to be made on a case by case basis. For instance, while we find no evidence that a special component of the definition of CELIBATAIRE_N is “responsible” for the particular **Magn** [*célibataire*] *endurci*, we think that the definition of Fr. OURAGAN (‘hurricane’) has to include the component ‘that can destroy’ not only to account for the single **Magn** [*ouragan*] *dévastateur* (lit. *devastating* [*hurricane*]), but also because a significant part of its combinatorial

² Compare CELIBATAIRE with the locution VIEUX GARÇON (lit. *old boy*), which denotes a male bachelor of a certain age; this locution necessarily entails some characterization of the mind frame of the individual in question.

behavior is linked with the notion of destruction: for instance, *l'ouragan dévaste, ravage une ville* (lit. *the hurricane devastates, ravages a city*). The approach we propose seems less artificial in that it allows for different ways of establishing connections between syntagmatic LFs and the definition of the base. At the same time, this approach is formal in nature; as we will show, we are interested in making explicit LF-definition connections by using a formal encoding of lexicographic definitions. For a previous study of the same kind of problem, see Kahane (2003).³

Our presentation is articulated into three sections. Section 2 introduces our methodology for describing LF-definition connections; section 3 presents the different types of LF-definition connections that we have until now been able to identify; finally, section 4 shows how LF-definition connections can be formally encoded.

2. Methodological approach

2.1. Working with a new type of formal definitions

The problem of modeling LF-definition connections can only be solved by using a formal and well-structured encoding of lexical definitions. For this reason, we work with lexicographic definitions written in a very specific format, namely that of the BDéf database, as presented in Altman and Polguère (2003). The BDéf approach can be roughly characterized as follows.

- The BDéf is a database that contains definitions of French lexical units.
- It was first developed by making formal translations of definitions taken from the published French Explanatory Combinatorial Dictionary (ECD) — see Mel'čuk *et al.* (1984, 1988, 1992, 1999).
- Its defining language functions as some sort of “pseudo-code” for both reasoning on word senses and formally encoding their description in a machine-readable format.⁴

For the sake of illustration, we present below the BDéf definition for Fr. ASSURER¹**I.1** [*Zoé a assuré à Luce que Félix viendrait*. ‘Zoé assured

³ The focus of Kahane (2003) is on the **Real/Fact** family of lexical functions for Fr. BLESSURE (*wound*_N), most types of **Magn** for this lexeme being seen as variants of **A₁Fact₁**.

⁴ This is quite similar to the DiCo's formal approach, where lexical records are automatically compiled into representations that are tailor-made for consultation and automatic processing — see Steinlin *et al.* in these proceedings.

Luce that Félix will come’].⁵ As we cannot enter into the detail of formal conventions used in the BDéf, we start with an English gloss of this lexeme’s definition in linear form (ECD-like definition). Additionally, English glosses/comments are provided next to the BDéf tabular definition.

English linear gloss of the ECD definition of ASSURER¹I.1

X assure que Y a Z = Person X says to person Z that Y took, takes or will take place in order for Z to be sure that Y had taken, takes or will take place.

BDéf definition of ASSURER¹I.1

English tabular gloss

Central component

*/*Communication_langagière*/*

1: X dire *2 à Z

Linguistic communication

Prop. 1: X says *2 to Z

Body

*/*Contenu_informationnel*/*

2: Y _avoir_lieu_.passé/présent/futur

Informational content

Prop. 2: Y took, takes or will take place

*/*But*/*

3: *1 dans_le_but *4

4: Z sûr#Ila de *2

Goal

Prop. 3: *1 in order to *4

Prop. 4: Z be sure of *2

Semantic type of variables

X: personne

X is a person (i.e. individual or social group)

Y: fait

Y is something that takes place (i.e. not an entity)

Z: personne

Z is a person

Actancial status of variables

~ [X, Y, Z]

X is the 1st, Y the 2nd and Z the 3rd argument of ASSURER¹I.1

The main advantage of using BDéf definitions in our work is that it allows us to precisely pinpoint specific components of the lexicographic definition and refer to them in order to model patterns of LF-definition connections. This will be demonstrated later, in Section 4.

2.2. What questions need to be answered?

In order to make clearer the problem we are trying to solve, let’s state it in lexicographic terms. In the French ECD and DiCo articles for

⁵ For additional examples commented in English (MAISONI.1 ‘house’ vs. MAISONI.2 ‘home’), see Barque and Polguère (2004).

ASSURER¹I.1, we find the following data about (**Anti**)**Magn** and (**Anti**)**Ver** LF links:⁶

Magn	: <i>catégoriquement, fermement; avec force, avec véhémence</i>
AntiMagn	: <i>mollement, timidement</i>
Ver	: <i>avec raison</i>
AntiVer	: <i>à tort</i>

The above encoding is convenient in that it is both compact and allows us to characterize in a very general way collocations controlled by ASSURER¹I.1. However, a learner of French looking at these data could legitimately ask the following questions.

- There are two groups of **Magn** (separated by a semicolon), but what is the actual semantic difference between the two?
- There is only one group of **AntiMagn**; should it be related to the first or to the second group of **Magn**?
- To what specific component(s) of the definition of ASSURER¹I.1 does each collocative meaning connects within the semantic representation of the whole collocation?

This latter question is very important as it boils down to asking what the speaker **actually** expresses when he utters *assurer catégoriquement ~ fermement ~ avec force ~ avec véhémence ~ mollement ~ timidement ~* etc. For instance, what specific parameter in the situation denoted by ASSURER¹I.1 does the speaker characterize as **AntiVer** — i.e. as being not the way it should be — when he says *X assure à tort que Y*? What is not as it should be? Was X wrong in assuring Y in the given context? Was Y untrue? Was X not believing in Y?

There are at least three reasons why we need to establish a better connection between LF and definitions:

1. We believe it will allow us to identify general subtypes of LFs, thus increasing the potential “granularity” of the system of standard LFs.⁷
2. It is a means of validating the structure and content of the ECD definition of the collocation base as all collocatives should find a natural semantic hooking to the base meaning.
3. It will help us build ECD articles for lexical units that mainly function as collocatives.

On this latter point, it is clear that lexical units such as CATEGORIQUEMENT, FERMEMENT, etc. that appear above as collocatives of ASSURER¹I.1

⁶ Data presented here are in fact richer than what can be found in the actual published article for ASSURER¹I.1 (volume IV of the ECD).

⁷ On the notion of LF granularity, see Kahane and Polguère (2001).

are full-fledged lexical units and, therefore, have a lexicographic definition of their own.⁸

3. *Types of LF-definition links*

According to the very definition of the notion of standard LF, the meaning of such LFs as **Magn**, **Ver**, **Bon**, **Real**₁, etc. has to remain extremely vague in order to isolate general and universal lexical relations. This means that each standard LF is associated with one “basic” meaning: ‘high degree’ for **Magn**, ‘as it should be’ for **Ver**, etc. However, it is often necessary to increase the granularity of LF encoding in order to account for actual lexical links: see, for instance, the use of superscripted and subscripted indices on LF names present in ECD entries. We see no contradiction in the fact that the system of standard LFs postulated by the theory may be different from the system of standard LFs actually used in lexicographic descriptions, the latter having a greater granularity. The first system reflects universal properties of the semantic-syntax interface in natural languages; the second system reflects the lexical structuring of particular natural languages.

In what follows, we focus on **Magn**, the prototypical representative of the **Magn** LF family. We do not propose to increase the granularity of the LF system by identifying different, more specific senses of **Magn**⁹. Rather, we propose different standard means of connecting its basic meaning (‘high degree’) to the definition of the lexeme that is the base of the collocation described by means of **Magn**. We will postulate that each type of connection, for any meaningful syntagmatic LF, can be characterized based on three parameters:

- the LF basic meaning (‘high degree’ for **Magn**);
- the pattern of connection, which can be direct or indirect;
- the target, which is the semantic component of the definition of the base where the hook up actually takes place.

Taking into consideration these parameters, we have identified the following types of **Magn** connections.

⁸ On the writing of lexicographic articles for lexical units that function as collocatives, see Alonso Ramos (2003).

⁹ Apresjan (2001) proposes to distinguish between two different basic meanings for **Real**₁, each of which combines with semantically distinct lexical units: artefacts, on the one hand, and lexical units that incorporate a modal meaning, on the other hand.

Table 1. Types of **Magn** connection to lexicographic definitions

I	Direct connection of Magn: ‘high degree’ —1→‘⟨base meaning⟩’. This is so to speak a “pure” Magn. I.1 With ‘high degree’ connected to the central component of the base definition. <u>Example</u>: <i>forte pluie</i> ‘heavy rain’. The meaning ‘high degree’ is directly connected to ‘fall’, which is the central component of the definition of <i>PLUIE</i> [≈ ‘fall of drops of water from the sky...’]. I.2 With ‘high degree’ connected to an embedded component of the base definition. <u>Example</u>: <i>grave accusation</i> ‘grave/serious accusation’. Given that an accusation is the communication by X that Y is responsible for Z, which is something bad, this collocation can be paraphrased as <i>accusation whose Z is very bad</i>. Clearly, the meaning ‘high degree’ is here directly connected to the embedded component ‘[Z is] bad’.
II	Indirect connection of Magn: ‘high degree’ —1→‘⟨intermediate meaning⟩’—1/2/...→‘⟨base meaning⟩’. In such cases, one can say that a specific perspective on the base meaning is taken. II.1 With intermediate meaning connected to the central component of the base meaning. <u>Example</u>: <i>haine tenace</i> lit. ‘persistent hatred’. The meaning ‘high degree’ connects to the intermediate meaning ‘duration’, which itself connects to the central component ‘negative feeling’. II.2 With intermediate meaning connected to an embedded component of the base meaning. <u>Example</u>: <i>pluie fine</i> ‘thin rain’, which is here a case of AntiMagn. The meaning ‘low degree’ connects to the intermediate meaning ‘dimension’, which itself connects to the embedded component ‘drops’.

Note. Each specific type of connection identified above could eventually be divided further into two subtypes: with or without the addition of a meaning to ‘high degree’. An example of type **I.1** with extra meaning added to the basic meaning of **Magn** is *pluie drue/serrée* ‘dense rain’. This collocation can be paraphrased as ‘heavy rain, with drops of water being close to each other’.

What is really new in the Table 1’s typology of LF-definition links is the notion of indirect link (type **II**). A **Magn** that indirectly applies to the base meaning is more than just an intensifier: it offers a special perspective on the situation denoted by the base. Prototypical

perspectives are: dimension, duration, quantity, ways of doing (emotion, energy...), etc. See, for instance, our analysis of the **(Anti)Magn** collocations controlled by ASSURER^{I.1} in section 4 below.

4. Encoding of LF-definition connections

Indications on how syntagmatic LFs connect to the base definition can already be found, though in a non systematic way, in published ECDs. We refer here to subscripted and superscripted indices that often follow the names of standard LFs. Subscripted indices have been used in ECDs to encode target of syntagmatic LFs that are not the central component of the base definition. For instance, *pluie fine* (lit. ‘thin rain’) has been encoded as **AntiMagn**_[‘gouttes’] in Mel’čuk *et al.* (1988). Superscripted indices (_{quant}, _{dim}, _{temp}) have been used to specify intermediate meanings, but without lexicographers being aware of the specific “linking” nature of these meanings: they function as semantic bridges between the LF basic meaning and the meaning of the collocation’s base. For instance, the example given above for type **II.1** of link — *haine tenace* — is encoded in Mel’čuk *et al.* (1984) as **Magn**^{temp}.

We propose to systematically encode LF-definition connections identified in section 3 by means of the following conventions¹⁰.

Table 2. Formal conventions for encoding **Magn** connection to lexicographic definitions

Connection	Patterns of encoding	Sample collocations	LF encoding
I.1	⟨LF⟩	<i>forte pluie</i>	Magn
I.2	⟨LF⟩[⟨Target⟩]	<i>grave accusation</i>	Magn [Z is bad]
II.1	⟨LF⟩-⟨intermediate meaning⟩	<i>haine tenace</i>	Magn-duration
II.2	⟨LF⟩-⟨intermediate meaning⟩[⟨target⟩]	<i>pluie fine</i>	AntiMagn-dimension [drops]

The above formalization, with its well-specified semantics, is already an improvement, in terms of accuracy, when compared to traditional ECD encoding. It can be made even more rigorous once we use definitions written in the BDéf format. Because these definitions are precisely structured in terms of elementary propositions and definitional blocks, it

¹⁰ As one can see in Table 2, no indication is given about the target when it is the definition’s central component (cases **I.1** and **II.1**).

is possible to refer directly to these propositions and blocks while encoding syntagmatic LF links. We will demonstrate this by encoding the **Magn** and **AntiMagn** links controlled by ASSURER¹I.1, whose BDéf definition has been given in section 2.

Table 3. Encoding of (**Anti**)**Magn** connections to the BDéf definition of ASSURER¹I.1 (given in section 2)

collocatives	BDéf definition-based encoding
<i>avec véhémence</i> lit. 'with vehemence'	Magn-emotion&energy
<i>avec force</i> lit. 'with strength'	Magn-energy
<i>fermement, catégoriquement</i> lit. 'firmly' 'categorically'	Magn-conviction [Informational content: Prop. 2]
<i>mollement</i> lit. 'softly'	AntiMagn-energy
<i>timidement</i> lit. 'shyly'	AntiMagn-conviction [Informational content: Prop. 2]

5. Conclusion

Results presented here are only the initial step toward a better integration of lexicographic definition of lexical units and LF formulas encoding collocations these lexical units control. There are of course many issues that should be studied further, such as logical links between the content of BDéf definitions and the presence of certain patterns of LF-definition connection.

Here are two examples of such logical links:

1. Some semantic families of lexical units will tend, by their very nature, to control indirectly connected **Magn** collocatives. For instance, verbs and nouns of communication have a non gradable central component ('say/communicate'), which cannot be the direct target of 'high degree'. Therefore, **Magn** for such units either connects to an embedded component or, more often, connects to the central component through the mediation of an intermediate gradable meaning (see **Magn** and **AntiMagn** of ASSURER¹I.1 in Table 3 above).

2. Many lexical units will have in their BDéf definition a non-central defining block entitled *Potential effect: OURAGAN* 'hurricane', *BLESSURE* 'wound', *COLERE* 'anger', *PEUR* 'fear', etc. For all these units,

one can predict the existence of **Magn** collocatives that target this specific block: *ouragan dévastateur*, *blessure mortelle*, *violente colère*, *peur panique*, etc. However, it is also necessary to consider here the interrelation between **Magn**, **Ver** and the **Fact** family: if the potential effect of L is taking place, then we are in presence of a “true” L (**Magn**= **Ver**), or an L that actually “functions” (**Magn**=**A₁Fact₀**) — see Kahane (2003).

This latter observation shows that the present study has to be extended to the whole system of standard syntagmatic LFs.

References

- Alonso Ramos M. (2003) La nature des collocatifs: leur statut en tant qu’unités lexicales. In F. Grossman & A. Tutin (eds.): *Les Collocations. Analyse et traitement*, Amsterdam, De Werelt, 2003. P. 45—60.
- Altman J., Polguère A. (2003) La BDéf: base de définitions dérivée du Dictionnaire explicatif et combinatoire, *Actes de la première conférence internationale de théorie Sens-Texte (MTT 2003)*, Paris, 2003. P. 43—54.
- Apresjan Yu. D. (2001) O leksičeskix funkcijax semejstva REAL-FACT (About lexical functions of type REAL-FACT). In W. Gruszczynski (ed.): *Nie bez znaczenia... (Prace ofiarowane profesorowi Zygmuntowi Saloniemu z okazji jubileuszu 15000 dni pracy naukowej)*, Białystok (Poland): Wydawnictwo Uniwersytetu v Białym Stoku, 2001. S. 23—40.
- Barque L., Polguère A. (2004) A definitional metalanguage for explanatory combinatorial lexicography. Unpublished presentation made at the *International Colloquium on Word Structure and Lexical Systems: models and applications*, Pavia (Italy), 2004. P. 16—17.
- Hausmann F. J. (1979) Un dictionnaire des collocations est-il possible ? *Travaux de littérature et de linguistique de l’Université de Strasbourg*, XVII: 1, 1979. P. 187—195.
- Kahane S. (2003) Une “blessure profonde” dans le Dictionnaire Explicatif et Combinatoire : Sur le lien entre la définition lexicographique et les fonctions lexicales. In F. Grossman & A. Tutin (eds.): *Les Collocations. Analyse et traitement*, Amsterdam, De Werelt, 2003. P. 61—73.
- Kahane S., Polguère A. (2001) Formal foundation of lexical functions. In *Proceedings of the Workshop “COLLOCATION: Computational Extraction, Analysis and Exploitation”*, 39th Annual Meeting and 10th Conference of the European Chapter of the Association for Computational Linguistics, Toulouse (France), 2001. P. 8—15.
- Mel’čuk I. (1995) The Future of the Lexicon in Linguistic Description and the Explanatory Combinatorial Dictionary. In I.-H. Lee (ed.): *Linguistics in the Morning Calm 3 (Selected Papers from SICOL-1992)*, Seoul, 1995. P. 181—270.

- Mel'čuk I. (1996) Lexical Functions: A Tool for the Description of Lexical Relations in the Lexicon. In L. Wanner (ed.): *Lexical Functions in Lexicography and Natural Language Processing*, Amsterdam/Philadelphia: Benjamins, 1996. P. 37—102.
- Mel'čuk I. (1998) Collocations and Lexical Functions. In A. P. Cowie (ed.): *Phraseology. Theory, Analysis, and Applications*, Oxford: Clarendon Press, 1998. P. 23—53.
- Mel'čuk I. *et al.* (1984, 1988, 1992, 1999) *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques, vol. I—IV*, Montréal, Les Presses de l'Université de Montréal.
- Mel'čuk I., Clas A., Polguère A. (1995) *Introduction à la lexicologie explicative et combinatoire*, Louvain-la-Neuve, Duculot, 1995.
- Polguère A. (2000) Towards a theoretically-motivated general public dictionary of semantic derivations and collocations for French. *Proceedings of EURALEX'2000*, Stuttgart, 2000. P. 517—527.

Between Inflection and Derivation Paradigmatic Lexical Functions in Morphological Databases

Maarten Janssen

ILTEC, Lisboa, Portugal
maarten@janssenweb.net

This article shows the use of lexical functions in the MorDebe lexical database system. Lexical functions are used in MorDebe to bridge the gap between derivation and inflection — to model relations between word-forms that are derivational from one perspective but inflectional from another. However, lexical functions in MorDebe range over lemmas rather than word-meanings and are hence significantly different from the normal lexical function.

Между словоизменением и словообразованием: парадигматические лексические функции в морфологических базах данных. В этой статье обсуждается использование лексических функций в морфологической базе данных MorDebe. Лексические функции используются в MorDebe для того, чтобы сократить разрыв между словообразованием и словоизменением — для моделирования отношений между словоформами, которые можно рассматривать с одной точки зрения как деривационные, а с другой — как словоизменяемые. Однако, лексические функции в MorDebe определяются скорее для лексем, т.е. словарных входов, а не для смысловых единиц и, следовательно, отличаются от лексических функций в обычном понимании.

1. Introduction

Despite its central role in linguistics, the distinction between inflection and derivation is far from clear-cut. Modern morphological theories often posit in-between cases — behaving partly like inflections, but partly like derivations. However, the introduction of marginal cases is problematic from the perspective of morphological database: an actual database which lists inflections and/or derivations, either has to treat them on a par, or model any individual form either as derivational or inflectional. This article will show how this problem manifests itself in the MorDebe morphological database, and how lexical functions are used to (partially) overcome the problem.

This article consists of three parts: the first part will show the set-up of the MorDebe database and its treatment of derivation and inflections. And the second part will illustrate the problem of distinguishing inflection from derivation from the perspective of the database, and how lexical functions are used to remedy this. But the application of lexical functions in this way implies a fundamentally different use of lexical functions: lexical functions as ranging over lemmas rather than word senses. The third part of the article will show how this affects lexical functions, and discusses the problems of this use of lexical functions.

2. MorDebe, OSLIN and SIMuLLDA

MorDebe is a large volume morphological database, set up language independently, but currently only used for (European) Portuguese. The database currently contains some 125.000 Portuguese lemmas, and around 1,5 million word-forms. MorDebe is integrated with a semi-automatic neologism detection system called NeoTrack, which keeps the database up-to-date and ever expanding. MorDebe models only the formal properties of words — and currently contains no semantics. But the MorDebe database is designed as the formal part of the SIMuLLDA system (Janssen, 2002): an interlingual lexical database system focusing primarily on semantic content.

SIMuLLDA has a layered set-up, distinguishing different lexical levels. The basic lexical unit is roughly the headwords in dictionaries — referred to as the lemma¹. All lemmas are related to a set of inflectional forms on the one hand, and a set of word-senses or meaning on the other. The lemma itself is strictly an abstract, formal unit, adorned with a citation form mostly for practical purposes.

The lemma is defined by its inflectional paradigm — if two lemmas have different inflectional paradigms, they have to be considered different lemmas, as in the case of the Dutch noun *band* which can have *banden* (tires) or *bands* (bands) as its plural — or the Portuguese verb *redar* which can be inflected as *redou* (I give again) or *redo* (I catch with a net). However, within an inflectional paradigm there can be alternative forms: the first person singular present tense of *ouvir* (to hear) can be either *ouço*

¹ In SIMuLLDA, the basic entries are actually referred to as *lexemes*, whereas the *lemma* is the set of word-forms under the lexeme — but to avoid confusion with the sense-dependent notion of lexeme in MTT, the term *lemma* is used in this article.

or *oiço*. The word-forms in the inflectional paradigm have an orthographic representation, an inflectional code, and possibly a dialectic restriction.

Every lemma can express one or more word-senses. The word-senses are stored in a separate database, explicitly linked to the lemma by means of unique identifiers. Apart from meanings, it is possible to relate other kinds of lexical information to the MorDebe database in the same fashion: pronunciation data, morphological and grammatical descriptions, etc. By linking all these different types of information to the same set of lemmas, the different databases form a network of lexical information, in which every lexical resource could be maintained by a different institute. At the moment, the idea of using MorDebe for distributed lexical databases is being developed into a proposal for an open source lexical information network (OSLIN).

One of the databases currently under development for Portuguese is a database of derivations. Derivations are modelled as relations between lemmas. Like inflectional forms, derivations are stored explicitly rather than being generated by rules. The fact that inflections are stored as part of the lemma, whereas derivations are modelled as relations between different lemmas implies that there is a clear distinction between the treatment of derivations and inflections in MorDebe. In the case of a number of morphological relations, this leads to a problem, since certain forms should be considered inflectional from one perspective, but derivational from another.

3. Between Inflection and Derivation

Inflectional morphology and derivational morphology are two traditionally distinct fields, attributed many types of differences — such as the claim that derivation is pre-grammatical, whereas inflection is part of the grammar itself (Split Morphology, Perlmutter 1998). Various distinctive traits have been formulated by which inflection can be distinguished from derivation. A good summary of the distinctive traits between inflection and derivation is given by Booij (*to appear*): (1) inflections are obligatory and fully productive, (2) inflection does not change the lexical category, (3) derivation has no impact on grammar, (4) inflection is paradigmatic, (5) inflection produces no semantic change.

But it has been argued by many authors, including Schultink (1962), Bybee (1985), and Booij (1995), that none of the proposed criteria define a clear-cut division between derivation and inflection. Proposed solutions to this lack of clear distinctions include the introduction of an in-between

category (Booij, 1995), or a class of in-between cases (Bauer, 2004), viewing the distinction as a gliding scale (Bybee, 1985), or drop the distinction altogether (Lieber, 1980).

This section does not discuss the differences between inflection and derivation in detail, but mainly illustrates the problems of a clear distinction from the perspective of that which is most relevant for lexical databases: inflection creates word-forms, whereas derivation creates new lemmas.

3.1. *The derivational nature of inflection*

There are clear examples of inflectional forms: gender and number on adjectives is inflectional in every sense of the word — the realisation of the adjective is completely determined by the noun it applies to (which is why Booij 1995 calls this *contextual inflections*), and the form of the adjective itself has no semantic, syntactic, or even pragmatic implication. But many other inflectional forms are less clear — those that Booij (1995) calls the *inherent inflections*: inflection that is “*syntactically relevant but not constrained by sentence structure*” (Bauer, 2004).

Inherent inflection has many derivation-like features. Inherent inflection forms often have a clear lexical meaning. For instance, female forms of nouns refer to female rather than male entities; if you play a piano piece *piano* it is not the same as when you play it *pianissimo*. And inflection is said to always relate to lemmas — not to meanings, but often inflectional forms only apply to specific meanings of a word: the female form *amarela* is only related to the word *amarelo* in its meaning of *pale person*, not in its more common meaning *yellow*.

Inflection should be paradigmatic, but diminutives, plurals and female forms can stack: the form *pequeninas* is the plural of the female form of the diminutive of *pequeno*. The superlative in Portuguese has the full paradigm of gender and number, and for languages with a richer nominal inflection, such as Russian, the number of inflectional forms can really explode because of this stacking.

Various types of morphological variants that are often considered inflectional are far from obligatory: the superlative only exists for gradable adjective, which is the minority of adjectives; and the female forms of nouns only exist for nouns denoting animate objects, which is an even smaller group.

Derived forms, being lemmas by themselves, can attract meanings of their own: the fact that a *movement* can be an ideological group is not related to it's being a derivation of *move*. But for inflections this should not

hold. However, the word *cadeirinha* — which is the diminutive of *cadeira* (chair), can also mean a kind of saddle made of wood, or a *rickshaw* (in Macau). And the word *glasses* is no longer just the plural of *glass*.

3.2. *The inflectional nature of derivation*

There are several morphological processes that are almost completely inflectional in nature, but are still considered derivational, merely because they are category changing. For instance, various types of derivations are so productive that they are almost obligatory — almost any adjective has an adverbial form, and these adverbial forms are highly predictable as well: except for a small group of exceptions, adverbs in English are formed by putting *-ly* at the end of the adjective (with the appropriate phonological modifications); and in Dutch the adverbial form is simply identical to the adjective.

The nouns or adverbs derived from adjectives are hardly different from their original form in meaning — even more so if the derived noun is used attributively, or if the adjectival form applies to a deverbal noun: there is little difference between the sentences *he moved slowly* and *his movement was slow* except for a change in perspective. In the case of the adjectival use of past participles, it is sometimes not even clear whether a particular use of the term is a past participle use, or an adjectival construction.

Given these considerations, Haspelmath (1995) proposes to expand the class of inflections with word-class changing relations, or *transpositional inflection*: a process that is “*inflectional in the sense that it is regular, general, and productive, but nonetheless transpositional*” (Bauer, 2004). In the remainder of this article, *inherent inflection* will be used as possibly including transpositional inflections.

3.3. *Paradigmatic Lexical Functions*

The above discussion illustrates the fact that there is no clear boundary between inflection and derivation — the question whether a given relation should count as derivational or inflectional depends at least partly on the purpose of the question. But this duplicity creates a tension on their treatment in databases like MorDebe. On the one hand, for the purpose of consultation, it would be most useful if female nouns and superlatives were modelled paradigmatically as part of their root. But since only lemmas can be related to meanings, they should be listed as separate lemmas.

The use of a paradigmatic lexical function can very nicely relieve this tension. To take the female nouns as an example: the words *actor* and *actress* can be listed as separate lemmas, both with their own singular and plural forms. The two distinct lemmas are then explicitly linked by a lexical function (the lexical function is taken from the DECIDE project: **female**). This explicit linking of the lemmas makes it possible to generate a larger paradigm for *actor* including both the male and female forms if the application requires it — the use of a functional link makes it possible to shift between an inflectional and a derivational perspective. The use of lexical functions in this fashion can also nicely account for the stacking nature of inherent inflection because of the recursiveness of lexical functions.

In MorDebe, the set of relations that is considered strictly inflectional is largely reduced in this manner: only the number and gender for adjectives; tense, aspect, number and person for verbs; and number for nouns (despite its problems). All other modifications, some of which are traditionally taken as inflectional, are modelled as separate lemmas — these include female forms of nouns, superlatives, diminutives, but also deverbal nouns, deadjectival adverbs, etc.

In the Portuguese MorDebe database, all such inherent inflections are being modelled as derivational relations between different lemmas, where the relation between the lemmas is modelled by means of lexical functions. For practical purposes, the **s0** are separated into deadjectival nouns (**s0a**) and deverbal nouns (**s0v**); the LF's currently used are: **s0a**, **s0v**, **adv0**, **a2**, **able2**, **max**, **min**, **dim** (diminutive), **aum** (augmentative), **female**, and **alt** (for orthographic variants: **alt**(*dourar*) = *doirar*). For Portuguese, about 15.000 LF relations have been recorded thus far, and all words that are added to the database are added together with their inflectional relations when applicable. This makes MorDebe one of the few large-scale LF databases in existence.

4. Lexical Functions over lemmas

Despite the attractiveness of the use of lexical functions as described in the previous section, there is a problem with it: the lexical functions are applied to *lemmas*, whereas LF's are by default defined over word-senses. Lexical functions are normally semantic derivations in at least three senses: firstly, they are not form based relations, since “*no morphological link needs to exist between the units involved*” (Polguère, 2000). Secondly, the relation they express is a semantic relation — even in the case

of the so-called syntactic functions: **s0** expresses a nominalisation *with the same meaning* as its original. And thirdly, their arguments are word-senses rather than lemmas.

This last point is a direct consequence of the definition of a lexical function: lexical functions are defined as relations of *lexical units*, which in MTT are either *lexemes* or *phrasemes*, and a *lexeme* is “a word taken in one well-specified sense and supplied with all the information fully determining its behavior when it is used in this sense” (Mel’cuk, 1993, p. 250)². The word-sense nature of LF’s is easy to see: **Magn**(*hand*) = *dab*, but only in its meaning of “someone with skill”.

4.1. Inflectional Lexical Functions

An important difference between derivations and inflections is that inflections operate on the level of lemmas: the past tense *jumped* is not the past tense of a specific meaning of *jump*. Inflectional paradigms belong to the lemmas, and not to the word-senses. Because of this, the inflection-like derivations discussed in the previous section also have to be modelled over lemmas, and not over word senses. Hence, the lexical functions described in 3.3 range over lemmas — they are a different type of lexical functions, which will be called *inflectional functions*³. Given the fact that their arguments are entities of a different type, inflectional functions are strictly speaking not lexical functions at all, but different kinds of relations. This despite the fact that they are modelled in the same way as lexical functions, even use the same function names where available.

Inflectional functions are in a sense midway between morphological derivations and the semantic derivations used in MTT. On the one hand, they represent semantic relations, and not morphological processes. But on the other hand, the functions relate to lemmas, and not to their meanings. This is best illustrated by some examples.

In Portuguese, the derivational suffix *-mento* turns verbs into nomina — modelled by the inflectional function **s0v**: **s0v**(*arrolhar*) = *arrolhamento* (corkage). But it is not the morphological derivation that is mapped by **s0v**, since the same function also maps different patterns: **s0v**(*canalizar*) = *canalização* (canalisation), **s0v**(*adastrar*) = *adastragem* (straightening), **s0v**(*assar*) = *assadura*, **s0v**(*abafar*) = *abafação* = *aba-*

² In this article, the term *word-sense* will be used for sense-specific lexical units to avoid confusion with the sense-independent use of the term *lexeme*.

³ The term *inflectional lexical function* was dropped on the suggestion of one of the reviewers.

famento = *abafadura* = *abafadela* (suffocation). The relation between the argument and value does not even have to be a productive derivational process: **s0v**(*propor*) = *proposição* (to propose), **s0v**(*fundir*) = *fundição* = *usão* (foundation). In certain cases, there does not even have to be a morphological relation between the two: **max**(*bom*) = *ótimo* (good; best) — similar to irregular inflections: *went*, *mice*.

The other way around, not all relations which look derivational are modelled as such: although *dentadura* could be the **s0v** of *dentar* (to bite) — it is not. It is morphologically derived from *dentado* (toothed) and means ‘set of teeth’ or ‘denture’. And even though *embocadura* (mouth of a river) is morphologically the nominalisation of *embocar* (to put to the mouth), it is not the **s0v** of it, because semantically, it is not the *action or effect of embocar*. In English, *ignorance* is no longer transparently related to *ignore* (Bauer, 2004).

But inflectional functions are not as semantic as the (normal) lexical functions. They have to be relations over the lemma and not its meaning: *abafação* (suffocation) can never be the **s0v** of *sufocar* (suffocate), despite the fact that *abafação* expresses the nominal content of *sufocar*. Although the relation does not have to be strictly derivational, there has to be a strong paradigmatic link between the value and the argument.

And inflectional functions have to be productive over the entire lemma, and not restricted to one of its senses. For instance, *embicadura* is not the **s0v** of *embicar*, since it is only the nominalisation in a very specific sense: ‘heading into a port’. Although it should be said that the relation can be blocked by certain word-senses: **female**(*amarelo*) = *amarela*, even though this female form can only be used for animate reading of the word (pale person), in the same sense that the plural of *agua* (water) can only be used in count-noun readings. Inherent inflections are more likely to be partially defective than contextual inflections.

4.2. Asymmetry

Although the inflectional functions operate on lemmas, the derived term is clearly not a lemma: although the way of expressing the **s0a** for *agudo* (sharp; biting; acute; serious; keen; in-depth) is *agudo* rather independently of its meaning, the reverse is not true: although *promoção* can be the **s0v** of *promover*, it also has a number of other meanings, including *sales*. Only in a specific meaning of the word is the derived form the **s0v** form of the root — although it might of course happen to be the only meaning.

The derivational meaning does not even have to be a prominent meaning of the derivate: the **s0a** of *cavo* (*hollow_A*) is *cavidade*, but the

meaning *hollowness* is only a marginal meaning of *cavidade*: it much more prominently means *cavity* (or *hollow_N*), which bears the same root as *cavo*, but is not derived from it — it is an adapted form of the Latin word *cavitate* (*cavity*).

The fact that the derived form can have other meanings besides being the transparent derivation is not surprising, and is in fact one of the reasons why inflectional functions are treated as derivations (in the sense of generating new words) in the first place. But it does mean that inflectional functions are asymmetrical: they take lemmas as their arguments, but yield word-senses as their values. This has to be taken into account in the interpretation of inflectional functions: **female**(*pensador*) = *pensadora* expresses the fact that the word for referring to a female *pensador* (thinker) is *pensadora*, not that the word *pensadora* is identical to the female form of *pensador*.

5. Conclusion

The problem with the gradual distinction between derivation and inflection in a morphological database can nicely be bridged by using lexical functions: because of the use of LF's, MorDebe can provide a rich set of paradigmatic information for every lemma in the database, including contextual inflections, but also inherent inflections and even category changing "inflections".

Lexical functions for this purpose (inflectional functions) are fundamentally different from normal lexical functions in that they range over lemmas rather than word-senses. However, for those inflectional functions that have an LF counterpart, there is a strong relation between the two: an inflectional function over a lemma implies that the corresponding LF holds for all related word-senses.

References

- Bauer, Laurie. 2004. The function of word-formation and the inflection-derivation distinction. In: Aertsen, Hannay & Lyall (eds.) *Words and their Places. A Festschrift for J. Lachlan Mackenzie*. Amsterdam: Vrije Universiteit, 2004.
- Booij, Geert. 1995. Inherent versus contextual inflection and the split morphology hypothesis. In: Booij & van Marle (eds.) *Yearbook of Morphology 1995*. Dordrecht: Kluwer, 1995.

- Booij, Geert. Inflection and derivation. In: Brown (ed.) *Encyclopedia of Language and Linguistics*. Oxford: Elsevier. *To appear*.
- Bybee, Geert. 1985. *Morphology*. Amsterdam: John Benjamins, 1985.
- Haspelmath, Martin. 1995. Word-class-changing inflection and morphological theory. In: Booij & van Marle (eds.) *Yearbook of Morphology 1995*. Dordrecht: Kluwer, 1995.
- Janssen, Maarten. 2002. *SIMuLLDA: a Multilingual Lexical Database Application using a Structured Interlingua*. PhD Thesis, Utrecht University, 2002.
- Lieber, Rochelle. 1980. *On the organisation of the lexicon*. PhD Thesis, MIT, 1980.
- Mel'cuk, Igor A. 1993. The Future of the Lexicon in Linguistic Description. In: Ik-Wan Lee (ed.) *Linguistics in the Morning Calm 3: Selected papers from SICOL-1992*. Korea: Seoul, 1993.
- Perlmutter, David. 1988. The Split Morphology thesis: evidence from Yiddish. In: Hammond & Noonan (eds.) *Theoretical Morphology*. San Diego: Academic Press, 1988.
- Polguère, Alain. 2000. Towards a theoretically-motivated general public dictionary of semantic derivations and collocations for French. In: Heid et al. (eds.) *Proceedings of the Ninth EURALEX International Conference*. Stuttgart: IMS, 2000.
- Schultink, Hendrik. 1962. *De morfologische valentie van het ongelede adjectief in modern Nederlands*. Den Haag: van Goor & Zonen, 1962.

Meaning-Text Unification Grammar: Modularity and Polarization

Sylvain Kahane,

Modyco, U. Paris 10 /
Lattice, U. Paris 7
sk@ccr.jussieu.fr

François Lareau

OLST, U. de Montréal /
Lattice, U. Paris 7
francois.lareau@umontreal.ca

This article presents the Meaning-Text Unification Grammar's current state, now that its formal foundations have been clarified with the development of Polarized Unification Grammar. Emphasis is put on the model's architecture and the role of polarization in linking its modules — semantics-syntax interface and the well-formedness grammars of each representation level.

1. Introduction

This article presents the current state of Meaning-Text Unification Grammar [MTUG], building on the recent development of Polarized Unification Grammar [PUG]. PUG can simulate most formalisms that are based on structure unification, such as rewriting grammars, TAG, LFG or HPSG (Kahane, 2004). However, although PUG was initially developed as a formal basis for MTUG, how it can actually be used for writing an MTUG has never been discussed.

MTUG's architecture is based on the Meaning-Text Theory [MTT] (Mel'čuk 1988; Kahane 2001b). Fewer levels of representation are considered in the current version of MTUG: semantic, syntactic, morphotopological and phonological. Each one has its own grammar. A semantic structure is a graph, a syntactic structure is a dependency tree, a morphotopological structure is an ordered tree (Gerdes & Kahane 2001) and a phonological structure is a chain. On these can be added other structures such as a logical or communicative structure at the semantic level or else a prosodic structure at the phonological level, but we will not discuss these points here. These representation levels are ordered (semantic < syntactic < morphotopological < phonological), which yields three interface modules: semantics-syntax, syntax-morphotopology and morphotopology-phonology).

MTUG therefore requires a formalism that can handle several types of structures (graph, dependency tree, ordered tree, chain) and link them. As we will see, PUG is a mathematical formalism tailored specifically for handling such structures as simply as possible. It relies on polarities that have to be neutralized for driving sentence synthesis or analysis in a way similar to how atoms' valence controls molecule formation. Our implementation of MTUG as a PUG makes extensive use of polarization: each module (well-formedness grammars as well as interface grammars) controls the construction of objects and the saturation of structures only via polarization. Each has its own polarity and uses the polarity of the adjacent modules to control its interaction with them. The order in which those polarities are neutralized induces a particular procedure. Neutralizing all polarities of a given level before considering those of the next level induces a sequential synthesis/analysis procedure (the one typically used in classic MTT implementations), whereas neutralizing all the polarities introduced for a given word before considering other polarities allows for an incremental synthesis/analysis procedure, despite a stratified architecture.

Due to the lack of space, we will leave aside the morphotopological and phonological levels as well as their interfaces, and we will focus only on the semantic and syntactic structures and the semantics-syntax interface (see Kahane & Lareau 2005, for a brief discussion of the morphotopological structure and its interface with syntax in PUG). We will briefly introduce PUG in section 2. Section 3 will introduce well-formedness grammars for the structures considered in this paper. Finally, in section 4 we will show how the PUG formalism can be used for writing MTUG interface grammars.

2. Polarized Unification Grammars [PUG]

Polarized Unification Grammars are devices that generate sets of finite structures. A structure is made of *objects*. For instance, an oriented graph is made of *nodes* and *arrows*. Each arrow is linked to two nodes by the functions *source* and *target*. It is these *functions* that provide the structure proper.

A *polarized structure* is a structure whose objects are polarized, *i.e.* linked through a function to a value that belongs to the finite set P of *polarities*. The set P is provided with a commutative and associative operation called *product* (noted “ $\cdot$ ”). A subset N of P contains *neutral polarities*. A polarized structure is said *neutral* if all of its objects are neutral. We will use in this paper a set $P = \{\bullet, \circ, \blacklozenge\}$ (we will call these polarities:

● = black = saturated, ○ = white = obligatory context and ● = grey = absolute neutral), with $N = \{●, ●\}$, and a product defined in the following table ($\perp$ indicates that two polarities cannot combine):

•	●	○	●
●	●	○	●
○	○	○	●
●	●	●	$\perp$

Table 1. The product of polarities

Structures are combined by *unification*. Unifying two structures A and B gives a new structure $A \oplus B$ by matching some of the objects of A with some of the objects of B. When A and B are unified, the polarity of an object of $A \oplus B$ that was obtained by matching an object of A with one of B is the product of the polarities of these two objects. All functions of objects that are unified must themselves be unified (just as the features when we unify feature structures).

A *polarized unification grammar* [PUG] is defined by a finite set T of object types (with functions attached to each kind of object type), a polarity system $(P, \cdot)$, a neutral polarities subset N of P and a finite set of elementary polarized structures, whose objects are described by T and of which one can be marked as an initial structure. The structures *generated* by the grammar are the neutral structures obtained by combining the initial structure and/or a finite number of elementary structures. This formalism is monotonic (with the following order on polarities: $● < ○ < ●$) and structures can be combined in absolutely any order (due to the commutativity and associativity of $(P, \cdot)$).

We will introduce in the next section two examples of PUG that respectively generate semantic graphs and syntactic trees. Examples throughout this paper will be based on (1):

(1) *Peter ate two apples.*

3. Well-formedness grammars

3.1. Semantic grammar

Our semantic representations are based on a graph of predicate-argument relations between semantemes (*i.e.* meanings of lexical and

grammatical units of a sentence). One could superimpose on it other kinds of structures that encode various types of information, such as an information structure or a logical scope structure (Polguère 1992, Mel’cuk 2001, Kahane 2005). The nodes of a semantic representation encode semantemes and the edges represent predicate-argument relations. Figure 1 below shows a simplified fragment of $\mathcal{G}_{\text{sem}}$ — a graph grammar which generates the semantic representation of (1). Each of its objects has a polarity, noted p_{sem} in the following discussion¹, that indicates which objects are actually constructed by the rule (objects polarized in black) and which only give the semantic valence of predicates (objects polarized in white).

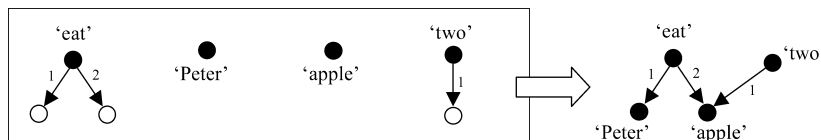


Fig. 1. A fragment of $\mathcal{G}_{\text{sem}}$ generating the semantic representation of (1)

3.2. Syntactic grammar

Our syntactic representations are dependency trees whose nodes are labeled with the name of a lexeme and are linked to other objects representing the grammemes via so-called *grammatical functions*². The edges are labeled with syntactic functions (which are language-specific). The rightmost part of Figure 2 shows the syntactic representation of (1).

Our syntactic grammar $\mathcal{G}_{\text{synt}}$ uses a double polarization, each verifying a different property of syntactic representations. The polarity p_{synt} constructs all the objects and checks that each of them has all the functions required by its type: *label*, *source* and *target* for edges, part of speech (*pos*) for the nodes, and eventual grammatical functions for some

¹ We will refer to polarities by the name of the functions that associate them to the objects. Thus, any object of a rule in $\mathcal{G}_{\text{sem}}$ is associated to a function p_{sem} which returns as its value one of the polarities in the set P described above.

² In our figures, we indicate these functions in *italic*, followed by the grammemes. Grammmemes are in turn linked to the lexemes through opposite functions. This double linking makes lexemes and grammemes inseparable since the unification of one forces the unification of the other. However, we will not discuss this mechanism here.

nodes (*number* for nouns, *mood*, *tense*, *number* and *person* for verbs, etc.)³. The polarity $p_{\text{synt-gov}}$, which we hide in our figures, is used to verify that the generated structure is a tree (*i.e.* that every node except one has a unique governor).

Figure 2 shows a fragment of $\mathcal{G}_{\text{synt}}$ that generates the syntactic representation of (1). The very first rule in this figure is the initial structure, which has to be used once (and only once) and which corresponds to the root of the tree (it has a black $p_{\text{synt-gov}}$ polarity, hidden in Figure 2, indicating that it cannot have a governor). The following four rules are lexical rules; they give the part of speech of lexemes as well as the grammemes that are necessary to them. Note that these rules do not control the syntactic valence of lexemes, as it will be controlled by the semantics-syntax interface. The next three rules are sagittal rules; they describe various possible syntactic relations (they also control the tree structure by having a white $p_{\text{synt-gov}}$ polarity on the governor and a black one on the dependent — again, hidden in Figure 2). The other rules are grammatical rules. They may be context-sensitive: thus, the indicative requires a subject (and vice versa), a numeral imposes the plural on its governing noun, or a non pronominal subject implies the 3rd person on the verb. Note that the introduction of a grammeme might depend on another grammeme: the indicative, for instance, will impose a tense, while infinitive will not.

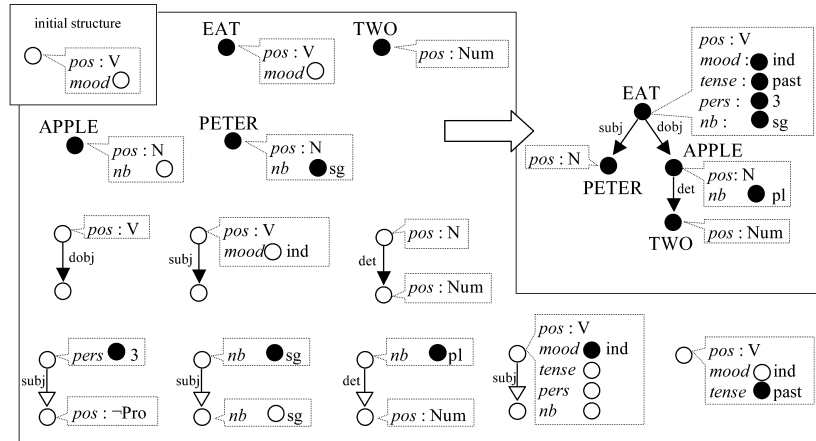


Fig. 2. A fragment of $\mathcal{G}_{\text{synt}}$ generating the syntactic representation of (1)

³ We use a simplified grammatical system, leaving aside, for instance, definiteness for nouns or finiteness and voice for verbs.

4. Interfaces

4.1. Correspondence grammars

A correspondence grammar $\mathcal{G}$ is a grammar that establishes a correspondence between structures belonging to two sets, that we will call $\mathcal{A}$ and $\mathcal{B}$. The rules of $\mathcal{G}$ establish a correspondence between elementary structures that make up the elements of $\mathcal{A}$ and $\mathcal{B}$. There are three possible ways of considering $\mathcal{G}$: as a transductive, equative, or generative grammar, depending on whether it is given one, two or no structures as input (Kahane 2001a). A transductive grammar translates a given structure into another, an equative grammar verifies that two given structures correspond to each other, while a generative grammar creates a pair of structures that correspond to each other.

Let us consider the equative functioning of $\mathcal{G}$, which is the simplest: $\mathcal{G}$ partitions both structures in an equal number of fragments that are in a one-to-one correspondence. In terms of polarities, it means that both structures given as input have to be declared as resources that $\mathcal{G}$ must “consume” (thus ensuring that both structures have totally been put in a correspondence). The objects of both structures bear a white $p_{\mathcal{G}}$ polarity that the rules of $\mathcal{G}$ must neutralize. The rules of $\mathcal{G}$ contain corresponding objects (from the levels of representation of $\mathcal{A}$ and $\mathcal{B}$) that bear a black $p_{\mathcal{G}}$ polarity. Hence, by neutralizing the two given structures, $\mathcal{G}$ establishes a correspondence between them.

Let us now turn to the transductive functioning of $\mathcal{G}$. It is this mode that actually models linguistic synthesis or analysis. We give as input to $\mathcal{G}$ a structure A belonging to $\mathcal{A}$, all objects of which bear a white $p_{\mathcal{G}}$ polarity. This triggers a number of rules from $\mathcal{G}$ (which all contain at least one black element) in order to neutralize A . These rules also build a new structure B , synchronized with A . A well-formedness grammar of the structures of $\mathcal{B}$ must then verify that B is an actual element of $\mathcal{B}$. The structure B , while being neutral for $\mathcal{G}$ (all its objects now bearing a black $p_{\mathcal{G}}$ polarity), must force the application of the grammar of $\mathcal{B}$. In order to do so, all of its objects must also have a white $p_{\mathcal{B}}$ polarity. In sum, every module has its own polarity controlling the construction of the structures it describes, but in order to call the adjacent modules, the objects of $\mathcal{A}$ and $\mathcal{B}$ must have, besides their respective $p_{\mathcal{A}}$ and $p_{\mathcal{B}}$ polarities, a white $p_{\mathcal{G}}$ polarity. We also have to add to the rules of $\mathcal{G}$ a white $p_{\mathcal{A}}$ polarity for the objects of the level of $\mathcal{A}$ and a white $p_{\mathcal{B}}$ polarity for the objects of the level of $\mathcal{B}$. Each object constructed by $\mathcal{A}$ will therefore have a double polarization $p_{\mathcal{A}}p_{\mathcal{G}}(\bullet, \circ)$, while the corresponding elements constructed

by $\mathcal{G}$ will have a double polarization $p_{\mathcal{A}}p_{\mathcal{G}}(\circ, \bullet)$. This yields a system with four polarities $\{(\circ, \circ), (\circ, \bullet), (\bullet, \circ), (\bullet, \bullet)\}$, equivalent to the system $\{\circ, \text{—}, +, \bullet\}$ of Bonfante *et al.* 2004 and Kahane 2004. These couples of polarities are called *articulation polarities*.

4.2. Semantics-syntax interface

The semantics-syntax interface $\mathcal{I}_{\text{sem-synt}}$ is a correspondence grammar between the set of semantic representations described by $\mathcal{G}_{\text{sem}}$ and the set of syntactic representations described by $\mathcal{G}_{\text{synt}}$. While discussing this interface we will illustrate its transductive mode in synthesis. It then takes as input a semantic representation similar to the resulting structure in Figure 1, except for that each object will bear a double polarity $p_{\text{sem}}p_{\text{sem-synt}}(\bullet, \circ)$ or $(\circ, \circ)$. The objects of $\mathcal{I}_{\text{sem-synt}}$ all bear a $p_{\text{sem-synt}}$ polarity (white or black, depending on whether or not they are constructed by $\mathcal{I}_{\text{sem-synt}}$), but those of the semantic level (resp. syntactic) will also have a white p_{sem} polarity (resp. p_{synt}). Figure 3 shows a fragment of the semantics-syntax interface grammar. Only the $p_{\text{sem-synt}}$ polarity is shown here (all p_{sem} and p_{synt} being white). The dotted lines are objects representing the correspondence between objects of adjacent levels and their polarity is shown by the rhombs that are superimposed on them.

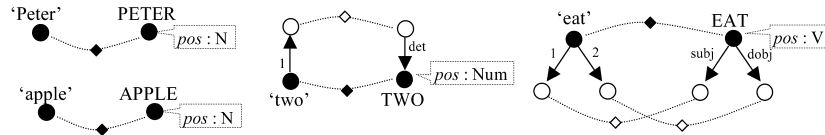


Fig. 3. A fragment of the interface grammar $\mathcal{I}_{\text{sem-synt}}$

In Figure 4, we give three structures: the semantic representation of (1) with the double $p_{\text{sem}}p_{\text{sem-synt}}$ polarization, the result after $\mathcal{I}_{\text{sem-synt}}$ has neutralized all the polarities that it could (*i.e.* $p_{\text{sem-synt}}$ and p_{sem} , but not p_{synt}), and finally the result after applying $\mathcal{G}_{\text{synt}}$ on the syntactic tree constructed by $\mathcal{I}_{\text{sem-synt}}$. The polarities that allow the interaction with the adjacent module are shown as backgrounded. Note that agreement grammemes (such as *person* for verbs) do not have a direct semantic counterpart, therefore they must not have a $p_{\text{sem-synt}}$ polarity.

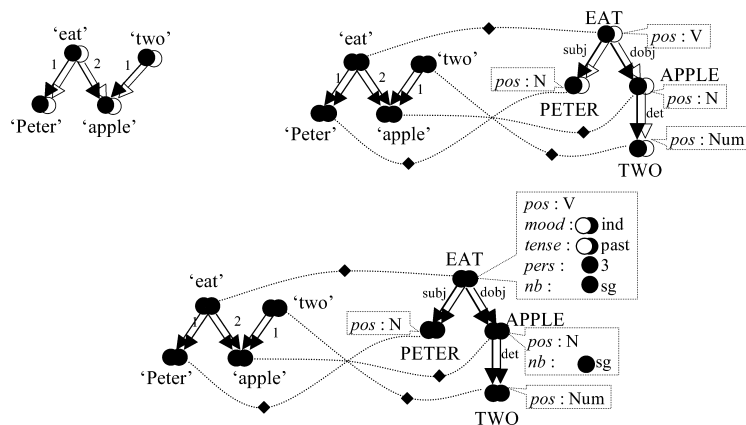


Fig. 4. The semantic representation of (1) before and after the application of $\mathcal{I}_{\text{sem-synt}}$ and $\mathcal{G}_{\text{synt}}$

Just as $\mathcal{G}_{\text{sem}}$, the syntactic grammar $\mathcal{G}_{\text{synt}}$ must be enriched in order to interact with both the semantics-syntax interface and the syntax-morphotopology interface. As the last structure of Figure 4 shows, applying $\mathcal{G}_{\text{synt}}$ on the output of $\mathcal{I}_{\text{sem-synt}}$ neutralizes all the p_{synt} polarities of the syntactic objects created by $\mathcal{I}_{\text{sem-synt}}$. However, it is not possible to introduce the lexeme EAT without assigning a mood to it (*cf.* Figure 2), and if we assign the indicative, for instance, we are then forced to use a tense, as well as agreement grammemes for person and number. The grammar $\mathcal{G}_{\text{synt}}$, while neutralizing the p_{synt} polarities of the output of $\mathcal{I}_{\text{sem-synt}}$, introduces grammemes that were not in this output and which (unless they are agreement grammemes) bear a white $p_{\text{sem-synt}}$ polarity (see Figure 4). These mood and tense grammemes must then be neutralized by rules of $\mathcal{I}_{\text{sem-synt}}$. However, there is no semanteme in the semantic representation given as input that could correspond to these grammemes. This is not a problem with MTUG, since the formalism makes it possible to modify on the fly the input representation in order to satisfy constraints that come from other levels of representation. Let us take for granted that our model includes the two rules shown in Figure 5. The first one is a rule of $\mathcal{G}_{\text{sem}}$ that gives a simple representation of one of the meanings of the past tense. The second rule belongs to $\mathcal{I}_{\text{sem-synt}}$ and establishes the correspondence between this meaning and the past tense grammeme. These two rules can be applied in the opposite direction from the undergoing synthesis process and modify the initial semantic structure, adding to it the meaning of the past

tense (we leave aside the mood for the sake of simplicity). We then obtain the last structure in Figure 5.

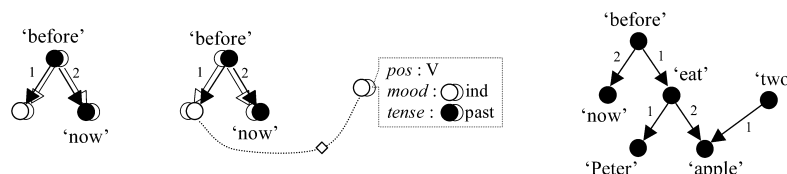


Fig. 5. The past tense

Of course, there are other rules that could satisfy the needs for a mood and a tense, and one would obtain as many modified semantic structures as there are ways to satisfy these syntactic constraints. A text generation system would then have to choose, among all the semantic structures proposed, which is the most suitable according to what it has in its world-knowledge database. In the case discussed here, such a system would have to answer the question “When does Peter’s eating two apples take place: in the present, in the past or in the future?”. This mechanism modelizes in an elegant way the fact that the expression of grammatical meanings is not necessarily triggered by a communicative need from the speaker, but might very well be imposed by the language (see Polguère 2000). This kind of mechanism, in addition to being plausible from a cognitive point of view, might prove particularly useful for machine translation, where the semantic representation obtained by analysis from the source language is not always directly expressible in the target language. It might as well be useful for analysis. It is then the sound chain that is the input and which can be modified if it is not totally well-formed. This might be a useful workaround for problems related to noise or performance flaws from human speakers. However, these hypotheses have not yet been validated through experimentation.

Conclusion

This article mainly focuses on the architecture of a linguistic model. We suggest a modular model (two well-formedness grammars and one interface grammar in the simplified version presented here, but a full-fledged model would include four well-formedness grammars and three interface grammars) that can handle several levels of linguistic representation and several types of structures with a single formalism.

This makes it possible to combine the rules in absolutely any order and allows the various modules to constantly interact with one another. This interaction and the saturation of structures are driven by the (generally multiple) polarization of every object. Each polarity born by an object having a well defined role, two types of algorithms can be considered: neutralizing the polarities specific to one module before considering other polarities would force the modules to apply in a sequential order, while neutralizing all the polarities associated with one object would force each element of the input structure to be processed from semantics to phonology (vice versa in the context of analysis). This second type of algorithm is possible only because one single formalism is used for all the modules and interfaces. We have also shown that polarities allow our model to accept as input underspecified structures lacking inflectional meanings, and still generate correct sentences by adding the missing semantemes to the initial representation. Finally, as Bonfante *et al.* (2004) have shown, polarization allows an efficient filtering of rules by reducing the amount of dead-end computing.

A model of French is currently under development. The heart of French topology has been described by Gerdes & Kahane (2004), the syntax of French tenses is described by Lareau (2004), other aspects of the semantics-syntax interface (mismatches between syntax and semantics and scope of quantifiers) are discussed by Kahane (2001b, 2005). An implementation of MTUG in the PUG formalism is also under process. More than coverage, we put emphasis on theoretical elegance and adequacy of formalization.

References

- Bresnan J. (1999), *Lexical-Functional Syntax*, Blackwell, 1999.
- Bonfante G., Guillaume B. & Perrier G. (2004), Polarization and abstraction of grammatical formalisms as methods for lexical disambiguation, *Proceedings of CoLing*, Genève, 2004. P. 303—309.
- Gerdes K. & Kahane S. (2001), Word order in German: A formal dependency grammar using a topological hierarchy, *Proceedings of ACL*, Toulouse, 2001. P. 220—227.
- Gerdes K. & Kahane S. (2004), L'amas verbal au cœur d'une modélisation topologique du français, *Proceedings of Journées de la syntaxe — Ordre des mots dans la phrase française, positions et topologie*, Bordeaux, 2004. 8 p.
- Kahane S. (2001a), What is a natural language and how to describe it? Meaning-Text approaches in contrast with generative approaches, *Computational Linguistics, Proceedings of CICLing 2001*, Mexico, Springer Verlag, 2001. P. 1—17.

- Kahane S. (2001b), A fully lexicalized grammar for French based on Meaning-Text theory, *Computational Linguistics, Proc. CICLing 2001*, Mexico, Springer Verlag, 2001. P. 18—31.
- Kahane S. (2004), Grammaires d'unification polarisées, *Proceedings of TALN*, Fès, 2004. P. 233—242.
- Kahane S. (2005), Structure des représentations logiques polarisation et sous-spécification, *Proceedings of TALN*. 2005. 10 p.
- Kahane S. & Lareau F. (2005), Grammaire d'Unification Sens-Texte : modularité et polarisation, *Proceedings of TALN*. 2005. 10 p.
- Lareau F. (2004), *Vers un modèle formel de la conjugaison française dans le cadre des grammaires d'unification Sens-Texte polarisées*, Pre-doctoral exam report, Montréal, Université de Montréal, 2004.
- Mel'čuk I. (1988), *Dependency Syntax: Theory and Practice*, Albany, SUNY Press, 1988.
- Mel'čuk I. (2001), *Communicative Organisation of Natural Language*, Benjamins, 2001.
- Nasr A. (1995), A formalism and a parser for lexicalised dependency grammars, *4th Int. Workshop on Parsing Technologies*, State University of New York Press, 1995.
- Polguère A. (1992), Remarques sur les réseaux sémantiques Sens-Texte, in A. Clas (éd.), *Le mot, les mots, les bons mots*, Presses de l'Université de Montréal, 1992.
- Polguère A. (2000), A “natural” lexicalization model for language generation, in *Proceedings of SNLP'2000*, Chiangmai, Thailand, 10—12 May 2000.
- Shieber S. M. & Schabes Y. (1990), Synchronous tree-adjointing grammars, *Proceedings of the 13th Int. Conference on Computational Linguistics*. Vol. 3. 1990. P. 253—258, Helsinki, Finland.

On the use of “quasi-grammeme” in Athabaskan¹

Andrej A. Kibrik

Institute of Linguistics, Russian Academy of Sciences
kibrik@comtv.ru

Many prefixal positions in the Athabaskan verb template can be attributed stable labels “derivation” or “inflection”. In addition, there is a large group of quasi-inflectional positions; this was identified due to the illuminating notion of quasi-grammeme proposed by Mel’čuk. Generally, the inflection vs. quasi-inflection vs. derivation opposition is a key to understanding the intricate synthesis implemented in the Athabaskan verb.

Понятие квазиграммемы и атабаскские языки. Многие префиксальные позиции в атабаскском глаголе могут быть четко определены как словоизменяющие либо словообразовательные. Кроме того, есть большая группа квазисловоизменяющих позиций, идентифицируемых как таковые благодаря полезному понятию квазиграммемы, введенному И. А. Мельчуком. В целом, оппозиция словоизменения/квазисловоизменения/словообразования является ключом к пониманию атабаскского полисинтетизма.

1. Introduction

In this paper I am going to connect the intricacies of Athabaskan morphology with some ideas proposed in the morphological component of the Meaning-Text theory, particularly in Mel’čuk 1993. Athabaskan languages constitute the largest native American family of the USA and Canada, embracing some 40 languages in three separated areas: Alaska and western Canada; Oregon and Northern California; and the Southwest. Athabaskan languages are among the world’s morphologically most complex languages. Most of what is expressed by syntactic constructions or function words in more familiar languages is expressed by morphology in Athabaskan. Therefore, morphological issues are of more weight in Athabaskan grammar compared to a “typological average”. Athabaskan languages are

¹ Research underlying this paper was supported by grant #05-04-04240 of the Russian Foundation for the Humanities.

strikingly unusual typologically (Kibrik 2002); they are among the major language families of the world, and their data certainly cannot be neglected, if one strives for a typologically grounded theory of grammar.

2. Inflection vs. derivation

Among the central morphological problems is the traditional issue of the derivation vs. inflection opposition. Since attempts to distinguish between inflection and derivation encounter many difficulties (both universally and in particular languages), a heretical question may arise: is it really necessary to make this distinction? It is helpful to identify aspects in which this distinction is useful:

1) the practical, descriptive aspect: when describing a language we have to discriminate between the information listed in the dictionary and that reserved for the grammar; derivation would be accounted for in the dictionary, and inflection in the grammar (or, alternatively, inflection and productive derivation can be described in the grammar);

2) the theoretical (cognitive) aspect: what we think is stored as one lexeme (with its forms being, at least potentially, constructed on the basis of grammatical rules by the speakers in real usage), and what we think is definitely stored and remembered as separate lexemes.

Dealing with the first aspect we are obliged to stick to the binary inflection vs. derivation opposition: every form should be accounted for somewhere in our description. Dealing with the second aspect we can allow for a possibility that the dichotomy is a continuous one. Some evidence has been presented, first, for the storage of inflectional forms in the speakers' long-term memory (Stemberger and MacWhinney 1988), and second, for the ability of the speakers to produce new derived lexemes on-line (see e.g. Zemska 1992). Below I do not discriminate between the two aspects consistently, but the difference between them should be kept in mind.

I believe that the distinction between inflection and derivation captures an important generalization, even though there are certainly instances when a straightforward application of this dichotomy is not possible.

3. Athabaskan morphology

Athabaskan morphology is traditionally described through the notion of template: a linearly ordered sequence of morphological positions such

that each position is attributed to a functionally homogeneous set of morphemes. For a recent proposal of templatic morphology in Athabaskan see Kari 1989; for a critique see Rice 2000. In the chart below I present what can be called Standard Average Athabaskan template; it is based on the actual templates of Navajo (Young and Morgan 1992), Hupa (Golla 1970), Ahtna (Kari 1989, 1992), Sarcee (Cook 1984), and Slave (Rice 1989). The template includes those positions that occur at least in two languages out of the listed five.

- 18 proclitic
- 17 (b) Oblique [postpositional object] +
 (a) preverb [postposition]
- 16 various derivational
- 15 reflexive Accusative [direct object] pronoun (Navajo, Hupa)
- { 14 iterative
- 13 distributive
- 12 incorporate (Slave, Sarcee, Ahtna)
- 11 number (Hupa, Slave)
- { 10 Accusative [direct object] pronoun
- 9 3 person Nominative [deictic subject] pronoun
- 8 transitivity decrease (Navajo, Hupa)
- 7 qualifier
- 6 inceptive
- 5 qualifier
- 4 conjugation
- 3 mode
- 2 1/2 person Nominative [subject] pronoun
- 1 transitivity indicator [classifier]
- 0 root
- +1 mode/aspect suffix
- +2 enclitic

In the chart, I propose some non-traditional terms; their traditional correlates are indicated in square brackets. Bracket signs on the left indicate the pairs of positions that might be interchanged in some languages (or in some instances within one language). Two striking peculiarities obvious from this chart are:

- the almost exclusively prefixing character, which, to my knowledge, does not occur in other languages of this level of morphological complexity;
- the violations of the cross-linguistic scope, or relevance, ordering tendencies (see Rice 2000), in particular, placement of inflection closer to the root than derivation.

Violations of scope-based ordering in Athabaskan is not limited to the derivation/inflection distinction, but concerns also the relative order of inflectional categories. As has been demonstrated by A. E. Kibrik (1980) and Bybee (1985: 34–35), among others, the most common ordering of grammatical categories, in terms of proximity to the root, is: ROOT — (A) aspect — (B) tense — (C) mood — (D) person. Compare examples from the thoroughly scope-obedient (suffixing) Central Alaskan Yup’ik (Eskimo-Aleut family) and from the entirely unpredictable (prefixing) Navajo:

(1) Central Alaskan Yup’ik (Eskimo-Aleut, Mithun 1999: 407)

ROOT	DERIVATION			INFLECTION		
<i>iter-</i>	<i>ngnaqe-</i>	<i>rraar-</i>	<i>tur-</i>	<i>llru-</i>	<i>u-</i>	<i>q</i>
enter-try.to-	first-	repeatedly-	Past-	Indicative-	3Sg	‘he always wanted to enter first’
	ASPECT	TENSE	MOOD	PERSON		
	A	B	C	D		
	Scope/relevance ranks					

(2) Navajo *nihideestsil* ‘I will move on the buttocks to a point’

DERIVATION			INFLECTION			ROOT
<i>ni-</i>	<i>hi-</i>	<i>di-</i>	<i>ghi-</i>	<i>s-</i>	<i>l-</i>	<i>tsil</i>
Terminative-	Seriative-	Inceptive-	Progr-	1Sg.Nom-	TI↑↓-	move.sitting
DERIVASP	LEXASP	INFLASP	MODE	PERSON	VALENCY	ROOT
A	A	A	B/C		D	
	Scope/relevance ranks					

In Central Alaskan Yup’ik, the hierarchical proximity of derivation to the root, compared to the peripherality of inflection, is obvious, and the boundary between derivation and inflection is clear-cut. In Navajo the ordering is odd and morphemes such as inceptive and transitivity indicator (TI) do not lend themselves easily to the derivation vs. inflection dichotomy (see below).

In this paper I mostly concentrate on the derivation/inflection dichotomy that is represented in the Athabaskan template in a puzzling and unusual way. One cannot hope to understand Athabaskan morphology without addressing the issue of how individual morpheme positions can be characterized in terms of this dichotomy. I am not going to propose an ultimate explanation for the Athabaskan paradox below. Rather the goal of this paper is to find out to what extent the labels “derivational” and “inflectional” help to explain the Athabaskan templatic positions, and how the conceptual apparatus can be extended.

4. Inflection and derivation in Athabaskan

In Athabaskan linguistics, it is usually implied that certain morphological positions can be characterized in terms of the inflection vs. derivation dichotomy. For example, Hoijer, Young and Morgan, and others typically label certain positions «derivational». In Sapir and Hoijer (1967: 85) only two positions (corresponding to what are positions 18—16 and 6—5 in my chart above) were considered derivational, and all the rest (with one exception, see below) were labeled «paradigmatic» (inflectional). Kari (1989: 428) quoted an unpublished paper by Whorf who quoted Sapir's term «interrupted synthesis» that characterizes this property of Athabaskan.

Let us see how the derivation and inflection labels apply to the 18 prefixal positions in the Athabaskan verb. The morphological positions that are exclusively devoted to a particular type of morphology, are the following:

- Purely derivational positions: qualifier [5, 7] and various derivational [16].
- Purely inflectional positions: Nominative pronoun [2/9], mode [3], conjugation [4], Accusative pronoun [10], Reflexive Accusative pronoun [15], Oblique pronoun [17b].

But what about all other positions and the corresponding morpheme classes? Can they be characterized as either inflectional or derivational? Along with derivation and inflection, I propose to distinguish the third phenomenon in Athabaskan: quasi-inflection, following Mel'čuk's (1993) notion of quasi-grammeme, and Golla's (1970: 115) notion of «secondary inflection». In Sapir and Hoijer (1967) most of such morphemes were treated as paradigmatic (that is, inflectional). In order to see how these morphemes should better be treated, one needs to delve into the actual definitions of derivation and inflection.

5. Quasi-inflection

Mel'čuk (1993: 262ff.) defines inflectional categories on the basis of two necessary properties: obligatoriness of expression in every word to which the given category applies, and regularity of expression; he also mentions some additional typical properties of the inflectional meanings (= grammemes): abstractness of meaning and broad combinatorial potential. Mel'čuk further notes that the regularity of expression is a relative

(gradual) rather than absolute property, and therefore a category may be inflectional to a certain degree. The grammeme ultimately serves in Mel’čuk’s framework as a basis for the definition of the lexeme. The derivational meaning (= derivateme) is defined negatively with respect to the grammeme (p. 287 ff.): non-lexical meaning which is not inflectional. Mel’čuk lists seven distinctions that contrast grammemes and derivatemes (p. 293ff.), though none of these distinctions can be considered completely necessary. As other authors, Mel’čuk acknowledges the absence of a rigid boundary between inflection and derivation, and identifies one important intermediate class of meanings (p. 302–303): **quasi-grammmemes** which resemble genuine grammemes in being regular, but on the other hand are not obligatory. This notion turns out very useful for Athabaskan. (In addition, Mel’čuk draws a distinction between the strongly derived (productive derivation) and the weakly derived (non-productive derivation) lexemes, see below.)

On the basis of Mel’čuk’s terminological pair “grammeme (= inflectional meaning)” — “inflection” I propose a parallel pair “quasi-grammeme” — “quasi-inflection”. I claim that the following prefixal positions in the Athabaskan template belong to morphemes that are quasi-grammmemes and thus can be characterized as quasi-inflection: transitivity decrease [8], number [11], distributive [13], and iterative [14]. Note that all of these positions can only be occupied by one morpheme each. For example, the distributive morpheme adds the meaning that the event took place on more than one occasion and, by inference, suggests the plurality of the absolutive argument.

(3) Upper Kuskokwim, distributive

- a. *tsasja ts’a-di-ghi-l-joch*
 cup Pref-Pref-ghPf-TI↑-broke
 ‘I broke a cup/the cups’
- b. *tsasja ts’a-n-da-zi-l-joch*
 cup Pref-Distrib-Pref-zPf-TI↑-broke
 ‘I broke the cups’

The form in (3a) does not have the meaning of non-distributivity or singularity of the patient, rather it is unmarked in this respect. So the distributive morpheme does not belong to any paradigm but is in contrast only to the absence, or lack of any formal marking and meaning. This is very different from inflectional categories whose zero members do convey a specific meaning, as e.g. English *I work* is the present tense, contrasted to *I worked*.

Similarly, the iterative morpheme indicates the repeated character of an event, while its absence does not convey any specific meaning.

(4) Navajo, iterative

a. <i>yi-di-lé</i> 33.Acc-2Sg.Acc-handle 'He carries it along'	b. <i>né-i-di-dlééh</i> Iter-33.Acc-2Sg.Acc-TI↓:handle 'He repeatedly carries it along'

Still another quasi-grammeme, transitivity decrease, appears e.g. in Navajo semipassive, accompanying the indefinite nominative pronoun.

(5) Navajo, transitivity decrease

a. <i>lééchqáʔi ni-sh-hash</i> dog 2Sg.Acc-zPf-TI↑:bit 'The dog bit you'	b. <i>ni-ʔ-di-sh-ghash</i> 2Sg.Acc-Indf.Nom-Decr-zPf-TI↑↓:bit 'You have been bitten'

Mel'čuk's first requirement with regard to quasi-grammemes is that they should be regular. The morphemes such as distributive or iterative can be freely added to any verb, in accordance with the semantic demands. Considering these morphemes derivational, as sometimes proposed, seems very unnatural: they definitely do not create new lexemes, but are freely added to any verbal meaning. (Note that adding quasi-grammemes entails other changes in the verbs in (3)–(5), but this is a different story, and a very complicated story of Athabaskan polysynthesis.)

Mel'čuk's second principle is non-obligatoriness of quasi-grammemes. Obligatoriness is the defining feature of grammatical categories: one or another grammeme belonging to the category must always be present in the words of the relevant word class. But this does not hold for the morphemes in question. When a certain quasi-grammeme does not appear on the verb this does not qualify it semantically in any way. Therefore, there is no sense in postulating zero counterparts in these positions. Only the presence of the specific morphemes is meaningful. This is the crucial difference of quasi-inflection from inflection.

Perhaps one more prefix position can be treated as quasi-inflection, namely, incorporate [12]. Unlike the four positions discussed above, there are multiple morphemes potentially occurring in this position. But otherwise it seems that this category is similar, being both regular and non-obligatory. The lack of an incorporate has no specific meaning and thus cannot be considered the zero member of the opposition.

(6) Upper Kuskokwim, incorporation

- | |
|---|
| a. <i>ho-ghi-s-dlo</i>
Pref-ghPf-1Sg.Nom-took.mass/pl.solid
'I took it (anything mass or plural solid) out' |
| b. <i>ho-lats-ghi-s-dlo</i>
Pref-dirt-ghPf-1Sg.Nom-took.mass/pl.solid
'I took some dirt out' |

- c. *ho-tsa-ghi-s-dlo*
 Pref-rock-ghPf-1Sg.Nom-took.mass/pl.solid
 ‘I took some rocks out’
- d. *ho-tu-ghi-s-dlo*
 Pref-water-ghPf-1Sg.Nom-took.mass/pl.solid
 ‘I took some water (in buckets) out’

In my view, quasi-grammemes are the polysynthetic analog of various adverbs and circumstants in less synthetic languages. They are optional and can be added or not added freely, and for this reason they cannot be captured by the opposition of inflection and derivation. Forced pressing of quasi-grammemes into derivation, typical of Athabaskan lexicography, may be one of the reasons responsible for the huge size and complexity of Athabaskan dictionaries, such as Young and Morgan 1987, 1992; Jetté and Jones 2000.

6. More complex instances

The remaining four prefixal positions require some special comment. Positions of proclitics [18] and preverbs [17a] in many cases might best be treated as quasi-inflection, though sometimes they might be derivational. The attribution of a preverb as derivational vs. quasi-inflectional is parallel to the distinction between arguments and adjuncts (actants vs. circumstants), that is, those participants that are predicted by the semantics of the lexeme, and those that are not.

The inceptive morpheme occurring in position [6] — see ex. (2) above — is originally derivational. But in many Athabaskan languages it developed a secondary function as a part of the future mode morphology; in these forms it can be inflectional or quasi-inflectional, depending on individual language.

Finally, let us consider the transitivity indicator [classifier] position [1]. It is the only position not identified by Sapir and Hoijer (1967) as either derivational or inflectional. In fact, the transitivity indicators (TIs) are largely inflectional, especially as concerns the changes $\varnothing$ > $\underline{d}$ - and $\underline{l}$ > $\underline{l}$ - accompanying the regular transitivity decrease processes (such as passive, reflexive, etc., see Kibrik 1996). The change $\varnothing$ > $\underline{l}$ -, accompanying the transitivity increase processes (productive, such as the causative, see Kibrik 1993, or unproductive, such as the anticausative, see Kibrik 1996), is a case of the derivational usage of the TI. (Note that cross-linguistically the transitivity-increasing processes tend to pertain to deri-

vation, and the transitivity-decreasing processes to inflection.) The following types of TI occurrences can be differentiated:

- inflectional, fully predictable (passive, reflexive)
- derivational, quite predictable (causative) or poorly predictable (anticausative)
- lexical, normally zero; if not zero, then fossilized, fully unpredictable.

Thus the $\bar{\text{I}}$ - TI is primarily derivational; when used in constructions like passive or reflexive, the $\underline{\text{d}}$ - TI is typically inflectional, and I - simultaneously refers to both kinds of processes; in the constructions like anticausatives and historical passives (see Kibrik 1996) $\underline{\text{d}}$ - and I - are used derivationally. TIs constitute a typologically interesting example of an inherently inflectional/derivational category.

7. Conclusion

To conclude, I have demonstrated how the intricate morphological system of Athabaskan languages can be treated with respect to derivation vs. inflection dichotomy. As in other languages, we find groups of clearly derivational and clearly inflectional affixes. Many prefixal positions in the Athabaskan verb template can be attributed stable labels “derivation” or “inflection”. A striking feature of Athabaskan is that inflection is generally closer to the root than derivation. The question of whether this is related to the prefixal nature of Athabaskan morphology is left for further investigation. It may be that languages tend to place derivation linearly before inflection; this principle, if viable, coincides with the “derivation closer to the root than inflection” principle in suffixing languages that constitute the unmarked typological option.

There is a significant residue of positions in the Athabaskan template that we cannot readily identify as either derivational or inflectional. There are at least two reasons for that. First, there is a large group of quasi-inflectional positions; this was identified due to the illuminating notion of quasi-grammeme proposed by Mel’čuk. Athabaskan quasi-grammemes are isolated morphemes that resemble inflectionals in being fully regular; but unlike inflectionals they cannot be considered obligatory since they do not belong to any category, and can be contrasted only to absence.

The second reason for complications is template positions that can variously be associated with either derivation, or quasi-inflection, or inflection. Some morphemes are originally derivational, but they developed a secondary usage that is not. In other cases certain morphemes belonging

to a certain position are inherently inflectional, while others are more like derivational; still they constitute a natural class.

Generally, the inflection vs. quasi-inflection vs. derivation opposition is a key to understanding the intricate synthesis implemented in the Athabaskan verb.

Non-obvious abbreviations in glosses

33 — both arguments are 3d person

Acc[usative]

Distrib[utive]

ghPf — gh-perfective

Iter[ative]

Progr[essive]

Nom[inative]

Pref — derivational prefix of a meaning irrelevant for the present discussion

TI↓, TI↑, TI↑↓ — transitivity indicator of decrease, of increase, of increase plus further decrease

zPf — z-perfective

References

- Bybee, Joan L. 1985. *Morphology. A study of the relation between meaning and form*. Amsterdam: Benjamins, 1985.
- Cook, Eung-Do. 1984. *Sarcee grammar*. Vancouver: University of British Columbia Press, 1984.
- Golla, Victor. 1970. *Hupa grammar*. Ph. D. dissertation. UC Berkeley, 1970.
- Jetté, Jules, and Eliza Jones. 2000. *Koyukon Athabaskan dictionary*. Fairbanks: ANLC, 2000.
- Kari, James M. 1989. Affix positions and zones in the Athapaskan verb complex: Ahtna and Navajo. *IJAL* 55. 1989. P. 424—454.
- Kari, James. 1992. Some concepts of Ahtna Athabaskan word formation. In: M. Aronoff (ed.) *Morphology Now*. SUNY Press, 1992. P. 107—131.
- Kibrik, Aleksandr E. 1980. O semanticheskoy motivirovannosti rangovogo raspredeleniya morfem. In: *Tezisy rabochego soveshchaniya po morfeme*. Moscow, 1980.
- Kibrik, Andrej A. 1993. Transitivity increase in Athabaskan Languages. In: B. Comrie and M. Polinsky (eds.) *Causatives and transitivity*, Amsterdam: Benjamins, 1993. P. 47—67.
- Kibrik, Andrej A. 1996. Transitivity decrease in Navajo and Athabaskan: Actor-affecting propositional derivations. In: *Athabaskan language studies. Essays*

- in honor of Robert W. Young. Ed. by E. Jelinek, S. Midgette, K. Rice, and L. Saxon. Albuquerque: UNM Press, 1996. P. 259—304.
- Kibrik, Andrej A. 2002. A typologically oriented portrait of the Athabaskan language family. In: V. I. Podlesskaya (ed.) Third winter school in typology. Materials of lectures and seminars. M.: RGGU, 2002. P. 38—48.
- Kuryłowicz, Jerzy. 1965. The evolution of grammatical categories. In: *Esquisses Linguistiques* 2. 1965. P. 38—54.
- Mel'čuk, Igor A. 1993. *Cours de morphologie générale*. Les Presses de L'Université de Montréal, 1993.
- Mithun, Marianne. 1999. *The languages of native North America*. Cambridge: CUP, 1999.
- Rice, Keren D. 1989. *A grammar of Slave*. Berlin: Mouton de Gruyter, 1989.
- Rice, Keren D. 2000. *Morpheme order and semantic scope*. Cambridge: CUP, 2000.
- Sapir, Edward, and Harry Hoijer. 1967. The phonology and morphology of the Navaho language. UCPL 50. Berkeley: University of California Press, 1967.
- Stemberger, Joseph Paul, and Brian MacWhinney. 1988. Are inflected forms stored in the lexicon? In: Hammond and Noonan (eds.) *Theoretical morphology: Approaches in modern linguistics*. San Diego: Academic Press, 1988. P. 101—118.
- Young, Robert, and William Morgan. 1987. *The Navajo language. A grammar and colloquial dictionary*. Albuquerque: University of New Mexico Press, 1987.
- Young, Robert, and William Morgan. 1992. *Analytical lexicon of Navajo*. Albuquerque: University of New Mexico Press, 1992.
- Zemskaja, Elena A. *Slovoobrazovanie kak dejatel'nost'*. (Word derivation as an activity.) Moscow: Nauka, 1992.

Игорь Мельчук и наши экспедиции¹

А. Е. Кибрик

Московский государственный университет им. М. В. Ломоносова
ae_kibrik@mail.ru

Экспедиции ОСИПЛа (отделения структурной и прикладной лингвистики филфака МГУ)² стартовали в 1967 году. Первоначально они преследовали учебную цель — лингвистической практики студентов ОСИПЛа на незнакомом языковом материале. Такая практика проходит в экстремальных условиях, когда исследователю о языке практически ничего неизвестно и доступны наблюдению только образцы устной речи и их переводы на русский язык. Исследователь должен уметь зафиксировать эти образцы в письменной форме и установить, из каких значимых элементов они состоят, опираясь на переводные эквиваленты и лингвистическую теорию, претендующую на приложимость к произвольному естественному языку. Такая полевая работа является идеальным полигоном для проверки знаний в области теории, а также для проверки самой существующей лингвистической теории на прочность.

Некоторым аналогом такой постановки задачи являются практикующиеся в лингвистических олимпиадах³ задачи-билингвы. В билингвах решающему предлагается языковой материал исходно неизвестного языка с переводом на русский язык. Необходимо обнаружить структурные закономерности в данных неизвестного языка, опираясь на их смысловой эквивалент на известном (русском) языке и продемонстрировать это, переводя контрольный текст с неизвестного языка на известный или наоборот. При решении таких задач решающий опирается на некоторые предположения о том, какими грамматическими свойствами может обладать потенциальный есте-

¹ Работа выполнена при поддержке РГНФ, проект 04-04-00215а.

² Ниже я буду говорить только о тех экспедициях ОСИПЛа, которыми я лично руководил.

³ Традиция проведения лингвистических олимпиад возникла на ОСИПЛе двумя годами раньше, в 1965 году.

ственный язык. Практика олимпиад показывает, что такого рода задачи может успешно решать школьник, не имеющий профессиональных лингвистических знаний, исходя из того, что возможность правильного решения обеспечивается тем, что структура человеческих языков регулируется естественными универсальными грамматическими свойствами.

При сходстве исследовательских целей — открыть грамматику неизвестного языка, опираясь только на языковые данные, — между олимпиадными задачами и полевой работой имеются существенные различия. Если лингвистические задачи специально придумываются (сочиняются), и при «озадачивании» того или иного явления авторам задачи приходится существенно ограничиваться регулярными грамматическими правилами, восстанавливаемыми на минимальном объеме языковых данных, то в экспедиционных условиях изучения незнакомого языка задачи обычно формулируются спонтанно по ходу работы, и для их решения надо собрать столько и таких языковых данных, которые необходимы и достаточны для правдоподобного решения. Кроме того, олимпиадная задача всегда предполагает, что ее *можно* решить (найти единственное правильное решение). Обнаружение грамматики (или, точнее, ее фиксированного фрагмента) неизвестного языка нельзя априорно считать решаемой задачей. Оставляя в стороне множество прагматических факторов, нельзя сбрасывать со счетов возможную несостоятельность самой лингвистической теории, без перестраивания или достраивания которой адекватное описание данных произвольного языка невозможно.

В соответствии со сформулированными выше целями нужно было создать исследовательский аппарат коллективной полевой работы, которого в готовом виде не существовало. Отработка такого аппарата требовала привлечения данных различных языков. В 1967 году изучался лакский язык, в 1968 — арчинский (Дагестан), в 1969 — шугнанский (Памир), в 1970 — хиналугский (Азербайджан). За четыре года был накоплен и обобщен определенный опыт полевой работы, стало очевидно, что такая работа может быть организована достаточно эффективно, и возникло искушение попробовать довести исследование какого-нибудь языка до своего логического результата — грамматического описания. Таким первым экспериментом была грамматика хиналугского языка, написанная по горячим следам в 1970—1971 году⁴, была также начата работа по более амбициозному

⁴ А. Е. Кибрик, С. В. Кодзасов, И. П. Оловянникова. Фрагменты грамматики хиналугского языка. М., 1972.

проекту — грамматике арчинского языка, рассчитанному на несколько лет. Вместе с тем идея расширения языкового кругозора нас не покидала, и было решено обратиться к совершенно новому для нас языковому ареалу. В качестве объекта описания был выбран один из чукотско-камчатских языков — алюторский.

Естественно, что грамматическое описание всегда исходит из некоторой лингвистической теории. Такой теорией для нас в то время были структурализм и универсализм, существенно изменившие лингвистическую парадигму в отечественной лингвистике и продемонстрировавшие свои очевидные преимущества в наших полевых исследованиях в области фонетики и морфологии. В лингвистике второй половины 60-х возникли также новые интегральные модели языка. В нетрадиционной советской лингвистике признанным лидером в области лингвистической теории был Игорь Мельчук, разработавший оригинальную интегральную лингвистическую модель, получившую известность как модель Смысл — Текст. В то время эта модель была известна в основном не по публикациям, а по публичным выступлениям Мельчука на различных конференциях. У меня возникла авантюрная идея пригласить Игоря участвовать в предстоящей экспедиции. Я считал, что студенты ОСИПЛа обделены тем, что Мельчук им не преподает. Экспедиция давала уникальный шанс восполнить этот пробел хотя бы для небольшой группы студентов. Кроме того, поработать над алюторским языком с автором интегральной модели тоже было очень заманчиво.

С Игорем я был знаком прежде всего по его туристической деятельности: в 1957 году, учась на первом курсе филфака, я впервые участвовал в многодневном лыжном зимнем походе, руководимом Мельчуком. Это было для меня очень яркое событие, оставившее неизгладимое впечатление, — как от самого похода, так и от личности Игоря. Я многому у него научился и стал ежегодно организовывать аналогичные походы.

С Мельчуком как выдающимся лингвистом я стал знакомиться позднее, по окончании классического отделения, когда начал специализироваться в области лингвистики. Это было знакомство издалека, по публикациям и выступлениям Мельчука.

На одном из таких выступлений я и сделал ему предложение поехать с нами на Камчатку. Игорь сказал, что совершенный профан в полевой работе, но ему было бы интересно попробовать. Успеху моего предложения способствовал, видимо, врожденный авантюризм Игоря, готового к неожиданным сферам приложения его необузданной энергии.

В официальном списке участников были пять студентов, уже не первый год участвовавших в экспедициях, С. Кодзасов и я. Кроме того, с нами поехали два приглашенных лингвиста-волонтера: Игорь Мельчук и Сима Никитина.

Наша поездка началась не с Камчатки, которая завернула нас уже перед самой посадкой из-за погодных условий, а с Магадана. Мы приземлились в Магадане вечером, и вылет был назначен на следующее утро. Нам предстояло провести ночь в аэропорту. Но Игорь тут же предложил использовать исторический шанс посетить «столицу Колымского края». Это был незабываемый марш-бросок. Магадан представлял собой практически одну центральную улицу с 3—4-х этажными домами типичной сталинской застройки 50-х годов, окруженную деревенскими окраинами. Мирный провинциальный город, только его название напоминает мрачную эпоху ГУЛАГа, но забыть о ней тут невозможно. Белую магаданскую ночь мы провели на песчаном берегу морской бухты, по которой были рассыпаны молодежные компании, тихо и проникновенно жегшие костры из выброшенной морем дощатой тары. Мы тоже развели костер, смотрели на огни стоявших на рейде кораблей и слушали негромкие песни Игоря, соответствовавшие месту и настроению. Берег постепенно пустел, становилось холодно, мы вернулись в город и дожидались автобуса, греясь в подъезде какого-то дома.

Регистрация была назначена на пять утра, опять задержки, но наконец мы снова летим и в этот раз благополучно прибываем в Петропавловск-Камчатский. Город окружают снежные шапки сопки, над которыми царит величественный конус Авачинской сопки. У нас образовалась пауза в один-два дня в ожидании корабля, на котором мы должны были плыть в Вывенку — конечный пункт нашего путешествия, алуторское селение. Игорь не мог спокойно смириться с необходимостью пустого ожидания и в первый же вечер предложил подняться на Авачинскую сопку. Часть нашей группы его поддержала, они быстро собрались и отправились покорять гору-вулкан. В этом был весь Игорь. Его жизненный принцип — ставить перед собой максималистские цели, веря в свои силы и возможность преодоления любых препятствий. Я остался среди ренегатов, на глаз оценивая это предприятие как заведомо нереалистическое. В середине следующего дня авантюристы вернулись. Они проехали на автобусе километров 30-50, до ближайшей доступной точки, и обнаружили, что расстояние до сопки не уменьшилось: она по-прежнему стояла у горизонта. Какое-то время они шли к ней пешком, но сопка не приближалась. Они узнали также, что эта сопка — почти трехтысячник,

восхождение на нее возможно только с проводником и при наличии специального снаряжения и занимает несколько дней. Неудачу Игорь воспринял довольно легко. По крайней мере он доподлинно выяснил, что у этой задачи нет сиюминутного решения.

Нам предстояло трехдневное плавание на комфортабельном многопалубном корабле, удлинившееся на несколько дней из-за шторма в океане. Это время мы в основном посвятили курсу лекций Мельчука по модели Смысл — Текст. Мы собирались в самой большой каюте, и Игорь обстоятельно излагал свою теорию в режиме рабочего семинара. Он делал это увлеченно, убедительно, логично, демонстрируя великолепные качества лектора. Таких лекций наши студенты еще никогда не слушали. Мы прерывались только во время сильной качки, физиологические последствия которой Игорь очень смачно комментировал, несколько эпатируя девочек, не привыкших к такому юмору.

Преодолев переправу с корабля на берег с помощью вызванного по радиации катера, чуть не утопившего наше снаряжение и нас из-за хмельной неосторожности его команды, мы ступили, наконец, на землю аляutorцев. Мы достигли географической и научной цели нашего путешествия. Началась работа над аляutorским языком.

Игорь достаточно быстро освоил специфику полевой работы и погрузился в нее с головой, не разделяя важных и неважных, интересных и неинтересных, приятных и неприятных ее аспектов. Он самозабвенно учился слышать и имитировать звуки аляutorского языка, активно участвовал в обсуждении деталей и тонкостей фонологических трактовок трудных случаев, принципов транскрипции, правил перевода аляutorских слов, конструкций и грамматических морфем, увлеченно работал с информантами и обладал даром человеческого общения с ними, превращая нудный процесс ответов на вопросы в праздник игры, наряду со всеми дежурил по кухне, таскал воду, выносил помой, был центром культурной жизни в группе. Он принимал участие во всех видах лингвистической работы по сбору языкового материала: записывал аляutorские тексты и разбирал их с информантами, занимался сбором лексики и грамматических парадигм, собирал примеры на грамматические явления. С неменьшим энтузиазмом он отдавался и радостям досуга — прогулкам по сопкам, тундре, побережью, купанию в ледяном океане, ловле рыбы и, конечно, пению.

Сбор языкового материала составляет львиную долю работы в экспедиции, но он сопровождается постоянным его теоретическим осмыслением. Для Игоря это было совершенно естественно и, более того, наиболее интересно. Обсуждения наши проходили постоянно,

как на специальных «пятиминутках», так и во время прогулок, за трапезой, перед отходом ко сну и во время утреннего моциона. Здесь Игорь чувствовал себя безусловным лидером. Он до всех гипотез и решений хотел дойти сам, и у него для этого были все основания. Но должен сказать, что спорить с ним довольно трудно. Он плохо воспринимал идеи, которые сам не выстрадал. Правда, потом нередко приходил к тем же выводам самостоятельно, как к своим собственным.

А алюторский язык оказался не так прост, как Игорю сперва представлялось, он удивлялся, сердился, что не понимает, почему его предсказания, как должна выглядеть та или иная форма, не сбываются, и упорно стремился втиснуть данные в свои правила, предписанные его моделью. За этим процессом было очень интересно и поучительно наблюдать.

Можно сказать, что Игорь был в центре всей нашей экспедиционной жизни, и это особенно почувствовалось, когда настал день его отъезда. Это, пожалуй, был наиболее печальный день в экспедиции, но и его Игорь окрасил в бодрые жизнерадостные тона. Мы все шли на пирс под его пение «Агнешки» и брызги шампанского у борта МРТ (малого рыболовецкого траулера), который вез его в районный аэропорт.

В следующем, 1972 году, мы снова поехали в Вывенку. Нас снова не принял Петропавловск-Камчатский, и мы залетели в Иркутск, прогуляли там по деревянным мостовым и осмотрели этот насквозь деревянный город. Но на Авачинскую сопку Игорь уже не стремился.

Вообще он считал, что с алюторским языком почти все ясно, осталось задать 17 вопросов, так что долго там делать нечего. Но алюторская действительность преподнесла ему очередной сюрприз. Уже через несколько дней он с нескрываемым изумлением сказал, что ничего не понятно и времени нам не хватит. Но с еще большим рвением вгрызался он в гранит алюторского языка, надеясь его одолеть. Действительно, все грамматические единицы были найдены, все парадигмы исчерпывающе записаны, но почему они так устроены, теория ответа не давала.

Работа с алюторским языком вместе с Мельчуком, в правильность теории которого мы безгранично верили, была для нас очень поучительна. Нельзя было уповать на то, что мы просто плохо понимаем теорию Мельчука или неквалифицированно ее применяем. Нам крепко повезло, что нам попался такой удивительный язык и что вместе с нами был Мельчук. Это заставляло думать о высоких теоретических материях и о беспредельности языкового разнообразия, за кажущимся хаосом которого должна скрываться какая-то новая правда, не уловленная текущим состоянием лингвистической теории.

ГУЛЯТЬ ТАК ГУЛЯТЬ: between Grammar and Dictionary¹

Mikhail Kopotev

Department of Slavonic and Baltic Languages and Literatures,
University of Helsinki, Finland
mihail.kopotev@helsinki.fi

The article is devoted to the analysis of two Russian syntactic phrasemes. The restrictions on lexical variables are investigated, their discourse links are established, and frequency data on their usage are cited. Some definite realizations of the phrasemes in question can be generated using certain grammatical rules, however a part of them is stored in the Dictionary in the ready-to-use form.

ГУЛЯТЬ ТАК ГУЛЯТЬ: Между грамматикой и словарем. Статья посвящена анализу двух русских синтаксических фразем. Исследуются ограничения на лексические переменные, устанавливаются дискурсивные зависимости фразем, приводятся данные о частотности их употребления. Делается вывод, согласно которому конкретные реализации фразем могут порождаться грамматическими правилами, однако часть из них хранится в словаре в виде готовых к использованию единиц.

According to the definition of a syntactic phraseme given by I. A. Mel'čuk², Russian expressions like *Гулять так гулять* should be

¹ This research was supported by the Finnish Academy (a multidisciplinary research project, funded within the research programme "Russia in Flux"). I wish to express my gratitude to L. A. Biryulin, A. Mustajoki, A. Nikunlassi, and E. Ju. Protassova for their comments on drafts of the article.

² 'A syntactic phraseme is a surface-syntactic tree containing no full lexical nodes (its nodes are labeled with either lexemic variables or structural words) but possessing a specific signified, having as its signifier a specific syntactic construction, and a specific prosody, and featuring as well a specific syntactics' [Mel'čuk 1995, 215; see also Mel'čuk 1995a]. At the same time, the approach offered here is also based on 'construction-oriented' frameworks, above all Construction Grammar [Fillmore et al. 1988; Goldberg 1995; Kay, Fillmore 1999], and Experientialism [Lakoff 1987].

classified as *syntactic phrasemes*. They have been identified which manifest themselves in an almost identical way: namely, two repeated word forms, with the particle *TAK* in-between. The first type roughly means ‘*an uncontrollable choice*’, while the second one means ‘*the utmost quality*’.

The article is devoted to the analysis of these two syntactic phrasemes and a discussion of the question concerning the place of these phrasemes in Grammar or in the Dictionary. In the first part the difference in filling variables will be shown; in the second, the contextual dependence of phrasemes will be analyzed; the third part is devoted to an analysis of the most frequent realizations of the specified phrasemes, and a formal description will be presented in the last part.

1. Lexemic variables

The phrasemes in question possess very different constraints pertaining to lexemic variables.

1.1. The phraseme of an uncontrollable choice

Any lexeme from a previous context can be used as a filler for the variable *X* in the phraseme ‘...*X* так *X*’. Repeated twice, it serves for the expression of a situation where a choice is not under the close control of a speaker. Such uncontrollability depends on the lexical and morphological meanings of a word used, therefore it can be defined as a *radial structure* (J. Lakoff) represented in the following *central subcategory* and its *non-central extensions*: an unwilling (or insufficiently motivated) consent (1); nothingness of a choice (2), luck of a choice (3).

- (1) Беда была в том, что Зоя надумала меня женить на своей ученице. ⟨...⟩ Пройдя через легкомысленный ад собственной жизни, Зоя решила, что надо укреплять устои, в углу у неё висели иконы, у изголовья лежало Евангелие. Ну что ж, *жениться так жениться*, не обижать же старуху. [Volga, № 9, 1998]
- (2) — Я пешком пойду, — сказал Саша, — в сани только чемодан кину, он у меня легкий. — Наше дело сполнять, — пробормотал Нил Лаврентьевич, — *везти так везти, иттить так иттить*. [A. Rybakov. Strah]
- (3) Облава была спокойная. Солдаты ходили по хатам, брали всех мужчин, и все приходили спокойно, молча, как и я. Теперь уже никаких документов не смотрели, годы рождения

не играли роли. Все чисто и благородно: *попался так попался* — и заткнись. [А. Кузнецов. Babij jar]

Some Russian particles like *ну что ж*, *ну* along with others work as a test context for this phraseme. Some realizations of the phraseme are accompanied by a specific intonation with the tone falling towards the end of a phrase. Contrary to the stated opinion (RG 1980; Nikolina 1991; Veličko 1996), this phraseme includes a long list of variables and does not have any tense modifications.

An analysis of variables does not reveal almost any word class or other morphological constraints: any form can be used as a filler with the corresponding context support not only ‘infinitive, substantive in any case with preposition and not, adjective, adverb’ [Veličko 1996, 19]. See, for example, [Švedova 1960, 94—95]. The material collected has shown that finite verbs (4), participles (5), predicatives (6), numerals (7), and, probably, syntactic words (8) are served in addition to the mentioned forms³.

- (4) Крымов развёл руками, как бы говоря: ‘Что ж, *повезло так повезло*, ничего не поделаешь’. [V. Grossman. Žizn’ i sud’ba]
- (5) *Приучена так приучена*, — согласился ученый. — Пусть сторожит.... [G. Gorin. Skazka pro sobaku, kotoraja prožila trista let]
- (6) То, что Мустафе запрещено было кем-то плясать в литературном кабаке, позабавило Иуду Гросмана, но не озадачило: *ну нельзя, так нельзя*. [‘Oktjabr’, №1, 2001]
- (7) Когда мы подписывали контракт, я сразу сказал, что нужно подписывать на три года. Он же настаивал на пяти. В итоге я согласился — *пять так пять*. Сама церемония подписания проходила нервно. [‘Megapolis Èkspress’, 14.01.2005]

It was not possible to find any example of syntactic words usage either in the Integrum database or in the RNC. However, it seems that examples such as number (8) do not act as exceptions in Russian language usage and do not contradict the researcher’s own intuition.

- (8) — Мы зайдем в магазин не по пути в школу, а после нее.
— Ну, *после так после*...

Thus, the collected examples show that fillers of the variable X can interact with all word classes and have minimal restrictions regarding

³ The material has been collected from the Russian National Corpus (RNC), and the *Integrum* database; it consists of more than 4000 examples.

morphological forms (the subjunctive and the imperative mood of verbs and gerunds). At the same time, the assumption that there are modal-tense modifications (or a paradigm in terms of the [RG 1980]) for these phrasemes seems to be unacceptable. Examples such as number (9) seem completely artificial, and are not observed either in texts or in the speech of native speakers of Russian. (However, the second type of phrasemes under discussion does display such a possibility).

(9) *Ну что ж, *ехать так ехать было/будет*.

On the other hand, it is quite possible to analyze the following sentences (10—12) as examples of a united ‘modal-tense paradigm’.

(10) Ну что ж, *пить так пить*.

(11) Ну что ж, *пили так пили*.

(12) Ну что ж, *выпито так выпито*.

Due to the contextual dependence of the phraseme manifestation it looks unnatural that these phrasemes should have a paradigm at all: it is much more important that a certain word form, on which this type can be grounded, was (or was implied) in a previous context.

1.2. The phraseme of the utmost quality

The second type of phrasemes under discussion applies stronger restrictions. The structure of this phraseme is ‘*X так X*’, where *X* is an acceptable word form, and the meaning is fully described in [Veličko 1996, 19]: ‘a statement, accepting of an identified, or offered object, but in its best, fullest, most qualitative manifestation’.

Russian particles like *вот это, ну вот* work as a test context for this phraseme. The prosodic contour is determined by an increase in intonation towards the end of the phrase and can be accompanied by lengthening the stressed vowels. The nominative case of nouns (13) and adjectives (14); infinitives (15) and finite forms of verbs (16—17)⁴ serve as variables in the phraseme (see also [Švedova 1960, 95—96; Vsevolodova, Lim Su En 2002 115—116]).

(13) Вот *встреча так встреча!* — как обухом по голове ударил
Тараса бойкий голос невесть откуда взявшегося Простенки.
[O. Novikova. *Mušskoj roman*]

⁴ The verb forms are included in the non-strict opposition, where imperfective infinitives oppose perfective past tense forms: *гулять так гулять*, but *Погулял так погулял*. There is insufficient place to elaborate upon this subject here.

- (14) Ведь и там им не о чем бы было поговорить, вот штука-то. Вот уж *чужие так чужие* — на веки вечные. [B. Šukšin. Čužie (1974)]
- (15) *Казнить так казнить, жаловать так жаловать*: таков мой обычай. [A. S. Puškin. Kapitanskaja dočka]
- (16) Вот Чубайс *мочит так мочит* — сразу города с полумиллионным населением отключает, люди замерзают, голодают. [‘Sovetskaja Rossija’, 2003.04.05]
- (17) Сам буду выступать и уж *врезаю так врезаю*! [Ju. Trifonov. Dom na naberežnoj]

No example has been found consisting of numerals or syntactic word classes in the RNC or the *Integrum* database. Examples like (18—20) seem unnatural. Therefore we cannot claim that these word classes cannot be used in the phraseme.

- (18) *(Вот это) *вечером так вечером*!
- (19) *(Вот это) *тридцать пять так тридцать пять*!
- (20) *(Вот это) *из так из*!

In conclusion, the incorrectness of the examples below has to be mentioned (21—23).

- (21) *(Вот это) *романом так романом*!
- (22) *(Вот это) *красивым так красивым*!
- (23) ?(Вот это) *Москва так Москва*!

It is clear from the examples mentioned, that the nominatives⁵ of common nouns⁶ and adjectives can be used as variable fillings. It is worse to use adverbs and numerals; they are possible only if they have (or gain from the context) a certain value: *вот это много так много!* means ‘as much as it can be’; *Вот это пять так пять!* is possible in the substantive usage, and means, for example: ‘a firm handshake’. According to some special rules, infinitives and finite forms of verbs can be used as well; prepositional phrases and syntactic word classes cannot.

As it was mentioned, *the phrasemes of the utmost quality* can have modal-tense modifications (24—25).

⁵ The only example with a non-nominative case given in [Vsevolodova, Lim Su En 2002, 115] ‘*Я встретил вот женщину так женщину*’ seems to be peripheral; there is no a single clause in the material collected from Russian corpora.

⁶ There is only one possibility to use proper nouns. It is correct if the reference is split; for example, ‘*Вот это Москва так Москва, а то, что было раньше, — только пригороды*’.

(24) (Вот) была встреча так встреча!

(25) (Вот) будет встреча так встреча!

2. Discourse links

The behavior of the two types of phrasemes within the framework of a discourse, as far as the discourse links which these phrasemes possess, is fundamentally different.

2.1 The phraseme of an uncontrollable choice

This type of phrasemes leans upon the previous context. Agreeing with the interlocutor (or with the speaker her/himself) one often uses a word form already mentioned.

(26) Экипаж — шесть человек, — сказал Хурков, включая проверяющего. Всем — поровну. Разговор закончен. — Ну закончен — так закончен, — лениво проговорил Ильин, — но я считаю: за один только суп — многовато будет. [*‘Zvezda’*. № 3. 2002]

However, some change in the repeated forms is preferable when there are several word forms in the previous context, making it difficult which word form to choose. In the example below (27), the nearest form of the lexeme *ПАРИЖ* is a Genitive (the rest are in the accusative case). The speaker prefers a default nominative form rather than any of the possible alternative.

(27) Но тут объявили регистрацию на рейс, как ни странно, вылетающий до **Парижа**. ‘Ну что? — спросил я жену. — Может, остаться?’ Я и впрямь готов был выйти из игры, чтобы не становиться для всех обузой. Но тут кто-то спросил: ‘Ваш чемодан?’ — ‘Мой’. — ‘Ну так чего он тут стоит?’ — И потащил его на весы. Мы регистрировались по списку и по групповому билету. Ну что ж, *Париж*, так *Париж*... Будь что будет. [*‘Zvezda’*. № 4. 2002]

The synonymic replacement of a lexeme is also possible which, naturally, can lead to certain stylistic modifications. For instance, example (28), despite its stylistic awkwardness, exemplifies the permissible synonymic replacement of an initial lexeme by a final lexeme.

- (28) ‘Петрович! Ты должен, как Андрей Болконский, пасть со знаменем у финского дота!’ — дает мне последние указания наш командир Александр Федосов. *Умереть так умереть*: ‘Есть, товарищ лейтенант, не впервой!’ [‘Karelija’. 21.03.2002]

Thus, a possible discourse characteristic of phraseme usage rests on the fact, that there is contextual support for the phraseme (often explicitly mentioned in the interlocutor’s speech), upon which the phraseme is based, with possible grammatical, lexical, or stylistic modifications.

2.2 The phraseme of the utmost quality

It appears that the text links of the phraseme are much more independent. This is proved by the fact, that it can be used freely either in a position of the beginning of a text or in a heading.

- (29) *Меняться — так меняться*. Молдова обещает поставить точку в вопросе делимитации границы с Украиной завтра [‘Den’’, 11.07.2001, № 120]
- (30) Новый русский на отдыхе: *отрывать так отрывать!* [‘Delovaja Penza’, 12.01.2000, № 168]

However, it is clear that these phrasemes are frequently entered with a reference to precedent or pseudo-precedent texts. Markers of such ‘citation’ are the expressions like *как говорится, как говорят, N сказал* etc. (31—32).

- (31) Как поет Розенбаум: ‘*Стрелять так стрелять!*’ Время к этому афоризму добавило: ‘Взрывать так взрывать!’ [‘Moskovskaja pravda’, 28.09.1996, № 8 (178)]
- (32) Владимир Лукин: Конечно, как говорится, *трясти, так трясти*. Конечно, если сейчас решить проблему перезахоронения Ленина, а я не сторонник этого, но раз уж начали одни символы трясти, значит надо другие символы трясти. [‘NTV’: Geroj dnja, 06.12.2000, 19:30]

Furthermore, a speaker may label as a citation such expressions for which there are no real sources at all. For instance, example (33) has been found only once: it has been in vain trying to locate a source for the clause below, either in the RNC or in the Integrum Database or in the Ru-net (Russian Net) in whole.

- (33) ‘Как говорится, *складировать так складировать!*’ [‘Peterburgskij stroitel’nyj rynek’, 15.01.2002; № 1—2 (45)]

Thus, a possible discourse characteristic of phraseme usage is that they are usually free from the immediate context, but often have links to the overall mental space, including citations, precedent texts, or other common knowledge.

3. Some frequency data

This part of the article pursues the idea that the usage of the type of phrasemes under discussion is based on a dependent relationship between frequency data and a cognitive mechanism that directs the usage of the phrasemes.

A search for examples in the big corpora has revealed that the most frequent forms are infinitives. Establishing a bank of examples comparable in quantity for the other word classes and word forms would demand a considerable effort. For this reason only syntactic phrasemes including infinitives have been collected. The material collected contradicts the hypotheses that there is a dependent link between the number of ‘Inf *так* Inf’ representations and the number of corresponding verbal lexemes in the Russian language.

According to [Zasorina 1977] and [Šarov 2002] the ten most frequent Russian verbs are the following: *быть, сказать, мочь, говорить, знать, стать, есть, хотеть, видеть, идти* (Šarov gives the verb *видать* instead of *есть* given in [Zasorina 1977]). However, the most frequent clauses of the ‘Inf *так* Inf’ type using verbs from the list are *сказать так сказать* and *есть так есть*. There are no examples with the verbs *мочь, знать, стать, хотеть, видеть*⁷ in the data. This proves that there is no link between the frequency of the phraseme and verbs with both types of the phrasemes. In the case of the *phraseme of an uncontrollable choice* it seems that a long list of verbs (and other lexemes, in all probability) is freely used, without any correlations between the frequency of the characteristics of the phraseme and the lexemes that constitute it.

In the case of the *phraseme of the utmost quality*, the most commonly used is the clause *Гулять так гулять*. The Integrum database has more than one and a half thousand examples. That number exceeds that of all other lexemes within the framework of the syntactic phrasemes under discussion. The following clauses noticeably lag behind, but,

⁷ It does not appear that there is a clear explanation why the clauses like **Мочь так мочь* etc. are impossible.

nevertheless, often appear: *Стрелять так стрелять* (more than 350 examples), *Любить так любить* (more than 200 examples), *Летать так летать* (about 180 examples), *Играть так играть* (about 170 examples), *Работать так работать* (about 150 examples), *Пить так пить* (about 130 examples), *Лечить так лечить* (about 90 examples).

Naturally, it is useless searching for an explanation for this phenomenon by looking at the frequency with which verbs are used. (the verb *стрелять*, for example, does not even enter the top one hundred most frequently used Russian verbs). A much more plausible reason may be, that the most frequently used phrasemes are those that reflex back to particular precedent texts: five verbs from the list mentioned, namely, *гулять*, *любить*, *стрелять*, *летать*, *лечить* can be traced back to A. Rosenbaum's popular smash hit 'Duck Hunting'.

- (34) Я помню, давно учили меня отец мой и мать:
Лечить — так лечить! Любить — так любить!
Гулять — так гулять! Стрелять — так стрелять!
 Но утки уже летят высоко
Летать — так летать! Я им помашу рукой.

To be precise, the above-mentioned phrasemes can be found in the masterpieces of N. A. Nekrasov, N. M. Jazykov, A. Pristavkin, and others. However, these texts can hardly be considered responsible for the coining of the phrases in question. On the other hand, the fact that high-frequency phrasemes are often found in the context of reasoning about the Russian soul and, thus, refer to unclear, but common concepts (see 35—36).

- (35) Русская душа любит гулять так гулять, пить так пить.
 ['Birža', 14.06.2001; № 21]
 (36) У нас ведь как — ругаться так ругаться, спорить — так спорить, выпить — так выпить, работать — так работать.
 ['Večernij Novosibirsk', 30.11.2001]

In aggregate, these data reveal a dependence of *the phraseme of the utmost quality* not on the rate of use of a certain lexeme, but on speech usage. The observation shows that the frequency rate depends on a knowledge of popular texts, or on the aspiration of the speaker to motivate the usage by means of creating 'a well-known piece of information'. By contrast, the other phraseme, which depends on its immediate context, does not submit to any frequency trend.

4. The formal description

The phraseme of an uncontrollable choice relies on its direct context and includes free word forms of various word class in its structure. It is represented that generative rules, rooted in Grammar, allow the creation of a potentially infinite quantity of clauses, constructed on the basis of this idiomatic syntactic model. Apparently, within a linguistic description this type should be specified by certain rules at the Grammar level. At the same time, for accurate representation of the phraseme it is necessary to include information about conditions imposed by the particular frame. Without such a frame a syntactic phraseme can be created, but it will be impossible to use it correctly. This frame may be termed ‘The Realization of a Choice Frame’, denoting the presence of an alternative (probably not explicitly shown) and the required choice, that is the focus of the frame (compare [Nikolina 1991, 55]. A generalized description can be set out as follows:

Frame:	The Realization of a Choice
Structure:	‘... <i>X mak X</i> ’, where <i>X</i> is a variable (any word form with a short list of constraints), having a semantic antecedent (probably not-expressed or expressed with a synonym) and taking the form of an antecedent or default form.
Prosodic contour:	A falling intonation towards the end of a phrase.
Meaning:	‘ <i>an uncontrollable choice</i> ’: unwilling (or insufficiently motivated) consent; a lack of choice, a nothingness of choice.
Pragmatics:	Obligatory context (marked in the above-mentioned scheme as ‘...’).

The phraseme of the utmost quality is more independent from its direct context, but connected with more strict rules of usage. In this sense, seemingly, there is no specialized frame fixed to this phraseme. Its grammatical description may look something like this:

Structure:	‘ <i>X mak X</i> ’, where <i>X</i> is a variable, having the following: the nominative of adjectives, nouns (except common ones), finite verbs, and infinitives (rather imperfective infinitives and perfective past forms).
Prosodic contour:	An increasing intonation forward the end of a phrase (can be accompanied by a lengthening of stressed vowels).

Meaning:	‘utmost manifestation of a signified’.
Pragmatics:	probable connotation of a well-known experience, referring to a precedent (or pseudo-precedent) text.

The proffered generalized description allows a speaker (including a foreigner) to generate an infinite number of correct Russian phrases. Excluding pragmatic information, such phrasemes will be correct. At the same time, the existence of a ready-to-use high-frequency cliché like *Гулять так гулять* in real speech allows one to assume, that some clauses are not generated by the rules defined above, but stored in the memory. In such a case, it should be supposed, that phrasemes under consideration are very closely tied in with lexical idioms like *Умирать так с музыкой*, which cannot be generated by any rules, but stored in the Dictionary. From the semantic point of view there are some similar expressions in Russian, having certain idiomatical restrictions either to the Grammar (see the above-mentioned list of the morphological restrictions) or to the Dictionary (for example, a list of lexemes for *Умирать так с...* includes no more than dozen of items). This conclusion corroborates the conception that ‘a strict separation of lexicon and grammar, <...> may prove to be but a methodological prejudice’ [Jackendoff 1995, 155—156, compare Bybee 1985, 111—135].

It has been assumed, that such a continuum is stored as a ‘construction’ [Schultze-Berndt 2002, 302]. However, it seems more likely that there are two ways for syntactic phrasemes to be stored and used. One method is that of ready-to-use speech patterns which pull together with lexical phrasemes; they are not generated every time and are taken from the Dictionary. As a rule, they contain more connotative information that is typically for the Lexicon. Another method demands the use of some grammatical rules. Naturally, these clauses contain less connotative information, as it is usually located in the Grammar. Most likely, during speech generation a speaker refers to the Dictionary, and — if there is no help — the clause is generated by certain rules.

References

- Bybee, J. L., 1985: *Morphology: A Study of the Relation between Meaning and Form*. Amsterdam: John Benjamins, 1985.
- Fillmore, Ch., Kay, P., & O’Connor, M., 1988: ‘Regularity and Idiomacity in Grammatical Constructions: The Case of Let alone’. *Language*, 64/3. 1988. P. 501—538.

- Goldberg, A., 1995: *Constructions. A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press, 1995.
- Jackendoff, R., 1995: 'The Boundaries of the Lexicon'. M. Everaert, E-J. van der Linden, A. Schenk, and R. Schreuder (eds.), *Idioms: Structural and Psychological Perspectives*. Hillsdale, New Jersey, Hove, UK: Lawrence Erlbaum Associates, 1995. P. 133—165.
- Kay, P., Fillmore, Ch., 1999: 'Grammatical Constructions and Linguistic Generalizations: The *What's X doing Y?* Construction'. *Language*, 75/1. 1999. P. 1—33.
- Lakoff, G., 1987: *Women, Fire, and Dangerous Things: what Categories Reveal about the Mind*. Chicago: University of Chicago Press, 1987.
- Mel'čuk, I. A., 1995: 'Phrasemes in Language and Phraseology in Linguistics'. M. Everaert, E. J. van der Linden, A. Schenk, and R. Schreuder (eds.), *Idioms: Structural and Psychological Perspectives*. Hillsdale, New Jersey, Hove, UK: Lawrence Erlbaum Associates, 1995. P. 167—232.
- Mel'čuk, I. A., 1995a: 'Un affixe dérivationnel et un phrasème syntaxique du russe moderne'. *Русский язык в модели 'Смысл — текст'*. Vienna/Moscow: Wiener Slawistischer Almanach/Škola 'Jazyki ruskoj kul'tury', 1995. P. 325—346.
- Nikolina, N. A., 1991: 'Pragmatičeskij aspekt opisanija predložnij frazeologizirovannoj struktury'. *Funkcional'nyj sintaksis i semantika*. Moscow: Izdatel'stvo Universiteta družby narodov, 1991. P. 47—58.
- RG, 1980: *Russkaja grammatika*, N. Ju. Švedova (ed.), Moscow: Nauka, 1980.
- Šarov, S. A., 2002: *Častotnyj slovar*. Available at <http://www.artint.ru/projects/frqlist.asp>.
- Schultze-Berndt, E., 2002: 'Constructions in Language Description'. *Functions of Language*, 9/2. 2002. P. 269—310.
- Švedova, N. Ju., 1960: *Očerki po sintaksisu russkoj razgovornoj reči*. Moscow: Nauka, 1966.
- Veličko, A. V., 1996: *Sintaksičeskaja frazeologija dlja russkix i inostrancev*. Moscow, 1996.
- Vsevolodova, M. V., Lim Su En, 2002: *Principy lingvističeskogo opisanija sintaksičeskix frazeologizmov*. Moscow: MAKSPress, 2002.
- Zasorina, L. N. (ed.), 1977: *Častotnyj slovar' russkogo jazyka*. Moscow: Nauka, 1977.

Влияние МСТ на общую лингвистику (в свете подведения предварительных итогов «семантической революции»)

С. А. Крылов

Институт востоковедения РАН
krylov-58@mail.ru

Предметом рассмотрения являются те идеи теории моделей «Смысл $\Leftrightarrow$ Текст», которые «вышли за пределы» школы МСТ и стали достоянием общей лингвистики за рамками этой школы. Лейтмотив МСТ — трактовка языковых явлений в рамках описания, способного эффективно использоваться автоматом («компьютером») как источник сведений, для осуществления машинного перевода. Отсюда вытекают более частные особенности МСТ.

The impact of Meaning-Text Model on general linguistics (in the light of preliminary results of the “semantic revolution”). Our subject will be those ideas of MTT, which overcome its “borders” and became part of general linguistics outside the borders of this school. The main principle of MTT is the treatment of every linguistic phenomena in the framework of such a description which can be effectively used (implemented) by an automaton (a computer) as a source of data for the successful work of a machine translation system. Basing on this nuclear principle, one can explain all specific peculiarities of MTT.

Предмет доклада — обсуждение идей, внесенных в общую лингвистику из теории моделей «Смысл $\Leftrightarrow$ Текст» (МСТ)¹. Предметом рассмотрения являются лишь те идеи, которые «вышли за пределы» школы МСТ и таким образом стали достоянием общей лингвистики за рамками этой научной школы².

¹ В задачу доклада **не входит**: ни анализ причин возникновения и угасания «научной моды» на МСТ; **различий в перцепции** МСТ отечественными и зарубежными лингвистами; влияния идей МСТ на описательный аппарат **частной** лингвистики; ни библиография работ по МСТ.

² Поэтому предметом рассмотрения **не будут**: ни спорные положения МСТ, ни такие принципиально важные (и, более того, кардинальные с

Главный принцип МСТ — новый угол зрения на языковые единицы. С точки зрения МСТ, любую языковую единицу U (шире, любое явление E, и шире — любую проблему) имеет смысл рассмотреть под специфическим углом зрения, а именно: как именно следует произвольную единицу U (шире, явление E) трактовать в рамках такого описания данного языка, которое способно эффективно использовать некоторым реальным (= уже построенным), или мыслимым как реальный, автоматом (в терминах 1960-х гг.) или «компьютером» (в терминах 1990-х гг. и нашего времени) в качестве источника сведений, позволяющих успешно осуществлять процесс машинного перевода. Из этого лейтмотива вытекают все более частные особенности МСТ.

МСТ внесла значительный вклад в описание и морфологии, и синтаксиса, и семантики.

В морфологии это уточнения к определениям таких важных понятий морфологической теории, как разграничение грамматических и неграмматических значений, синтаксических и семантических значений, морфологических и неморфологических средств (начатые И. А. Мельчуком ещё в конце 1950-х гг.). В ходе этих уточнений в первой половине 1960-х гг. были предложены (в ходе уточнения понятия «аффикс») такие понятия, как «циркумфикс», «трансфикс»³. Во второй половине 1960-х гг. обоснована идея применения понятия конверсии за пределами перехода частей речи друг в друга (и введено понятие «субкатегориальной конверсии»), расширено понятие супплетивных отношений в словообразовании (и распространено на сферу «синтаксического» словообразования, ср. *полагать* — *мнение*), продемонстрирована возможность расхождения между иерархией формальной простоты/сложности (= формальной производности) и иерархией отношения семантической простоты/сложности (= семантической производности) (ср. *катить* — *катиться*, где возвратный глагол семантически проще, чем невозвратный, тогда как формальная производность — обратная). В 1970-е гг. было предло-

точки зрения внутренней самооценки авторами МСТ своих работ!) ее достоинства, как: глобальность модели (модулярность композиции плюс взаимная согласованность модулей); создание единой системы понятий для морфологии, разработка формального языка, в рамках которого можно делать строгие выводы; выбор направления описания «от смысла к тексту».

³ Сам термин «МСТ» стал употребляться в печати лишь начиная с 1967 г., когда была опубликована статья А. К. Жолковского и И. А. Мельчука «О семантическом синтезе». Однако большинство идей И. А. Мельчука, выдвинутых в 1958—1966 гг., впоследствии стало естественной частью МСТ.

жено распространить противопоставления «алло-» единиц «эмическим» единицам со сферы сегментных показателей на сферу несегментных знаков — таких, как апофония, значащая мена тона и др. В 1982 г. И. А. Мельчук издал краткий словарь морфологических терминов, сопровождавшийся итоговой схемой межпонятийных отношений. Схема была построена в виде связного ориентированного графа, рёбра которого, помеченные стрелками, отображали направленное отношение смысловой сложности между терминами, употреблёнными автором в «левой» (*definiendum*) и «правой» (*definiens*) частях дефиниции. Позже, в 1990-е гг., эта концепция легла в основу своеобразной энциклопедии морфологических терминов (6-томный «Курс общей морфологии» Мельчука).

Если в сфере морфологии вклад МСТ в общую лингвистику состоял преимущественно в уточнениях к определениям относительно известных понятий (и лишь во вторую очередь — во введении новых понятий), то в сфере синтаксиса и семантики МСТ привела к введению именно новых понятий, служащих инструментом синтаксического и семантического описания, то есть понятий, почти совсем не разработанных лингвистикой предшествующих периодов.

В синтаксисе новой была и формулировка основных положений «синтаксиса зависимостей» (СЗ), и разграничение поверхностного и глубинного СЗ.

Первые опыты построения графов зависимостей (ГЗ) проводились А. М. Пешковским и его школой ещё в 1920-е гг. (хотя и довольно эскизно). Затем, начиная с 1930-х гг. и до середины 1950-х — уже весьма тщательно продуманную и детально реализованную теорию ГЗ («стемм») развивал Л. Теньер.

Однако СЗ И. А. Мельчука был нацелен на достижение такой степени формализованности синтаксического описания, которая бы позволяла использовать его в системах автоматического перевода. Именно это привело и к новым решениям многих вопросов синтаксического описания, и к новой постановке самих вопросов.

Так, в МСТ (в отличие от концепций Пешковского и Теньера) в качестве синтаксического элемента (минимальной синтаксической единицы, т. е. отдельного узла графа) выступает не «минимальный член предложения» (= «минимальный ответ на вопрос»), а минимальная подвижная цельнооформленная единица — словоформа. Все служебные слова, связки и вспомогательные слова (с помощью которых строятся аналитические формы слов) трактуются как отдельные синтаксические единицы. Поэтому в предмет синтаксиса были включены правила взаиморасположения (и вообще распо-

жения в тексте) таких единиц, которые по протяжённости меньше, чем минимальный «член предложения» — например, компонентов «предложно-падежной формы» существительного или компонентов «аналитической формы» глагола ⁴.

Важным вкладом МСТ в синтаксис является теория нулевых синтаксических единиц. Если в работах А. М. Пешковского к их числу относилась лишь нулевая форма, заполняющая позицию сказуемого, то в работах И. А. Мельчука понятие нуля расширилось, охватив также нулевые подлежащие в безличных (*Всю улицу засыпало песком*) и неопределённо-личных (*Всю улицу засыпали песком*) предложениях.

Новым шагом в синтаксисе явилось введение расширенного понятия синтаксического отношения (СО), предложенное в МСТ. Оно позволило описать то общее, что есть у отношений, традиционно трактуемых в качестве синтаксических (предикативное, комплетивные, определительные, обстоятельственные...) и у других видов СО между словами, традиционно не включаемых в номенклатуру СО. «Традиционные» СО (напр., описанные А. И. Смирницким) усматривались лишь между «членами предложения». К «новым» же видам СО можно отнести, в частности:

- 1) ограничительное (между частицей и сферой действия частицы);
- 2) вводное (между вводным оборотом и самостоятельной предикацией, к которой он относится);
- 3) сравнительные отношения (между компаративной формой и сравнительным оборотом);
- 4) элективное отношение (между вершиной элективной конструкции и её подчинённой частью);
- 5) отпредложное и отсоюзное СО (между служебным словом и тем знаменательным словом, которое заполняет его валентность);
- 6) присвязочные СО (между связкой и именной частью сказуемого);
- 7) СО между частями аналитических форм слов.

Между тем в традиции фактически отсутствовало именно необходимое обобщённое понятие для разных видов межсловных СО.

⁴ На первый взгляд, такой подход кажется тривиально необходимым. Однако синтаксической традицией он не предусмотрен вовсе! Так, в русском языке (РЯ) предлог ставится в препозиции по отношению к той именной группе, к которой он относится; но найти «традиционную» грамматику РЯ, где бы содержалось вышеуказанное правило расположения слов, оказывается, нелегко. Искомое правило не входит ни в «синтаксис предложения», ни в «синтаксис словосочетания», ни в «морфологию» (повисая как бы «в вакууме»).

И само введение обобщённого понятия СО (безусловно, необходимое), и конкретные решения дилемм синтаксического представления, принятые И. А. Мельчуком в рамках СЗ, явились следствием принятия важнейшего принципа — постулата, лежащего в основе всей синтаксической концепции МСТ. Он касается фундаментального онтологического свойства синтаксических конструкций вообще (и, как следствие, синтаксических структур предложений), которое отражается в виде постулата, налагающего ограничение на формальные свойства синтаксического графа (отображающего синтаксическую структуру): «этот граф должен быть связным» (Мельчук 1974: 212).

Приняв «постулат связности», следует признать, что единственно осмысленный способ конструктивной критики конкретных синтаксических решений⁵, предлагавшихся в русле МСТ — это построение таких альтернативных способов описания синтаксиса, при которых «постулат связности» продолжал бы соблюдаться.

Разграничение единиц глубинного и поверхностного СЗ⁶ позволило описать ряд различий между формальными и содержательными характеристиками синтаксических единиц.

Так, идиоматическое словосочетание обладает известной синтаксической цельностью, что сближает его со словоформой. Этот факт фиксируется в терминах МСТ так: на глубинно-синтаксическом уровне представления идиома признаётся единым «узлом» синтаксического графа (т. н. «обобщённой лексемой») — подобно обыкновенной (однословной) номинативной единице.

С другой стороны, компоненты идиом обладают некоторой степенью синтаксической самостоятельности — они вступают в отношения согласования и управления, иногда отрываются друг от друга линейно или даже допускают перестановку местами; эти свойства сближают их с однословными номинативными единицами. В терминах МСТ, на поверхностно-синтаксическом уровне идиома признаётся сочетанием нескольких синтаксических «узлов», что сближает её со свободным словосочетанием.

В качестве таких «обобщённых лексем» оказывается возможным описывать также некоторые виды устойчивых сочетаний полуслужебных глаголов с существительными предикатного (ситуационного,

⁵ Английский СЗ описан в Mel'čuk & Pertsov 1987; русский СЗ рассеян по малотиражным публикациям его фрагментов (в 1970-е гг.).

⁶ Непосредственным источником влияния на МСТ был, по-видимому, американский генеративизм (в первую очередь — концепция Н. Хомского).

процессного, событийного) значения, где ядро лексического значения выражено подчинённым членом словосочетания — существительным. Так, в качестве «обобщённых лексем» описываются такие словосочетания, как *стоять на вахте*, *идти на компромисс*, *совершать преступление*, *быть в лихорадке*, *иметь обыкновение*, *вести переговоры*, *быть в лихорадке*. Это позволяет более систематично отразить перифрастические возможности соответствующих глагольно-именных сочетаний. Экстравагантным техническим средством, наглядно иллюстрирующим указанный принцип, стало введение так называемых «фиктивных лексем» (**вахтит*, **компромиссировать*, **преступать*₂, **обыкновенствовать*, **переговариваться*₂, **лихорадиться*).

Различение глубинного и поверхностного синтаксиса является естественным следствием принятия гипотезы о том, что у СО есть некоторый универсальный аспект (универсальный относительно различий между языками)⁷. На основании этой гипотезы авторами МСТ был разработан метаязык описания инвентаря «глубинных» СО, предположительно наличествующих во всех языках мира: к ним относятся актантные отношения⁸, атрибутивное и сочинительное СО.

Здесь, как и в теории поверхностного синтаксиса, следует различать два аспекта: с одной стороны, сам базовый принцип построения метаязыка глубинного синтаксиса на основе ограниченного инвентаря отношений, служащих общей смысловой основой для СО всех языков мира (этот принцип созвучен концепциям О. Есперсена, Ш. Балли, Е. Куриловича, Л. Теньера, Ю. С. Мартемьянова, В. Г. Гака), а с другой стороны, конкретные попытки реализации этого принципа при разработке конкретных версий такого метаязыка (здесь многое дискуссионно ср. разные версии З. М. Шаляпиной, Н. Н. Леонтьевой, Е. В. Падучевой, А. Е. Кибрика).

Важной частью концепции МСТ явились такие понятия, как «валентность» и «модель управления» (а также «диатеза»), обогатившие одновременно теорию грамматики⁹, лексикологию и лексикогра-

⁷ Эта гипотеза фактически восходит к идеям грамматики Пор-Рояля (1660 г.) и разделялась, по-видимому, всеми представителями «логических подходов» в синтаксисе.

⁸ Разных актантных отношений усматривалось в ранних версиях МСТ (в 1960-е гг.) всего 4, затем (с середины 1970-х) стало усматриваться 6.

⁹ В частности, на основе понятия диатезы была построена дедуктивная теория исчисления теоретически возможных типов залогов.

фию. Понятие «способа выражения валентности» позволило подвести общую базу под все синтаксические средства, сходные по функции: выбор синтетического падежа, выбор предложно-падежной формы, оформление подчинённой предикации инфинитивным оборотом, оборотом с отглагольным именем или придаточным оборотом.

Наиболее оригинальным оказался вклад МСТ в теоретическую лексикологию и семантику: были уточнены основные понятия, связанные с описанием смысловых сходств между словами; введено понятие лексической функции (ЛФ).

Уже в начале XX в. свои права завоёвывал ономаσιологический подход в лексикологии (направление «Слова и вещи») и в грамматике (Ф. Брюно, О. Есперсен, учение о понятийных категориях); но основные грамматические и лексические явления не получали должного освещения в ономаσιологическом аспекте. Для проведения последовательно ономаσιологического взгляда на языковые явления необходимо выйти за пределы чисто грамматической постановки задач (как это делалось в грамматической семасиологии) и чисто лексикологической постановки задач (как это делалось в лексической семасиологии), а рассматривать взаимодействие лексических средств с грамматическими при обозначении всей ситуации целым предложением. Вплоть до 1960-х гг. семантика целого предложения (а не только его частей) изучалась преимущественно логиками, а не лингвистами. Лишь в 1960-е гг. проблема сентенциальной семантики стала решаться с точки зрения ономаσιологического подхода (от ситуации — к её языковому оформлению). Важную роль в осуществлении такого поворота мыслей сыграли работы Ю. С. Мартемьянова и В. Г. Гака.

В работах А. К. Жолковского и И. А. Мельчука по разработке МСТ был предложен новый, во многом диаметрально противоположный традиционному, подход к самому феномену синонимии. Если в традиционной лексикологии понятие синонимии основывалось на идее полной или частичной взаимозаменяемости синонимов в некоторых (или во всех) контекстах (трактуемой как доказательство факта синонимичности), то в МСТ внимание было привлечено к тому, что синонимические отношения не только не предполагают взаимозаменяемости синонимов, но, строго говоря, вообще независимы от неё. Поэтому отношения синонимии (понимаемой как семантическая эквивалентность) нередко устанавливаются между словами, сочетаемостные возможности (так называемые «синтактики») которых не пересекаются вообще (таких, как *закадычный*, *заклятый*, *проливной*, *заядлый*, *горький*, *глубокий*)! Пафос той «научной револю-

ции»¹⁰, которую МСТ произвела в лексикологии, состоял в том, чтобы от изучения лексических синонимов, находящихся в отношениях свободного варьирования (таких, как *бросить* — *кинуть*), перейти к изучению дополнительного распределения лексических синонимов. Для изучения семантической эквивалентности была разработана теория «лексических функций» (ЛФ).

В результате той «Семантической Революции» (СемР), которую лингвистика пережила в 1959—1974 гг.¹¹, предмет и задачи семантики были радикально пересмотрены. До СемР семантика понималась как часть (раздел) лексикологии, методологически подчинённая лексикологии. Между тем в результате «СемР», наоборот, лексикология стала пониматься как «лексический» раздел семантики!

Сегодня, когда со времён «СемР» прошло более 30 лет, можно предположить, что истина лежит где-то посередине. Ни лексикология (теория нерегулярных, словарных, индивидуальных явлений) не составляет раздела семантики (как это стали думать «после

¹⁰ В смысле Т. С. Куна.

¹¹ Хронологические рамки происходившей «СемР» уместно очертить так: 1959 — «предвестье» СемР: выступления Р. О. Якобсона и вслед за ним Вяч. Вс. Иванова о смысле как инварианте перевода; статьи В. Г. Гака о сопоставлении структур предложения при переводе;

1960—1961 — начало СемР: манифест лаборатории 1 МГПИИЯ (А. К. Жолковский, Н. Н. Леонтьева, Ю. С. Мартемьянов) о принципиальном использовании смысла при машинном переводе; статьи В. Г. Гака о сопоставлении семантики слов при переводе;

1964 — пик СемР: сборник МПИП № 8 (со статьями А. К. Жолковского о семантическом анализе, Ю. С. Мартемьянова о записи ситуаций и др.);

1965—1973: постепенная фиксация результатов СемР:

1965—1967: система «семантического синтеза», теории «лексических функций» (А. К. Жолковский, И. А. Мельчук);

1965—1968: теория «лексико-грамматической организации предложения» (В. Г. Гак);

1965—1973: теория денотативного тождества (Е. В. Падучева);

1968—1976: толково-комбинаторный словарь (редакторы А. К. Жолковский, И. А. Мельчук);

1969 — 1972: теория «семантических примитивов» (А. Вежбицкая);

1974: подведение итогов произошедшей СемР (появление «итоговых» монографий И. А. Мельчука, Ю. Д. Апресяна, Е. В. Падучевой).

То, что к началу 1970-х гг. интенсивно шло именно закрепление результатов **уже завершённой** СемР, полуэксплицитно выражено Ю. Д. Апресяном: «Нынешняя эпоха развития лингвистики — это, б е с с п о р н о, эпоха семантики» (Апресян 1974: 3) (разрядка моя!).

СемР») ¹², ни семантика (теория значений) не составляет раздел лексикологии (как это думали «до СемР»).

Лексикология (наука о нерегулярном) видит синонимию (лексическую) только в парах типа «*глядеть* — *смотреть*», а для семантики синонимия (семантическая) включает также пары типа «*бодрствовать* — *не спать*» (лексикология такие пары не считает синонимическими, да и вообще игнорирует, так как второй член пары вообще не есть лексическая единица, а представляет собой свободное сочетание). Лексикология трактует пары типа «*бодрствовать* — *спать*» как антонимы (лексические), а семантика — как разновидность «синонимических средств языка», так как антонимы (лексические) образуют синонимическую пару (пару «семантических синонимов») в контексте отрицания. Зато лексикология не включает пары типа «*спать* — *не спать*» в число антонимических (т. е. пар «лексических антонимов»), ибо второй член пары есть не лексическая единица, а словосочетание; а для семантики такие пары (члены которых суть «семантические» антонимы) представляют первостепенный интерес.

Вкладом МСТ в лексикологию ¹³ и семантику явилась толково-комбинаторная лексикография ¹⁴.

Обратим внимание на двуплановость одного из центральных понятий МСТ — понятия ЛФ.

С одной стороны, ЛФ — это инструмент систематизации типовых отношений между словарными единицами (они подразделяются на два типа: 1) синтагматические, между компонентами сочетаний

¹² Так, проблемы лексикологии, имеющие к семантике довольно слабое отношение или вообще не имеющие отношения — это шкала «центр vs. периферия» в лексике (степень употребительности слов, методы ее измерения), омонимы (для семантики омонимы — явление сугубо случайное, если не считать периферийного раздела семантики — «семантики каламбура»), пути пополнения лексического состава (заимствования, кальки, исконная лексика), внутренняя форма и ее забвение, этимология (частично), ономастика (частично), терминология и терминография (частично), «формально-ориентированная» лексикография (орфографическая, орфоэпическая, формально-грамматическая, акцентологическая, морфемная, словообразовательная, рифмическая); а также инверсионная и частотная лексикография.

¹³ Своеобразный компендиум «обновленной» лексикологии (где главные понятия лексической семантики пересмотрены и уточнены под углом зрения МСТ) — Апресян 1974.

¹⁴ Ее итогом стал «Толково-комбинаторный словарь русского языка» (Вена, 1984).

типа *дождь* — *проливной* и 2) парадигматические, между коррелятами типа *дождь* — *ливень*, *полагать* — *мнение* и т. п.).

С другой стороны, ЛФ — это инструмент систематизации смысловых отношений между двусторонними единицами языка безотносительно к тому, являются ли эти единицы лексическими (словарными) или перед нами единицы регулярно образуемые. Авторами МСТ отмечалось, что грамматический способ выражения ЛФ Conv — это залог, ЛФ A₁ и A₂ — это действительные и страдательные причастия, ЛФ Imper — это императив и т. п. Этот ряд можно продолжить. Так, добавление частицы *не* — выразитель ЛФ Anti, номинализация предложения с помощью оборота *то, что...* (в русском) или *масдара* (в арабском, грузинском) или причастия (в алтайских языках) — выразитель S₀, номинализация с помощью оборота *тот, кто...* (в русском) или агентивного причастия (в монгольском) — выразитель S₁; множественное число — выразитель «склеенной» ЛФ //Mult; единственное число — выразитель «склеенной» ЛФ //Sing; наречный интенсификатор *очень* и адъективный интенсификатор *большой* — выразители ЛФ Magn; система пространственных предлогов (в предложных языках), пространственных послелогов (в послеложных языках) или локализационных падежей (в языках с богатой падежной системой) — выразители ЛФ Loc_{in}, Loc_{ab}, Loc_{ad} и т. п.

Применительно к регулярным выразителям ЛФ термин «лексический» должен считаться сугубо условным, ибо фактически уместнее говорить о некотором более общем виде отношений (например, об «ономасиологических функциях») (среди проявлений которых есть, в частности, и некоторые чисто «синтаксические функции», по терминологии Ю. С. Мартемьянова). С позиций «интегрального» (или «сентенциального») (а не «лексикологического») подхода к семантике, сформировавшегося в эпоху после «СемР», наличие регулярных выразителей функций означает, что данное явление отнюдь не уместается в рамки лексикологии, а принадлежит скорее общей ономасиологии¹⁵.

В целом, по сравнению с «революционной» эпохой 1958—1974, период усовершенствования МСТ в 1974—1980 гг. уместно считать периодом «реформ» (временем «эволюционного» развития теории в рамках одной и той же «научной парадигмы»), так как концептуаль-

¹⁵ Не случайно во II т. «Курсе общей морфологии» Мельчука некоторые виды грамматических значений (преимущественно словообразовательных) предлагается описывать с помощью аппарата ЛФ.

ные новшества в этот (как и в более поздний период) касались лишь частных деталей, а не принципиальных положений теории¹⁶.

В 1980 г. в русле усовершенствования теоретического аппарата МСТ¹⁷ Ю. Д. Апресяном было выдвинуто понятие «нетривиальных семантических признаков» и разработана классификация «правил взаимодействия значений» (среди которых — «лексемно-лексемные», «лексемно-граммемные» и «граммемно-граммемные»). Эти понятия, наряду с классификацией типов лексикографической информации (в рамках «интегрального подхода» к языку, лежащего в основе «системной лексикографии»¹⁸) означали существенное обогащение наших представлений о лексикографически релевантных свойствах слов.

Литература

- Апресян 1974 — Апресян Ю. Д. Лексическая семантика. М.: Наука, 1974. 367 с.
- Мельчук 1974 — Мельчук И. А. Опыт теории лингвистических моделей «Смысл $\Leftrightarrow$ Текст». Семантика. Синтаксис. М.: Наука, ГРВЛ, 1974. 315 с.
- Mel'čuk, Pertsov 1987 — Mel'čuk, Igor A. & Pertsov, Nikolaj V. Surface syntax of English. A formal model within the Meaning-Text framework. Amsterdam-Philadelphia: John Benjamins, 1987. 526 p.

¹⁶ Важнейшее из концептуальных новшеств, введенных в МСТ в книге Апресян 1974, — это фактическое включение в теорию МСТ проблематики семантической правильности (описываемой правилами «семантического согласования» и «взаимодействия значений»).

¹⁷ Некоторые шаги, направленные на такое усовершенствование, содержались уже в книге Апресян 1974 (так, было предложено включать в предмет изучения МСТ «правила взаимодействия значений» (ср. используемое понятие «семантического согласования») и, шире, вообще проблему «семантических аномалий».

¹⁸ Теоретическая концепция этого направления (существующего около двух последних десятилетий) составляет особую страницу в истории отечественной теоретической лингвистики и потому здесь не анализируется.

Underspecification of Quantifier Scope in MTT

Leonardo Lesmo, Livio Robaldo

Dipartimento di Informatica and Centro di Scienze Cognitive,
Università di Torino,
Email: {lesmo, robaldo}@di.unito.it

Sentences with more than one quantifier can have different interpretations depending on the scope order of their quantifiers. To our knowledge, this type of ambiguity has received a lot of attention in the Underspecification literature but not in the MTT one. In this paper, we define a new underspecified *SemS* of MTT which refers to all possible readings of a sentence. In order to specify a particular scope order, we introduce additional dependencies. They roughly correspond, in FOL, to Skolem constants and functions and are part of a new additional structure of *SemR* level, named *SemCtxDepS*.

1. Introduction — Meaning Text Theory (MTT)

The key idea of MTT ([Mel'čuk and Zolkovskij 67], [Mel'čuk 88], [Mel'čuk 00]) is that a natural language is an m:n relation between *meanings* and *texts*. Since this relation is very complex, it is modularized along several levels of representation, and correspondence rules between adjacent levels are defined. For written text, there are five levels: surface and deep morphological (*SMorphR* and *DMorphR*), surface and deep syntactic (*SSyntR* and *DSyntR*), and semantic (*SemR*). *SMorphR* and *SemR* are, respectively, symbolic representations of what we have called 'meaning' and 'text' above. Each level comprises several structures, one of which is predominant with respect to the other ones. For example, *SemR* consists of three structures: *SemS*, *Sem-CommS* and *RhetS*. *Sem-CommS* and *RhetS* are additional structures which specify, respectively, the semantic information the speaker puts in foreground rather than in background (theme vs. rheme) and the rhetorical style (colloquial, informal, poetic, etc.).

As far as *SemR* (the reference MTT level for this paper) is concerned, its main structure is *SemS*, i.e. a network representing predicate-argument relations. Every node of *SemS* is called a *semanteme* and corresponds to a

specific sense of a word. There are two classes of semantemes: *functors* (predicates, quantifiers and logical connectives) and *names of objects* (e.g. proper names). Functors have arguments; in the network, these are indicated by means of numbered arcs.

Instead, a *DSyntR* consists of four structures: *DSyntS*, *DSynt-CommS*, *DSynt-AnaphS* and *DSynt-ProsS*. *DSynt-CommS*, *DSynt-AnaphS* and *DSynt-ProsS* are additional structures which contain, respectively, rheme/theme, anaphoric and prosodic information. *DSyntS*, on the other hand, is the main structure of *DSyntR* and is a dependency tree which specifies the syntactic organization of the sentence in terms of dependencies between words. These dependencies appear in the tree as labelled arcs.

In this paper, we do not use the *DSyntS* label inventory originally defined in MTT (6 actant relations plus ATTR, COORD and APPEND), but the one defined in [Bosco et al. 00]. Moreover, we choose the determiner as governor for nominal groups (see [Hudson 90]). In this formalism, all NL quantifiers have an obligatory argument (complement) corresponding to their syntactic restriction. For space constraints, we cannot show in detail the correspondence rules $SemS \Leftrightarrow DSyntS$; We briefly state those we need as follows:

1. There is a one to one correspondence between the lexeme and the semanteme (except in case of pronouns and ellipsis, but we do not consider them here).
2. Verbs, common nouns and prepositions correspond to functors with a set of suitable arguments.
3. Determiners (like NL quantifiers) and proper names correspond to names.

For example, if we consider

(1) *two students study three theorems of four theories.*

The *DSyntS* and *SemS* of (1) are

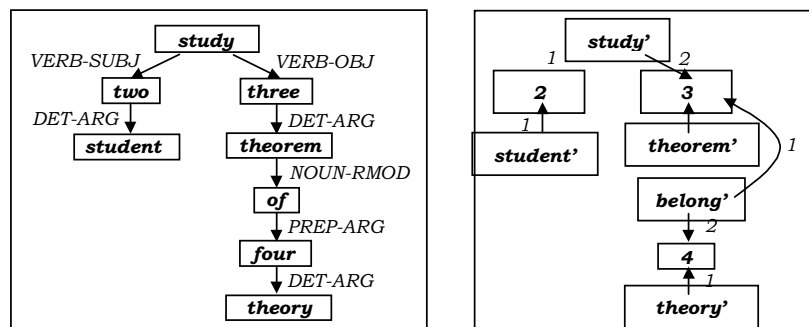


Fig. 1. *DSyntS* (on the left) and *SemS* (on the right) for
Two students study three theorems of four theories

2. Underspecification on quantifier scope

Underspecification techniques aim to devise a formalism such that all different readings are represented in a single structure. The underspecified representation should be ‘extended’ (and not reshuffled) by adding further information that expresses the various different readings. To this end, a number of underspecified representations have been proposed (main approaches are [Hobbs and Shieber 87], [Alshawhi 92], [Reyle 93], [Bos 95], [Musken 99], [Copestake et al. 99], [Koller 00] and [Egg et al. 01]. See [Bunt 03] for an overview). Recently, dependency structures have raised interest as a by-product of Lexicalized Tree Adjoining Grammars [Joshi and Schabes, 97]. In [Joshi and Kallmeyer, 03], it is shown how, in LTAG, an underspecified representation can be obtained from the derivation tree, a dependency structure representing the history of derivation.

In all frameworks, possible interpretations are derived starting from a single underspecified representation. This representation can be a partial formula with unsolved components or a flat set of labelled logical forms with particular free variables named holes. In any case, the subformulae are embedded in other ones so that, in a model-theoretic interpretation, the set of internally quantified entities vary on each externally quantified entity. For example, consider a simplified version of (1)

(2) *two students study three theorems.*

In 3a we show the underspecified representation of (2) in terms of Hobbs and Shieber’s algorithm [Hobbs&Shieber, 87], and, in 3b-c, the two possible disambiguations performed by the latter.

- (3) a. $\text{study}(\langle 2, x, \text{student}(x) \rangle, \langle 3, y, \text{theorem}(y) \rangle)$
 b. $2_x(\text{student}(x), 3_y(\text{theorem}(y), \text{study}(x, y)))$
 c. $3_y(\text{theorem}(y), 2_x(\text{student}(x), \text{study}(x, y)))$

In 3a, the terms $\langle q, v, p(v) \rangle$ are complex terms (there is one of them for each quantifier); they are “pulled out” and “unstored”, obeying to certain constraints, in order to obtain a wff (well-formed formula) in the form $q_v(p(v), s(v))$ in which $s(v)$ is another wff.

Depending on the order we choose for “expanding” the complex terms, we obtain the two interpretations 3.b and 3.c. The first is satisfied only by models in which the set of three theorems depends on the specific student (there are two students, each of which can study three different theorems) whereas the latter only by those in which the two students depend on a theorem. The reader can easily see that in most other approaches to the problem, we obtain the same results for sentence (2).

2. SemCtxDepS

In our approach, instead of building formulae with an embedded structure so that only models with certain dependencies between groups of involved entities will satisfy them, we explicitly show these dependencies in the representation. This is done by inserting dotted arcs, named *semCtxDep* arcs, between quantifiers. Moreover, the quantifier having wide scope is linked to a new node named *Ctx*. *Ctx* refers to the context, i.e. the domain of individuals with respect to which the semantic representation will be evaluated. It does not correspond to an actual word, but it has been inserted only for maintaining the homogeneity in the representation.

Ctx and *semCtxDep* arcs are part of a new additional structure of *SemS* named *SemCtxDepS*. Therefore, the two interpretations 3.a and 3.b correspond to 4.a and 4.b respectively.

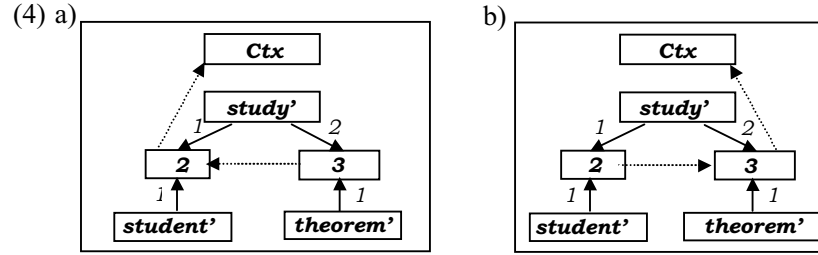


Fig. 2. SemS + SemCtxDepS of 2.a and 2.b

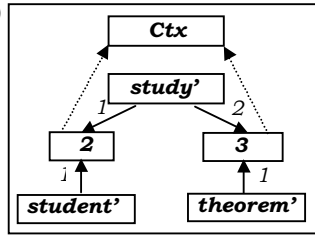
Clearly, we are implementing the same idea that lies behind Skolem's theory in First Order Logic (FOL). In other words, representations 3.a and 3.b correspond to FOL formulae 5.a and 5.b respectively, whereas our representations in 4.a and 4.b correspond to 5.c and 5.d, in which s_1 , s_2 , t_1 , t_2 , t_3 are Skolem constants and $f_1(x)$, $f_2(x)$, $f_3(x)$, $f_1'(y)$, $f_2'(y)$ are Skolem functions.

$$\begin{aligned}
 (5) \text{ a) } & \exists x_1 \exists x_2 [x_1 \neq x_2 \wedge \\
 & \forall x [(x = x_1 \vee x = x_2) \rightarrow \text{student}'(x)] \wedge \\
 & \forall x [(x = x_1 \vee x = x_2) \rightarrow \exists y_1 \exists y_2 \exists y_3 [y_1 \neq y_2 \neq y_3 \wedge \\
 & \quad \forall y [(y = y_1 \vee y = y_2 \vee y = y_3) \rightarrow \text{theorem}'(y)] \wedge \\
 & \quad \forall y [(y = y_1 \vee y = y_2 \vee y = y_3) \rightarrow \text{study}'(x, y)]]]] \\
 \text{ b) } & \exists y_1 \exists y_2 \exists y_3 [y_1 \neq y_2 \neq y_3 \wedge \\
 & \forall y [(y = y_1 \vee y = y_2 \vee y = y_3) \rightarrow \text{theorem}'(y)] \wedge \\
 & \forall y [(y = y_1 \vee y = y_2 \vee y = y_3) \rightarrow \exists x_1 \exists x_2 [x_1 \neq x_2 \wedge \\
 & \quad \forall x [(x = x_1 \vee x = x_2) \rightarrow \text{student}'(x)] \wedge \\
 & \quad \forall x [(x = x_1 \vee x = x_2) \rightarrow \text{study}'(x, y)]]]] \\
 \text{ c) } & s_1 \neq s_2 \wedge \\
 & \forall x [(x = s_1 \vee x = s_2) \rightarrow \text{student}'(x)] \wedge \\
 & \forall x [(x = s_1 \vee x = s_2) \rightarrow (f_1(x) \neq f_2(x) \neq f_3(x) \wedge \\
 & \quad \forall y [(y = f_1(x) \vee y = f_2(x) \vee y = f_3(x)) \rightarrow \text{theorem}'(y)] \wedge \\
 & \quad \forall y [(y = f_1(x) \vee y = f_2(x) \vee y = f_3(x)) \rightarrow \text{study}'(x, y)]) \\
 \text{ d) } & t_1 \neq t_2 \neq t_3 \wedge \\
 & \forall y [(y = t_1 \vee y = t_2 \vee y = t_3) \rightarrow \text{theorem}'(y)] \wedge \\
 & \forall y [(y = t_1 \vee y = t_2 \vee y = t_3) \rightarrow (f_1'(y) \neq f_2'(y) \wedge \\
 & \quad \forall x [(x = f_1'(y) \vee x = f_2'(y)) \rightarrow \text{student}'(x)] \wedge \\
 & \quad \forall x [(x = f_1'(y) \vee x = f_2'(y)) \rightarrow \text{study}'(x, y)])
 \end{aligned}$$

The two configurations in (4) suggest the existence of another reading for this sentence, represented in 6.a (in 6.b we show the corresponding logical formula).

(6)

a)



b) $t_1 \neq t_2 \neq t_3 \wedge s_1 \neq s_2 \wedge$

$$\begin{aligned} & \forall y[(y=t_1 \vee y=t_2 \vee y=t_3) \rightarrow \text{theorem}'(y)] \wedge \\ & \forall x[(x=s_1 \vee x=s_2) \rightarrow \text{student}'(x)] \wedge \\ & \forall x \forall y[(y=t_1 \vee y=t_2 \vee y=t_3) \wedge (x=s_1 \vee x=s_2)) \\ & \quad \rightarrow \text{study}'(x, y)] \end{aligned}$$

In this reading, we have only two students and only three theorems and each of the two students study the same set of three theorems. Note that this reading is the only acceptable one for the sentence

(7) *two students, **John and Jack**, study three theorems: **the first three of the book***

in which we have added (in boldface) some context elements in order to force 6.a to be the only acceptable one. Note also that in all cited underspecification approaches, as the one we briefly analyzed in 3, we cannot represent 6.a because they require that a quantifier occurs inside the scope of the other one whereas, for 6.a, we need to assign wide scope to both quantifiers.

3.1. Constraints on SemCtxDep arcs

Clearly, we need constraints for the insertion of semdep arcs (e.g. if two nodes are linked in both directions, each quantifier should occur in the scope of the other). Our strategy was to devise a set of reasonable semantic constraints and to check their predictions by evaluating the admissibility of the readings. On all the examples we considered, the predictions are confirmed by linguistic intuition. The constraints are:

1. There are no cycles.
2. The same pair of nodes cannot be linked via two different paths.
3. A node n_1 cannot be linked to a node n_2 if they do not share a predicate unless:
 - a. There is another node n_3 , which shares a predicate with n_1 and is linked to n_2 .

- b. There is another node n_3 which shares a predicate with n_2 and there is a path of links from n_3 to n_1 .

The first constraint is rather obvious. The second constraint specifies that there must be no redundancy. The third constraint states that a quantifier can depend on another one just in case there is a semantic relation along which the narrow-scope set can be perceived as “variable” with respect to the individuals belonging to the wide-scope set. So, in (1), we can only have such a dependency between 2 and 3 (the students can vary on the theorems or viceversa) or between 3 and 4 (the theorems can vary on the theories or viceversa), but not between 2 and 4 (neither the students can vary on the theories nor the theories can vary on the students).

Nevertheless, the ‘unless’ clause in 3 partially relaxes this constraint. 3.a says that the quantifier associated with n_1 can occur in the scope of the quantifier associated with n_2 if another quantifier (associated with n_3) is in that scope and shares a predicate with n_1 (e.g., in fig. 4, on the left, 3 and 4 have the same scope level and are both within the scope of 2). 3.b says that the quantifier associated with n_1 can occur in the scope of the quantifier associated with n_2 in case there is another quantifier n_3 which shares a predicate with n_2 and which occurs in the scope of n_1 (e.g., in fig. 4, on the right, 4 could not be linked to 2, but since 3 shares a predicate (*study*) with it, 3 can ‘carry with itself’ 4 within the scope of 2. So, the two students chose the theories to which the theorems they will study belong (possibly three different theorems for each theory). With these constraints, sentence (1) has thirteen readings, five of which are reported in Fig. 3 and 4.

Note that a quantifier can depend on more than one quantifier, as in fig. 3, on the right: we have a set of three theorems for each pair (student, theory). We will see another example of this configuration below.

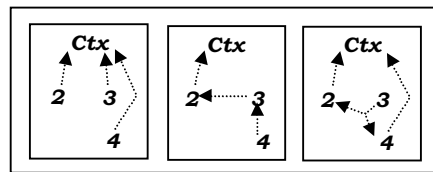


Fig. 3. three representations
licensed by clauses 1-2-3

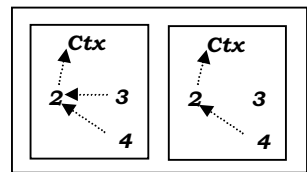


Fig. 4. two representations
licensed by clause 3.a and 3.b

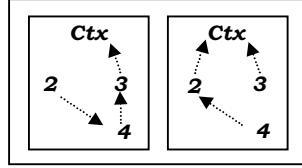


Fig. 5. Two representations excluded by clause 3.b

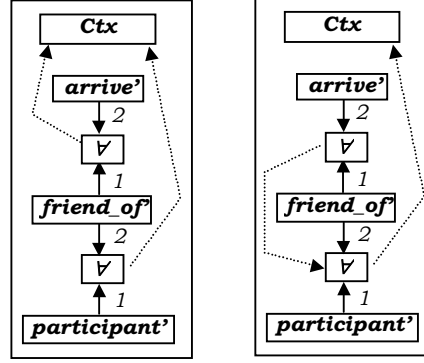


Fig. 6. SemS and the two *SemCtxDepS* of *Every friend of every participant arrives.*

3.2. Universal and Existential quantifiers

Existential and universal quantifiers are treated as special cases of the other ones. Universals, since they range on the whole domain of individuals, cannot exhibit any dependency on another quantifier, i.e. their inclusion within the scope of another quantifier does not affect the selection of individuals they range on, so that the *SemCtxDep* arc exiting them can be assumed to be fixed and to enter *Ctx*. For analogous reasons, no arc can enter an existential quantifier. Anyway, there is an exception to the rule about universals. It concerns an universal acting as a modifier for another quantifier in its restriction. For example, in the case of *every friend of every participant arrives* (fig. 6), there are two possibilities, according to the intention of referring to ‘all friends of *any* participant’ or to ‘all friends of *all* participants’¹. In the first case, we enable the upper universal to be linked via a *SemCtxDep* arc to its modifier as in fig. 6, on the right. Clearly, the standard link to *Ctx* is possible, so also the second case is accounted for (fig. 6, on the left).

4. Comparison with other approaches

We can now test the prediction of the DTS on a well known example

(8) *Two politicians spy on someone from every city*

¹ Of course, the same situation arises in case of other modifiers, e.g. *all friends of two participants*.

Just two *semdep* arcs can vary: the one exiting *two*, and the one exiting *some* ($\exists$). The rules presented above allow for 6 interpretations, reported in fig. 7.

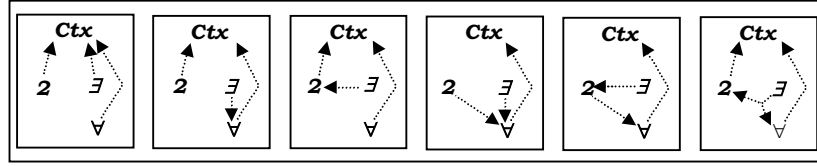


Fig. 7. The six disambiguated readings
for *Two politicians spy on someone from every city*

In [Joshi et al. 03], underspecification is dealt with via a formalism similar to MRS [Copestake et al. 99] that makes use of *quantifier sets*. It accepts just 4 readings for (8), expressed as

$$\exists\forall\forall, \exists\forall\exists, \forall\exists\forall, \forall\forall\exists$$

while $\exists\forall\forall$ and $\forall\forall\exists$ are unacceptable (see also [Larson 85]).

Nevertheless, consider the following sentences of which, because of semantic preferences, the preferred reading may correspond to the one (with the same order) of the previous figure.

- Two boys, John and Jack, went (together) to a show about all rockstars.*
- Two boys, John and Jack, have visited (together) a museum in every region of Italy.*
- Two boys, John and Jack, bought a dress of all colors.*
- An answer to every riddle has been guessed by two different persons.*
- A ticket for any show is bought, at least, by two people.*
- Two boys, John and Jack, ate a portion of every course of the menu.*

In [Joshi et al. 03], (a) is not a valid interpretation because «it may be banned due to general architectural reasons» (p.3). The logical representation they propose for (a) is ([Joshi et al. 03], footnote 2):

$$\begin{aligned} &\exists_y[\text{person}'(y) \wedge \\ &\quad 2_x[\text{politicians}'(x) \wedge \\ &\quad \quad \forall_z[\text{city}'(z) \wedge \text{from}'(y, z) \rightarrow \text{spy}'(x, y)]]] \end{aligned}$$

And the comment is that « $\exists\forall\forall$ involves having the quantifier $\exists_y$ separated from its restrictor *from*(*y*,*z*), a configuration that should probably be banned...». On the contrary, our translation is :

$$\begin{aligned}
& p_1 \neq p_2 \wedge \\
& \forall_x [(x = p_1 \wedge x = p_2) \rightarrow \text{politician}'(x)] \wedge \\
& \forall_y \forall_z [(y = pe_1) \wedge (\text{city}'(z)) \rightarrow (\text{person}'(y) \wedge \text{from}'(y, z))] \wedge \\
& \forall_x \forall_y [(x = p_1 \wedge x = p_2) \wedge (y = pe_1) \rightarrow \text{spy}'(x, y)]
\end{aligned}$$

And here, the separation mentioned above does not arise. Reading (e) is more complex, but it seems possible to express that situation via (8), although [Larson 85] sees it as unacceptable. Anyway, we follow [Pereira 89]: “*expressing restrictions on possible scopings in terms of formal constraints on the logical form goes against the spirit of compositionality [...] they can be expressed as constraints on the derivation of possible meanings for sentences rather than constraints of the alleged forms of those meanings*”. Our constraints have linguistic/semantic motivations, rather than formal ones.

5. Conclusions

In this paper, we have presented how it is possible to modify *SemR* in MTT in order to deal with underspecification of quantifier scope ambiguity. In particular, we have moved all its representational aspects into a new additional structure named *SemCtxDepS*: in *SemS*, determiners (in particular, quantifiers) appear as if they were proper names. In our opinion, this simplifies the correspondence rules $SemR \Leftrightarrow DSyntR$. Moreover, we think that the disambiguation operators acting to specify a particular scope order (i.e. *SemCtxDep* arcs) are more motivated, linguistically and semantically, than methods used in most current approaches to underspecification, mainly based on the subformula encapsulation mechanism, which, in some cases, can prevent some acceptable readings.

References

- [Alshawhi 92] H. Alshawhi, ed., “The Core Language Engine”. Mit Press, 1992.
- [Bos 95] J. Bos: “Predicate Logic Unplugged”, Proc. 10th Amsterdam Colloquium, 1995. P. 133—142.
- [Bosco et al. 00] C. Bosco, V. Lombardo, D. Vassallo, L. Lesmo: “Building a Treebank for Italian: a data-driven annotation schema”, in Proc. 2th Int. Conf. on Language Resources and Evaluation (LREC 2000). P. 99—105, Athens, 2000.
- [Bunt 03] H. Bunt: “Underspecification in Semantic Representations: Which Technique for What Purpose?”. Proc. 5th Workshop on Computational Semantics (IWCS-5), Tilburg, 2003. P. 37—54.

- [Copestake et al. 99] A. Copestake, D. Flickinger, I. A. Sag: "Minimal Recursion Semantics. An introduction". Manuscript, Stanford University, 1999.
- [Egg et al. 01] M. Egg, A. Koller, J. Niehren. "The Constraint Language for Lambda Structures". *Logic, Language, and Information*, 10. 2001. P. 457—485.
- [Joshi and Schabes 97] A. K. Joshi, Y. Schabes: Tree-Adjoining Grammars. In G. Rozenberg, A. Salomaa (eds.): *Handbook of Formal Languages*. Springer Verlag, Berlin, 1997. P. 69—123.
- [Joshi et al., 03] A. K. Joshi, L. Kallmeyer and M. Romero (2003) "Flexible composition in LTAG: quantifier scope and inverse linking". In Kluwer, editor, *Computing Meaning*. Vol. 3. R. Musken and H. Bunt (eds). 2003.
- [Kallmeyer and Joshi 03] L. Kallmeyer, A. K. Joshi: "Factoring Predicate Argument and Scope Semantics: Underspecified Semantics with LTAG". *Research on Language & Computation* 2003. P. 3—58.
- [Koller et al. 00] A. Koller, K. Mehlhorn and J. Niehren. "A polynomial-time fragment of dominance constraints. In *Proceedings of the 38th Annual Meeting of the Association of Computational Linguistics*, Hong Kong, 2000. P. 368—375.
- [Hobbs and Shieber 87] J. R. Hobbs, S.M. Shieber: "An Algorithm for Generating Quantifier Scoping", *Computational Linguistics* 13, 1987. P. 47—63.
- [Hudson 90] R. Hudson: *English Word Grammar*. Basil Blackwell, Oxford and Cambridge, MA, 1990.
- [Larson 85] R. K. Larson: "Quantifying into NP", Ms. MIT, 1985.
- [Mel'čuk 88] I. Mel'čuk: *Dependency syntax: theory and practice*, SUNY University Press, 1988.
- [Mel'čuk 00] I. Mel'čuk: *Semantics and the Lexicon in modern Linguistics*. In: A. Gelbukh (Ed.), *CICling-2000*, 1st International Conference on Intelligent Text Processing and Computational Linguistics. 2000. P. 6—18; www.CICLing.com.
- [Mel'čuk and Žolkovskij 67], "O semantičeskom sinteze" [On semantic synthesis], *Problemy kibernetiki* [Problems of Cybernetics]. 19. 1967. P. 177—238. [Fr. transl.: 1970, T. A. Information. 2. P. 1—85].
- [Muskens 99] Muskens, R. "Talking about Trees and Truth-conditions". In *JFAK, Essays Dedicated to Johan van Benthem*. 1999. Vossiuspers/Amsterdam University Press.
- [Pereira 89] F. Pereira. "A calculus for semantic composition and scoping". In *27th Annual Meeting of the Association for Computational Linguistics*. P. 152—160. University of British Columbia, Vancouver, Canada, April, 1989.
- [Reyle 93] U. Reyle: "Dealing with ambiguities by Underspecification: Construction, Representation and Deduction", *Journal of Semantics*. 10 (2). 1993.

The Particles One Cannot Do Without

I. B. Levontina

Institute of the Russian Language,
Russian Academy of Science
irina.levontina@mail.ru

A. D. Shmelev

Moscow State Pedagogical University
shme.lev@ru.net

It is sometimes the case that in a given communicative situation, a certain particle is pragmatically obligatory even though the speaker does not seek to express the associated meaning. In the absence of the relevant particle, the wrong implicatures may appear, or the text may cease to have coherence, or the utterance is perceived as a manifestation of arch use of language. The lack of the particle where the communicative situation calls for it every so often gives the impression that we deal with a word-for-word translation of an original text containing no similar marker.

If a particle is pragmatically obligatory in a certain type of contexts in a given language, and a context of this type has to be translated from a language with no such marker into the language in question, the translator may be confronted with a problem. If s/he translates the text literally (that is, with no particle), s/he will convey the precise meaning, but the translation will be inadequate (the wrong implicature will appear). However, if s/he adds the needed particle, the translation will be more adequate even though a new semantic component is added.

In the paper, we put emphasis on the use of the particle *еще* in combination with a verb in future tense; we have established a type of contexts where *еще* is pragmatically obligatory.

Some time ago some Russian television program presented the results of some sociological research and it turned out that the most frequently cited expression is *Я вернусь* (literally, 'I'll return!'). The sentence sounded weird and funny because no one could immediately recognize the source. Eventually though people guessed that it was Terminator's expression *I'll be back!* (usually quoted in English, even by people who do not know English). The Russian translation carries neither the right connotations nor relevant associations. But this is not the only trouble. However strange it may sound, the expression *Я вернусь* and *I'll be back*

are not exact, idiomatic equivalents. The Russian translation should say *Я еще вернусь*, with a particle being added, which is how the sentence is usually rendered into Russian. Here emerges the question as to why the Russian text requires a particle in this case.

The concept of pragmatic obligatory use

There are other cases in Russian when a particle is obligatory from the pragmatic point of view. Pragmatic obligatory usage implies that the lack of a proper particle may lead to wrong implicatures, or disrupt the coherence of the text, or else the utterance may acquire stylistic coloring and demonstrate equivocation.

Thus, the repeated request in Russian requires a special marker, such particles as *ну, да, же* [Левонтина, 1999]. For instance, if a request to turn the television off is ignored, it will seem strange if a person just repeats *Выключи, пожалуйста, телевизор* 'Please, turn off the television'. Instead one would rather say *Да выключи же телевизор*; *Ну пожалуйста, выключи телевизор* ('Do turn off the television' or 'Turn the television off, will you?'). If the request is repeated, but contains no markers of a repeated action, the coherence of the dialogue will be lost since the person formulates the request as if for the first time. S/he may deliberately ignore the fact that her/his request was not even considered.

Another example of pragmatic obligatory particles is the particle *вдруг* as a part of conditional construction (its meaning may be roughly conveyed by the English 'what if'). By using the construction *если вдруг*, the speaker implies, "I know that there are no reasons for this to happen, but because anything can happen, I would still like to consider this option." The use of the word *вдруг* in such a context is pragmatically obligatory, its absence in the conditional clause creates a conversational implicature 'the speaker considers the realization of this condition quite probable'. In other words, the utterance starting with *Если Петя позвонит, скажи ему...* 'If Peter calls, tell him...' implies that the speaker views Peter's call as likely to occur (otherwise it should have sounded *Если вдруг Петя позвонит...*). The interlocutor who does not agree with this implicature may argue using the word *вдруг* to object (e.g. *Если Петя позвонит, скажи ему... — С чего это он вдруг позвонит?* 'Why would he call?') or express the same hypothesis in another way, by adding the lacking *вдруг* (*Ну хорошо, если он вдруг позвонит, я скажу ему...* 'Well, if he should call, I'll tell him'). [Шмелев, 2001]

*Еще in combination with future tense**Problem definition*

Going back to the Terminator case, it is worth noting that, on the other hand, the expression *Я вернусь* ‘I’ll return’ does not fail to be both natural and idiomatic. If we take the poem written by Simonov “Жди меня” (“Wait for me”) with its famous opening line *Жди меня, и я вернусь* ‘Wait for me, and I’ll come back’, the use of *еще* (*Жди меня, и я еще вернусь*) would be highly inappropriate. The same is the case with a famous Russian song “Я сегодня до зари встану” (“I’ll get up before the dawn today”, lyrics by R. Rozhdenstvensky) where the sentence *Даже не был я знаком с парнем, / Обещавшим: «Я вернусь, мама»* ‘I wasn’t even acquainted with the guy / Who promised, “I’ll come home, mother”’ cannot be modified with *еще*.

The conclusion that follows is that the Russian language is sensitive to some context parameter, which specifies how the speaker refers to some future event. It will be useful to answer such questions as:

1. What are the characteristic features of the context that determine the usage of *еще*?
2. When is *еще* optional?
3. When *еще* cannot be used?
4. What is the meaning alteration due to the use of *еще* in the utterances related to future events?
5. What are the reasons for wrong usage or pragmatic ambiguity of utterances which lack obligatory or almost obligatory *еще*?

The distinctions to be made

Let us note, first of all, that the particle *еще* is a polysemous one. Some meanings are analyzed in [Boguslavsky 1996: 258—291]. In particular, the continual meaning of *еще* (*Он еще жив* ‘He is still alive’) and the additive *еще* with its two variants (stressed and unstressed) are considered. The stressed additive *еще* can be illustrated by *В доме есть ↓еще раненые* ‘There are still some more wounded in the house’ and the unstressed one by *В доме есть еще ↓раненые* ‘There are also some wounded in the house’.¹ *Еще* that is used in the sentences like *Я еще вернусь* ‘I’ll be back’ is neither additive nor continual. It does not imply

¹ For differences between the stressed and unstressed *еще* see also [Апресян, 1990: 102].

that the situation taking place at the moment of speech will continue as it happens with continual *еще*. Neither it presupposes that the situation in question (or some similar situation) has already taken place and will be reproduced; in this respect, it is opposed to additive *еще*.

One should distinguish the meaning of *еще* under analysis from other meanings of *еще* in future contexts, and in particular from the following ones.

First, it is different from the additive stressed *еще* in the meaning of 'once again' or 'some more'. Consider: *Я приеду к вам еще!* 'I'll visit you once again!'; *Мы передохнем и попробуем еще* 'We'll have some rest and try once again'; — *Да ешьте, ешьте, — настаивал я. — Мне сейчас еще принесут.* 'Eat, please do — I insisted. — I shall get some more soon' (G. Baklanov). *Еще* here functions as an adverb rather than a particle. It can be seen that *еще* in the context of referring to future events does not differ much from *еще* in the context referring to past events. Consider: *Они приехали еще* 'They came once again'; *Мы передохнули и попробовали еще* 'We had some rest and tried again'.

Second, it differs from additive unstressed *еще*; consider: *Будете еще что-нибудь заказывать? — Да, я еще съем десерт* 'Would you like anything else? — Yes, I'll have dessert'; *Книгу я пока не нашел, но я посмотрю еще в Ленинской библиотеке* 'I haven't found the book yet, but I am going to look it up in Lenin library as well'. Consider: *Я еще съел десерт* 'In addition, I ate dessert'; *Книгу я еще посмотрел в Ленинской библиотеке* 'I tried to find the book in Lenin library as well'.

Third, there are contexts where *еще* means 'for some time' (the continual *еще*). Consider: *Я еще здесь побуду* 'I will stay here some more time'; *Главком ВВС Владимир Михайлов заявил, что самолет Ан-70 еще полетает* 'Commander-in-Chief Vladimir Mikhailov declared that the aircraft AN-70 will fly for a while longer' (Izvestia, 2003.01.15); *Начальник столичного метеобюро Алексей Ляхов сказал «Известиям», что торфяники скорее всего еще будут гореть* 'The head of the city weather bureau told Izvestia that peatlands would still be on fire for some time' (Izvestia, 2002.08.16). Consider: *Он еще там побыл* 'He spent some more time there'; *Самолет еще полетал* 'The aircraft flew for some more time'; *Торфяные пожары еще продолжались* 'The peat-fire still continued'.

Finally, a totally different type of contexts, which we will not consider in detail here, is made up by the particle *еще* in combination with the verbs *успеть* 'to have time, to manage to make it in time', *смочь* 'to be able', *удаться* 'to succeed', *хватить* 'to be sufficient for, to last' (all glosses are of course approximate) and the like: *Сейчас 5 часов, я еще*

успею на поезд 'It's 5 o'clock now, I'll still manage to catch the train';
Имеющегося запаса прочности теплосетей еще хватит на какое-то время 'Low safety margins of heating network will be enough for some time' (*Izvestia*, 2003.01.15). Consider: *Он еще успел на поезд* 'He still managed to catch the train'; *Запаса еще хватило на какое-то время* 'Safety margins were enough for some time'.

Thus, the enumerated types of future tense usage of *еще* manifest its different meanings, which can be also realized in other tenses. The meaning of *еще* under consideration is realized only or almost exclusively in the context of future.

Not so fast!

Before proceeding to the future meaning explication of *еще*, let's have a look at some typical examples:

Но можно не сомневаться, что Цымбаларь еще удивит всех нас.
 'There are no doubts that Tsymbalar may surprise us yet.' (*Izvestia*, 2001.08.07)

Даже издалека Швейк совершенно отчетливо слышал пронзительный голос подпоручика Дуба: — Вы меня знаете?! А я вам говорю, что вы меня не знаете!.. Но вы меня еще узнаете!.. Может, вы меня знаете только с хорошей стороны!.. А я говорю, вы узнаете меня и с плохой стороны!.. Я вас до слез доведу! Ослы! 'Even from afar, Schweik distinctly heard officer Doub shrieking and screaming: "You think you know me?! And I tell you — you know nothing about me! But you will see for yourself! You may have seen me at my best! I tell you — you will see at my worst! I'll make you cry! Assheads!"' (J. Hašek, *The Good Soldier Schweik*, Russian translation by P. Bogatyrev)

*Мы еще оденем вас шелками,
 плечи вам согреем соболями.
 Мы построим вам дворцы большие,
 милые красавицы России.*

'We will garment you in silk some day'

'We'll cover your shoulders with sable'

'We'll build great palaces for you'

'Sweet beauties of Russia'

(Yaroslav Smelyakov)

*Но мы еще дойдем до Ганга,
 Но мы еще умрем в боях,
 Чтоб от Японии до Англии
 Сияла Родина моя.*

‘But we will go as far as the Ganges some day’
 ‘And we will die in battle some day’
 ‘So that from Japan to England’
 ‘My native land will be glorified’
 (Pavel Kogan)

Уходя в очередной раз в сторону от очередного комплимента, как боксер от удара, он сказал: «Погоди, я еще напишу что-нибудь стоящее, вот увидишь». ‘Dodging, as if a blow in boxing, yet another compliment, he said, “Wait a minute! I’ll write something noteworthy, wait and see!” (A. Bitov)

The examples show the following points. First, they are polemical: *еще* introduces an implication that there is a different, contrary hypothesis on how the events will unfold. Therefore it is not by chance that the sentence containing *еще* begins with *но* ‘but’. Consider: *Но мы еще дойдем до Ганга* ‘But we will go as far as the Ganges some day’; *Но вы меня еще узнаете* ‘But you will see for yourself!’. The typical stress on the verb in utterances with *еще* is also polemically conditioned. Consider: *Мы дойдем до Ганга* ‘We will go as far as the Ganges’; *Мы еще дойдем до Ганга* ‘But we will go as far as the Ganges some day’.

Secondly, *еще* implies that the event described is not going to happen right after the moment of speech. The sentence *Мы еще оденем вас шелками* ‘We will garment you in silk some day’ does not indicate that this action is to be undertaken in the near future. When a person is told *Мы еще повоюем* ‘We will still fight’, s/he understands that now the things look black. This meaning is so evident that, for instance, the sentence *Мы еще поговорим об этом* ‘We’ll talk about this later’ is used to change the subject. Consider: *Мы еще поговорим о результатах этой амнистии..., я хотел бы все-таки [вернуться] к выборам* “‘We will discuss the results of the amnesty in detail later..., I’d like to revert to elections” (Venedictov, The Echo of Moscow, 2003.07.25)

Finally, thirdly, *еще* means that at the moment of speech the situation is developing in such a way that it is impossible to make any predictions and forecasts. Consider *Заканчивалась поэма в том смысле, что, дескать, вы победили, но мы еще отомстим* ‘The poem ended on the note that you have won, but we’ll revenge some day’ (Khodasevich).

Thus, with the help of the particle, the speaker admits that the addressee has some grounds to think that it is going to happen “no longer” or “no more”, but warns that the latter should not make a final conclusion yet. In other words, *еще* means, “it is not true that the situation is irreversible.” In addition, *еще* refers to a relevant period of time, starting from the moment of speech, and indicates that this change will occur at some indefinite moment within this period, but not right after the moment of speech.

The meaning of ‘it is too early to make any conclusions yet’ determines two pragmatic contexts for *еще* in the meaning considered: ‘a day before the fair’ and ‘everything has not been lost, don’t get upset yet’. 1) ‘A day before the fair’: *Ты у меня еще попляшешь!* ‘You will live to regret it!’; *Ты еще пожалеешь* ‘I’ll make you squirm yet!’ and the like; 2) ‘Everything has not been lost, don’t get upset yet’: *Ты еще всех нас переживешь* ‘You will outlive us all!’; *Все еще будет хорошо, все еще наладится* ‘Don’t worry, it is sure to turn out well!’. Sometimes it may be unclear if it is a threat or consolation or these both shades of meaning are realized simultaneously. Consider: *Это мы еще посмотрим!* ‘Never is a long day!’; *Не огорчайтесь, — услышал Малиновский и понял, что разговаривает с блондинкой в голубом халате. — Они ещё пожалеют.* “Don’t get upset,” Malinovsky heard and realized that he was talking with a blonde in a blue dressing gown. “They will live to regret it!” (Dovlatov)

On the basis of the explication suggested, we may identify the contexts when *еще* is pragmatically obligatory. In particular, *еще* is obligatory or almost obligatory if a person has to forgo something (if s/he is not able to fulfill some plan) and comforts her/himself that this is not final and in some indefinite future the situation may change.

For example, if someone’s plans misfired, you may say, *Да ты еще 10 раз съездишь на Кунг* ‘You will go there a dozen times!’ When comforting a person who had some disappointment in love, they say, *Ты еще встретишь настоящую любовь* ‘You will still meet someone!’; *Да она еще сама к тебе прибежит* ‘She will be the one to run after you!’ Without *еще* the sentences would sound strange.

On the other hand, there are contexts which make the use of *еще* impossible. For example, *еще* cannot be used if the utterance specifies the terms under which the situation will take place, as it contradicts the added semantic meaning of *еще* ‘at this moment nothing signifies that the situation will actually occur’. Consider: *Стучите, и откроют вам* ‘Knock, and it shall be opened unto you’; *Позвоните, и я приеду* ‘Call and I’ll come’; *Если что, я помогу* ‘I’ll help if needed’. This is the reason why

еще is incompatible with the context of Simonov's "Wait for me"². The reason is that returning home in this case is determined by expectation and thus *еще* cannot be used. It is interesting that if you take away the conjunction *и* 'and' which denotes cause-effect connection, *еще* will become not only possible but even preferable. Consider *Жди меня: я еще вернусь* 'Wait for me since I'll come back'.

Еще cannot be used if a person is about to perform a certain action, Consider: *В глаз дам!* 'I'll plug in the eye!'; *Стой, стрелять буду!* 'Don't move or I'll shoot!' On the whole, *еще* in this meaning doesn't go together with the exact timing. Consider a strange sentence *Я еще ↓ поеду на море через неделю* 'I'll still go to the seaside in a week' and the natural *Я еще съезжу на море, какие мои годы* 'I'll still have an opportunity to go the seaside, I am young enough'.

Let's return to Terminator's formula and its translation.

So what is the difference between *Я вернусь* 'I'll return' and *Я еще вернусь* 'I'll come back'?

The utterance with *еще* is a deferred forecast. At the moment there are signs that the situation *p* is not going to happen. It is forecast that in future these signs will remain unchanged for some time or even intensify until the situation *P* occurs. In other words, a person goes away and will be absent for some indefinite period of time but will eventually come back.

Simple expression of future tense without *еще* means that either the action will take place now or the period when it will happen is not even considered. The absence of the particle *еще* with a verb in future tense may also imply that either there are already some grounds for forecast or one may indicate the time and conditions of *p*. It may be irrational faith as in the poem of *Rozhdestvensky* when the soldier promises his mother: *Я вернусь, мама* 'I'll come back, mother'. Such sentences are often used as studied reassurance (as in the cheerful soldier's song *Солдат вернется, ты только жди* 'The soldier will come back, only wait for him'). With insufficient context, it is unclear which implicature is meant because the sentence without *еще* may turn out to be pragmatically inadequate, as it happens with the sentence from Terminator.

² Consider also:

Ничего, ничего, ничего!
Сабля, пуля, штыки — все одно.
Ты, родимая, ты дожидись меня,
И я приду.

'Nothing, nothing at all, nothing much!'
'Sables, bullets, bayonets — all the same.'
'You, my own, wait till I come,'
'And I'll come.'

(Y. Kim)

The absence of *еще* in those contexts when the communicative situation carries markers of its presence often gives the impression that we deal with a word-for-word translation of an original text containing no similar marker. Sometimes this is the result of non-cooperativity. Consider repeated forecasts *Он тебя обманет* 'He'll betray you' in response to proofs of serious intentions of the person in question. With *еще* the sentence will be more adequate: yes, now everything is fine, nothing augurs ill, but some day the situation will change.

Pragmatically obligatory particles in translated texts

Here we encounter a paradox. If a particle is pragmatically obligatory in a certain type of contexts in a given language and the translator renders it in a language with no such markers, s/he finds her/himself at the cross-roads. If s/he gives word-for-word translation, that is, without a particle, s/he will convey the precise meaning but the translation will be inadequate (the wrong implicature will appear). On the other hand, if s/he adds the needed particle and thus imparts new shades of meaning which the original text doesn't contain, s/he will sacrifice accuracy for adequate translation and equivalence.

It can be best illustrated when analyzing another typical Russian particle (namely, *же*) in the so-called emphatic meaning. It is usually pragmatically obligatory as its removal violates the idiomatic nature of the utterance and changes its illocutionary force. For example, the sentence *Ну что же ты!* 'Bother!' or 'There you go!' is perceived as reproach, while *Ну что ты!* 'Indeed!' can express disagreement, or serve to hurry someone up or perhaps be used in some other meaning. The question *Как так?* 'How come? Say, how is that?' is taken only as a sign of perplexity while the expression *Как же так?* 'Oh, come now!; How could you...?' has more shades of meaning and different interpretations.

The particle *же* is often used in Russian translations of original English texts. In most cases it is an added element, with no equivalent in the English text as in the following examples from *Winnie-the-Pooh* by Alan A. Milne, narrated by Boris Zakhoder:

А Пух все думал, как же этот дедушка выглядит. (*in the original text*: ...and Pooh was wondering what a Grandfather was like.)

— Да это же Кристофер Робин! — сказал он. (*in the original text*: He saw a friend of his. "It's Christopher Robin," he said.)

— А теперь подумаем, — сказал он, в последний раз облизав доннышко горшка, — подумаем, куда же это я собирался идти.

(*in the original text*: “Now let me see,” he thought, as he took his last lick of the inside of the jar, “where I was going?”)

One may notice that for example the English sentence *And what about me?* is most naturally rendered into Russian as *А как же я?*

There are cases, however, when the Russian sentences with *же* are translations of English sentences with some corresponding markers. When analyzing *Winnie-the-Pooh* in particular we can see a funny correlation: in Zakhoder’s translation *же* often appears when the words are singled out by italics. Consider:

— Так ведь это же я! — сказал он. (*in the original text*: “But this is Me!” said Bear)

— А как же обедать?! — Обедать, дорогой мой, не придется! — сказал Кристофер Робин. (*in the original text*: “What about meals” “I am afraid no meals,” said Christopher Robin.)

Иногда он грустно думал: «Почему?», а иногда: «По какой причине?», а иногда он думал даже так: «Какой же отсюда следует вывод?» И не удивительно, что порой он вообще переставал понимать, о чем же он, собственно, думает. (*in the original text*: Sometimes he thought sadly to himself, “Why” and sometimes he thought, “Wherefore?” and sometimes he thought, “Inasmuch as which?” — and sometimes he didn’t quite know what he was thinking about.)

— Что мне делать? Я же должен подарить ему что-нибудь! Ай-ай-ай-ай-ай! (*in the original text*: “Bother!” said Pooh. “What shall I do? I must give him something.”)

Thus the investigation of contexts where particles are pragmatically obligatory is of paramount importance in understanding their functioning.

References

- Апресян Ю. Д. Типы лексикографической информации об означаемом лексеме // Типология и грамматика. М., 1990.
- Богуславский И. М. Сфера действия лексических единиц. М., 1996.
- Левонтина И. Б. Стратегии уговаривания: частицы в повторных просьбах // Язык. Культура. Гуманитарное знание: Научное наследие Г. О. Винокура и современность. М., 1999.
- Шмелев А. Д. Некоторые тенденции семантического развития русских дискурсивных слов (*на всякий случай, если что, вдруг*) // Русский язык: пересекая границы. Дубна, 2001.

Русские прилагательные-определения: валентностный анализ

Леннарт Лёнгрен

Университет г. Тромсё, Норвегия
lennart.lonnngren@hum.uit.no

В основу предлагаемого сообщения положен некомпонентный анализ лексического значения. С учетом позиции предиката, объединяющего две части словосочетания, различается четыре типа валентностных структур. Предикат может быть выражен 1) прилагательным: *новый дом*; 2) существительным: *головная боль*; 3) сегментом, инкорпорированным в прилагательное: *⟨внутри⟩партийные конфликты*; 4) сегментом, инкорпорированным в существительное: *Божьи ⟨созда⟩ния*.

Valency analysis of adjectives in attributive position. This analysis is based on a non-componential approach to lexical meaning. With respect to the position of the predicate uniting the two parts of the syntagm four types are identified. The predicate may be expressed by 1) the adjective: *novyj dom* '(a) new house'; 2) the noun: *golovnaja bol'* 'headache'; 3) a segment incorporated in the adjective: *⟨vnutri⟩partijnye konflikty* 'conflicts within the party'; 4) a segment incorporated in the noun: *Bozh'i ⟨sozda⟩nija* 'creations of God'.

В основе принятого здесь подхода лежит отказ от компонентного анализа значений языковых знаков (ср. Murphy 2003, с. 15). Признание всех знаков атомарными влечет за собой и отказ от толкования как ключа к значению знака. Вместо толкования можно пользоваться «внутренним переводом», т. е. опираться на парафразу. Удача анализа подтверждается изоморфизмом между соответствующими семантическими графами.

Приведем несколько примеров. В слово *читатель* не входит семантический компонент 'читать'. *Читатель* и *читать* (здесь в инкорпорированной форме: *⟨чита⟩*) — это две лексемы одного уровня, причем каждая из них занимает свой узел в соответствующем графе. Другая ситуация, конечно, в случаях т. н. синтаксической деривации: *читать* / *чтение*.

В принципе так же обстоит дело и с производными прилагательными. В словосочетании *каменный дом* прилагательное имеет то же значение, что *построенный* в парафразе *дом, построенный из камня*, а *дом* и *камень* представляют собой актанты данного предиката. *Лесной* в словосочетании *лесная тропинка* синонимично предлогу *в*; ср. *тропинка в лесу*. Но в словосочетании *лесной пожар* прилагательное — синтаксический дериват; ср. *Лес горит*.

При таком подходе вряд ли можно безоговорочно согласиться с мнением И. А. Мельчука (1974, с. 84), что *книжный* является синтаксическим дериватом. В примерах типа *книжный шкаф*, *книжный язык* это не так. Только по отношению к такому необычному выражению, как *книжная страсть его миновала* (В. Широков), перед нами, действительно, синтаксический дериват; ср. *страсть к книгам*.

Перейдем к анализу прилагательных — и не только производных.

В словосочетаниях типа *высокое дерево*, *новый дом* синтаксическая зависимость идет от существительного к прилагательному (1). Семантическая же зависимость идет в противоположном направлении: от прилагательного-предиката к существительному-актанту (2). Есть, однако, случаи, когда семантическая направленность совпадает с синтаксической (3). Такое словосочетание производное; ср. (4).

1. новый дом 2. новый дом 3. головная боль 4. Голова болит.
- 

Значит, в первом случае предикатом является прилагательное (*новый*), во втором — существительное (*боль*). Есть ли другие возможности?

Да, предикатом (минимум двухместным) может служить сегмент, инкорпорированный либо в прилагательное (5), либо в существительное (6):

5. <внутри>партийные конфликты 6. Божьи <созда>ния
- 

К последнему типу относим также случаи, когда нужный предикат имплицитный, но извлекается из значения именно существительного, например: *командный «в» пункт* (см. об этом ниже).

Вне этой классификации остаются случаи, когда в составе словосочетания нет объединяющего предиката. Например, в словосочетании *красный цвет* существительное является гиперонимом прилагательного. Возникают парафразы типа

7. платок красного (цвета)

8. красный платок.

Слово *цвет* дается в скобках, т. к. оно не занимает своего узла в графе. Такую «неузловую» лексему назовем синтаксическим.

Теперь рассмотрим каждый из четырех типов отдельно.

Первый тип (*новый дом*).

Помимо непроизводных словосочетаний, типа *новый дом*, есть немало производных. Правда, направленность деривации словосочетаний *быстрый полет* и *быстро лететь* неопределенна, но несомненно, что как *быстро*, так и *полет* подвержены синтаксической деривации. Это значит, что в соответствующих графах эти слова продолжают занимать свои «прежние» узлы. Другой случай синтаксической деривации представляют отглагольные прилагательные:

9. перелетные птицы

10. дойная корова

11. жилой дом

В приведенных примерах реализуются разные отношения к актанту. В первых двух случаях можно даже говорить о диатезах: активной (субъектной) и пассивной (объектной).

При двухместных предикатах-прилагательных иногда выражен и второй актант. Данный актант может быть инкорпорирован в прилагательное (12). И в отглагольное прилагательное нередко инкорпорирован второй актант глагола, т. е. дополнение (13):

12. <любве>обильное сердце

13. <трав>ядное животное

В примере *дальновидные правители* инкорпорированный актант в непроизводном виде выражен как наречие (*видящие далеко*). Однако опорный компонент *-видный* может соотноситься не с глаголом *видеть*, а с глаголом *выглядеть*; ср.: *серповидный предмет* — *Предмет выглядит как серп*. Как видно, морфема <серпо> восходит к союзной группе.

В редких случаях в инкорпорированном актанте, восходящем к предложной группе, может быть сохранен предлог: <сног>*сшибательное открытие*. Если же к предложной группе восходит главное слово, предлог, конечно, теряется (ср. *жить в доме* — *жилой дом*).

Инкорпорированный элемент не всегда представляет собой актант прилагательного. Отношение между ними может быть и обратное:

14. <низко>рослые деревья

15. <трудно>выездная Россия¹

Есть случаи, когда актантом прилагательного является не существительное, как обычно, а глагольный сегмент, инкорпорированный в существительное (16). Такие словосочетания находим преимущественно в синтаксических конструкциях типа (17), с синтаксическим сказуемым.

16. быстрый <преда>тель

17. Я — быстрый (<преда>тель)².

Ср. непроеизводную парафразу *Я быстро предаю*. Ср. также англ.: *I am an early riser*.

Теперь рассмотрим примеры, где предикат выражен «некорневой» лексемой. Речь идет о предикатах типа *каменный*, где корень обозначает не сам предикат, а один из его актантов. Содержание такого предиката задается непосредственно: оно извлекается из сочетания актантов. В более развернутых парафразах этим прилагательным соответствуют слова разных частей речи: предлогов, союзов и глаголов, напр.: *каменное лицо* — *лицо как камень*, *песчаный берег* — *Берег состоит из песка*. В паре

18. <пыль>ная дорога

19. <лес>ная тропинка

прилагательные синонимичны предлогам *на* и *в*, соответственно (*пыль на дороге*, *тропинка в лесу*). Приведенные примеры являются своего рода конверсивами. Есть и другие конверсивы подобного типа, например:

20. <май>ская ночь

21. <гриб>ная осень.

В основе этих фраз лежат предикаты временного значения (*ночь в мае*; *осень, когда много грибов*). В следующей паре представлен динамический предикат: «изготовить», «построить».

22. <золот>ое кольцо

23. <корабль>ные сосны

Здесь конверсивным отношением затронут и временной план (материал существует раньше продукта). Словосочетанию (22) соответст-

¹ Пример из прозы Е. Евтушенко.

² Пример из романа Г. Щербаковой.

вует парафраза *кольцо, изготовленное из золота* (с синтаксическим предлогом) или — проще — *кольцо из золота* (с семантическим предлогом). Словосочетание (23) немного сложнее: оно соотносится, скорее всего, с исходной парафразой *Из этих сосен можно (было бы) строить корабли*. Главное слово может обозначать не материал, а место:

24. <уголь>ные шахты



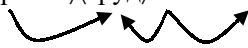
25. (В) шахтах добывают уголь.



Глагол *добывать* рассматриваем как трехместный, т. к. он близок к глаголам *искать*, *находить*, *обнаруживать*.

Прилагательное может содержать более, чем одну инкорпорированную морфему:

26. <красно><груд>ый снегирь



27. <сабле><как><зуб>ый тигр



28. <когт><ист>ые лапы



29. <голого><кружи>тельный вид



В примерах 26—28 прилагательные синонимичны предлогам *с* или *у*, в примере 29 — предлогу *от*; ср. парафразу *От этого вида кружится голова*. По поводу примера 28 можно отметить, что по своему значению морфема *-ист* более определенная, чем само прилагательное, но менее определенная, чем корневая лексема. Ср. парафразу *лапы с острыми когтями*.

Иногда не только прилагательное, но и существительное является предикатным словом. В словосочетаниях типа *круглосуточное наблюдение*; *четырёхчасовой сеанс* прилагательное синонимично флексивной морфеме или глаголу *длиться*; ср.: *вести наблюдение круглые сутки*; *сеанс длится четыре часа*. Вот еще два таких примера:

30. обувь <ино><стран>ного производства



31. <спирт>овая острота



Исходные парафразы: *обувь, произведенная в другой стране*; *острота, как у спирта* (если возможна парафраза *острота спирта*, пример относится ко второму типу, см. ниже).

Очень разнообразен по своему выражению предикат *использовать*³. В частности, он представлен такими прилагательными, как *отбойный* (*молоток*), *посудомоечная* (*машина*). Вот анализ последнего из этих примеров:

³ Благодарю И. М. Богуславского за мысль, высказанную в личной беседе, что данный предикат можно рассматривать как трехвалентный.

32. <посудо>(моеч)ная машина 33. машина для мытья посуды
-

Ср. с имплицитным предикатом «использовать»: *огнетушитель*.

Первым актантом пространственного двухместного предиката вместо главного слова может служить извлекаемый из него имплицитный предикат. Так, из названия артефакта можно извлекать способ его возникновения:

34. <англий>ская «изобрести» булавка
35. изобретенная в Англии булавка
-

Второй тип (*головная боль*).

Главное слово в большинстве случаев отглагольное или же оно имеет соответствующую семантику, напр. *лесной пожар* — *Лес горит*. Кроме существительного в роли подлежащего или дополнения (ср. *рыбная ловля*), прилагательное может в исходной парафразе соответствовать предложной группе или наречию:

36. орловское происхождение автора
37. корабли дальнего плавания
-

В примере (38) имеет место тавтология, поэтому лишний элемент обозначен как синтаксический. Ср. исходную парафразу (39).

38. его <скелето>(образный) вид 39. Он выглядит (как) скелет.
-

Предикаты, обозначающие параметры, всегда двухместные:

40. <полу>метровая глубина пруда
41. пруд глубиной (в) <пол>метра
-

Третий тип (*внутрипартийные конфликты*).

Наряду с очень распространенными случаями типа *загробная жизнь*, главное слово может быть объектом инкорпорированного глагола:

42. <склад>ные стульчики



43. Стульчики можно складывать.



44. <не><излеч>имая болезнь



45. Болезнь <не>возможно вылечить.



46. <машино><пис>ный текст



47. Текст напечатан на машинке.



В следующем примере главное слово — актант инкорпорированного прилагательного:

48. <бело><снеж>ные новостройки



49. белые, как снег, новостройки



Но сравните внешне схожий пример, относящийся, однако, к первому типу:

50. <бело><мрамор>ные ступени



Четвертый тип (*Божьи создания*).

Парафразами могут служить субстантивированные выражения; ср. *Божьи создания* — *созданное Богом* а также — с другой «диатезой» — *сестрины вздыхатели* — *вздыхающие по сестре*. Вот пример с двухместным же предикатом *останавливаться*:

51. автобусная <останов>ка



52. <останов>ка автобуса



Но парафраза (52), в отличие от (51), двусмысленна; ср.: *до полной остановки автобуса*.

В примере (53) добавлен третий актант:

53. Петин <учи>тель рисования



54. Он учит Петю рисовать.



К четвертому типу отнесем также случаи, когда название агенса инкорпорировано в название действия, а не наоборот.

55. кладбищенский <сторож>[ить] 56. карманный <вор>[овать]



Ср. с имплицитным «воровать»: *карманник*.

В пределах четвертого типа встречаются также примеры с инкорпорированным предлогом *с* (т. е. с предикатом-приставкой):

57. Петин <со>[служи]вец



58. Верины «в» <со>[камер]ницы.



Теперь рассмотрим случаи, когда объединяющий предикат имплицитен, но извлекается из значения существительного. Приведенный уже пример *командный пункт* можно соотнести с *в этом пункте командуют* или *пункт, где командуют*. Первым актантом может быть некорневая лексема:

59. <камен>ный «в» век



60. В [том] веке делали [орудия] (из) камня.



Из названия деятеля можно извлечь название действия:

61. Петин врач «лечить»



62. Он лечит Петю.



Что касается второго актанта, то особенно легко поддаются извлечению *verba efficiendi*:

63. Петин роман «написать»



64. Петя написал роман.



Но ср.: *Петин роман с соседкой* (этот пример относим ко второму типу).

Колеблются между двумя типами примеры, возводимые к альтернативным исходным парафразам; например, для словосочетаний *нефтеочистительный завод*, *книжный магазин* возможна парафраза с главным словом в роли подлежащего или в роли обстоятельства места (*магазин продает* / *в магазине продают*).

Литература

Мельчук И. А. Опыт теории лингвистических моделей Смысл ↔ Текст. М., 1974.

Murphy, L. Paradigmatic Relations and Semantic Information in MTT: some questions. MTT 2003 Proceedings. Paris. P. 11—20.

Наш сосед по старой даче:
предложные актанты и модификаторы
у предметных и реляционных имен *

О. Н. Ляшевская

ВИНИТИ РАН
olesar@mail.ru

Рассматриваются случаи предложного кодирования актантов и модификаторов имен лиц и предметов (ср. *сосед по даче, мясо на углях*). Сложность в определении статуса предложных сочетаний состоит в том, что здесь не применимы стандартные для глаголов синтаксические критерии. Предлагается семантический критерий различения актантов и модификаторов: если вводимый предложным сочетанием предикат и его аргументы отождествляются с (главным) предикатом толкования предметного имени, распространитель считается актантным.

Nash sosed po staroj dache: Prepositional actants and modifiers of names of things and relational nouns. Cases of prepositional coding of actants and modifiers of names of persons and things (cf. our old cottage neighbour, meat on coals) are considered. The status of the prepositional constructions is hard to define because syntactic criteria standard for verbs are of no use here. The semantic criterion to distinguish between actants and modifiers is offered: if a predicate introduced by a prepositional group and its arguments are identical with the (main) predicate of the lexical decomposition of the noun, then the preposition group is considered to be actantial.

Вопрос о том, есть ли валентности у имен существительных, решается по-разному в разных семантических теориях. Обычно признают наличие актантов у отглагольных и прочих имен, обозначающих ситуацию (S_0): *присутствие Петра, мечты о новом доме*, а также у реляционных имен типа *друг, отец*. Что касается предметных имен, то согласно западным формально-семантическим теориям (Dowty 1979, Grimshaw 1990) у них могут быть участники (partici-

* Исследование выполнено в рамках проекта РФФИ № 05-06-80396.

pants) в лексико-семантическом представлении, однако эти имена не могут присоединять синтаксические аргументы, и следовательно, все их зависимые считаются модификаторами. У предметных имен постулируется только один «внешний» аргумент, ассоциируемый с референтом данного имени, например, Dog $\langle R \rangle$, и с его помощью описывается формализм взаимодействия имени с прилагательными и прочими распространителями, ср.: «в сочетании *large dog* аргумент прилагательного *large* связывается (coindexed) с аргументом R имени *dog*» (Grimshaw 1990: 97)¹.

Модель «Смысл $\Leftrightarrow$ Текст» признает для предметных имен существование традиционных валентностей, как семантических, так и синтаксических, аналогичных валентностям глаголов и имен ситуаций (см., например, Mel'čuk 2004): *грузовик с кирпичами*, *окно в парк*. Однако, поскольку имена *грузовик* и *окно* не хочется считать предикатами, соответствующие участники фиксируются как аргументы при предикатах, встроенных в толкование предметных имен (Рахилина 2000)², а для определения состава валентностей используется принцип наследования актантов от предиката, ср.:

*грузовик*_Z $\langle X, Y \rangle \leftarrow$ перевозить $\langle X$ (кто), Y (что), Z (на чем), W (откуда), T (куда) $\rangle$ ³.

Наиболее распространенными видами актантов при предметных именах являются ПОЛЬЗОВАТЕЛЬ и другие агентивные участники: *топор*_Z (X рубит Y Z -ом), *лыжи*_Y (X перемещается... с помощью Y -а), *магазин*_W (X обслуживает... Y -ов), а для имен пространственно ориентированных объектов также различные ОРИЕНТИРЫ: *дыра в кармане*, *окно в крыше*, *пятно на скатерти*, *поезд Москва-Сочи*, *письмо из Одессы*, *забор вокруг участка* (примеры из СИИ 32). Бо-

¹ Ср. похожий формализм с аргументом a_0 , предложенный А. Н. Барулиным (1996). В ТКС (Мельчук, Жолковский 1984) валентность, замещаемую самим именем, можно обнаружить в толковании словосочетания УЧЕБНОЕ ЗАВЕДЕНИЕ: 'учреждение W , предназначенное для того, чтобы его сотрудники Z учили 1 X -ов Y -у', где W — четвертый актант глагола *учить* 1. Впрочем, подобное решение не проводится в словаре последовательно, и у других предметных имен такая насыщенная валентность не упоминается ни в толковании, ни в модели управления.

² О предикатах в семантике предметных существительных см. также Pustejovsky 1995.

³ Далее в работе для указания тождества между актантом глагола и предметным именем мы будем указывать при существительном соответствующий нижний индекс: *грузовик*_Z $\langle X, Y \rangle \leftarrow$ перевозить $\langle \dots, Z$ (на чем), $\dots \rangle$.

лее редкие актанты — это, например, ЗНАЧЕНИЕ ПАРАМЕТРА (*градусник, часы*: актанты «сколько градусов», «сколько часов» никогда не выражаются) или СРЕДСТВО: *дровяная печь, костер из книг*. Не имеют валентностей только нефункциональные имена: растения, природные объекты (*холм*), небесные тела (*звезда*) и некоторые простые предметы (*палка*).

В работе Рахилина 2000 указываются следующие основные отличия в моделях управления предметных имен и предикатов:

— актанты предметных имен как правило необязательны или их поверхностное выражение невозможно, ср. *убийца топором;

— разные актанты одного предметного имени часто имеют одинаковый способ выражения (ср. *бабушкин подарок* = *подарок от бабушки* vs. *подарок бабушке*) и не всегда могут быть выражены одновременно;

— актанты при предметных именах могут выражаться иначе, чем соответствующие актанты при глаголах: *соревноваться в беге* ~ *соревнования по бегу* (принцип ненаследования).

Существует два основных способа кодирования актантов при существительном: (1) с помощью генитива, поссессива или адъектива (*подарок бабушки, бабушкин подарок, обувной магазин*) и (2) с помощью предложных групп (*фильм о войне*). И тот и другой способ используются также для кодирования модификаторов: например, генитив и поссессив регулярно кодируют ПОСЕССОРА (*бабушкин сундук*), а сочетания с предлогом ПО, среди прочего, СТОИМОСТЬ предмета (*картошка по 10 рублей*).

Далее будут рассмотрены случаи предложного кодирования актантов и модификаторов имен лиц и собственно предметов (ср. *сосед по даче, мясо на углях*). Оговорим, что описания моделей управления обычно ориентированы на синтез и поэтому в них в центре внимания находятся способы выражения валентностей и ограничения на их заполнение для разных семантических типов предикатов и участников. Проблема противопоставления актантов и сирконстантов носит, тем самым, практическое применение: где в словаре отражать информацию о данном участнике: в модели управления (для обязательных участников) или в других зонах (в случае нетривиального выражения необязательных участников)? В предлагаемом исследовании мы хотим, напротив, отвлечься от проблемы способа выражения и сосредоточиться на том, как отличить актантные и сирконстантные распространители существительных в процессе анализа текста.

Конструкция с предлогом ПО

Распространители с предлогом ПО являются типичным обстоятельством при предметных именах: ср. *слесарь по имени Гриша*, *картошка по (цене) 10 рублей*, *костюм по моде 40-х годов* и др. В сочетаниях с реляционными именами типа *пришел мой сосед по даче / по подъезду / по парте*; *мои соседи по купе* этот распространитель выражает, на первый взгляд, также некоторую новую информацию: ср. фразу без распространителя *пришел наш сосед*; поэтому сочетания с ПО интуитивно хочется считать модификаторами. Сравним однако толкование предметного имени и толкование всей предложной конструкции.

В толкование имени *сосед* встроен предикат со следующими аргументами:

находиться рядом $\langle X, с Y-ом, в пространстве Z \rangle$.

Кроме того, имя вводит в рассмотрение некоторую ситуацию W (*жить, сидеть в классе, ехать поездом* и т. п.), конкретизируя тем самым предикат *находиться*. В примере *мой сосед по парте* референт имени *сосед* кодирует валентность X, прилагательное *мой* замещает валентность Y, предложная группа *по парте* связана с валентностью места Z и вводит в рассмотрение ситуацию W *сидеть в классе*. Таким образом, есть все основания полагать, что предложная группа кодирует обязательный семантический актанта.

Откуда же тогда возникает ощущение, что предложный распространитель модифицирует значение имени? Как кажется, дело в том, что у имени *сосед* существует стандартный способ насыщения валентностей Z и W. При невыраженных валентностях в нейтральном контексте (ср. *он позвонил соседу*) имя *сосед* интерпретируется как человек, который *живет* (W) в соседнем доме, квартире или комнате (Z). Соответственно, ситуация *сидеть в классе за одной партией* является новой информацией по сравнению с ожидаемой.

Кроме того, предложные сочетания могут нести дополнительную информацию, не сводимую к идее местонахождения, например, сочетание *мой сосед по парте* имплицитно несет информацию о ситуации 'я учусь в школе', а сочетание *мой сосед по даче* — информацию о ситуации 'у меня есть дача'. Предложные распространители могут также модифицировать время этой ситуации, ср.:

Наш сосед по старой даче $\supset$ *Раньше у нас была другая дача*

Наш приятель по Душанбе $\supset$ *Раньше мы жили в Душанбе*⁴

и в результате *сосед по старой даче* не является *соседом* в собственном смысле, также как *друзья по бане* не являются настоящими друзьями (ср. *Новый мэр назначает руководителями департаментов своих друзей по бане*).

Если мы рассмотрим весь список реляционных имен, допускающих распространитель с предлогом ПО: *друг, приятель, коллега, напарник, соратник, сосед, дед, предок*, — то обнаружим, что участник ситуации Z (иногда W: *друзья по переписке*) может иметь различный статус. Сочетание ПО + S_{dat} при именах *сосед, коллега, соратник* вводит в рассмотрение некий предикат P (*жить, находиться, работать* и др.) $\langle X \text{ (кто)}, Z \text{ (где)} \rangle$. Поскольку он совместим с предикатом *находиться рядом / работать вместе*, входящим в толкование указанных имен лиц, и не вводит никаких новых актантов, участника Z можно считать актантом существительного. Небольшое различие между именем *сосед* и именами *коллега, соратник* состоит в синтаксической обязательности валентности. У имени *сосед* актант Z не требует обязательного выражения (фраза *встретил соседа* обычно понимается как ‘того, кто живет рядом’, см. выше); имена *коллега, соратник* в большей мере требуют внешнего выражения, ср. тривиальное насыщение валентности *коллега по работе*.

Что касается имен *друг* и *приятель*, то здесь участника Z (W) следует трактовать как модификатор-сирконстант, ср. *друзья по школе; приятель по рыбалке*. Предикаты отношения и общения, занимающие центральное место в толковании имен, не содержат актанта Место, ср. *друзья по школе* и *общаться в школе*.

В работе Рахилина 2000 была рассмотрена конструкция с предлогом ПО в значении выбора из списка альтернатив, ср. *заместитель директора по кадрам, инструктор по плаванию, тетрадь по математике, пенсия по старости, токарь по дереву, пила по бетону* и т. д. В этих примерах сочетание с ПО заполняет некоторую валентность имени, например:

пила_Z по бетону_Y $\leftarrow$ *пилить* $\langle X \text{ (кто)}, Y \text{ (что)}, Z \text{ (чем)} \rangle$.

⁴ Заметим, что если в первом случае виновником сдвига во времени является прилагательное *старый*, модифицирующее значение имени *дача* (*наша старая дача₂* $\supset$ *у нас есть дача; раньше была другая дача*), то во втором случае причиной появления компонента ‘раньше’ может быть только сама конструкция с ПО.

Интересен случай, когда эта альтернатива, к которой апеллирует конструкция с ПО, уже заложена в толковании имени, ср.

дед по матери:

дед $\langle Y \rangle \approx$ 'отец отца или матери Y -а'

Требуется няня по уходу за больными / по уходу за ребенком:

*няня*_X $\langle Y \rangle \approx$ 'женщина (X), которая ухаживает за ребенком или больным Y '

В современном русском языке замечена тенденция к переходу от генитивного управления к модели с предлогом ПО⁵ для многих имен, у которых конструкция с ПО интерпретируется как выбор из ряда альтернатив. Наши наблюдения за употреблением в Интернете выражений типа *учитель по физике* и *учитель физики* показывают следующее соотношение:

ПО + Dat.		Gen.	
учитель по математике	4705	учитель математики	65553
учитель по физике	3244	учитель физики	55277
учитель по труду	1712	учитель труда	10901
учитель по языку	98	учитель языка	2569
учитель по гитаре	615	учитель гитары	252
учитель по игре (на фортепиано)	215	учитель игры (на фортепиано)	3478

Конструкция с ПО позволяет также обойти запрет на выражение актанта:

**студент вокала и студент по классу вокала:*

*студент*_X $\langle Z, Y, W \rangle \leftarrow$ *учиться* $\langle X$ (кто), Z (чему), Y (у кого), W (где));

ср. также ??*учитель языка Си++* и *учитель по языку Си++*; ?*министр чрезвычайных ситуаций* (825 употреблений в Интернете) и *министр по чрезвычайным ситуациям* (18385 употреблений).

«Альтернативная» конструкция с ПО распространяется также на имена лиц, в семантике которых отсутствует соответствующий слот, ср. объявление в метро:

Требуются женщины по ремонту радиоаппаратуры.

Естественно, что в этом случае мы имеем дело с чистым модификатором значения имени.

⁵ См. в этой связи работу Гловинская 1996.

Конструкция с предлогом НА + Loc

Ранее мы рассматривали взаимодействие аргумента предлога ПО с центральным предикатом толкования существительного. Аналогичным образом, среди сочетаний с предлогом НА обнаруживаются актантные распространители, взаимодействующие с главным предикатом: *пятно на скатерти* (предикат *покрывать поверхность*), *джинсы на бедрах* (предикат *носить*) и др.

В то же время, согласно теории Дж. Пустеевского (Pustejovsky 1995), в толковании предметных имен выделяются несколько зон, описываемых с помощью предикатов

- «быть частью / целым»;
- «функция (назначение)»;
- «создание».

У имени *джинсы* главный предикат толкования (*носить*) относится к типу «функция». Информация, связанная с предикатами других типов («быть частью / целым» и «создание»), не требует обязательной экспликации, так как и говорящий, и слушающий обладают соответствующими знаниями. Распространение предметного имени с помощью сочетания с предлогом НА позволяет актуализировать эту информацию и активировать «спящие» валентности. Так, мерееологический предикат и его аргументы актуализуются в сочетаниях *куртка на молнии* и *ботинки на толстой подошве*, а предикат создания и его аргументы в выражениях *каша на молоке* и *тесто на дрожжах*.

Поскольку предикаты типа «часть / целое» и «создание» не входят в центральную зону толкования, возникает сомнение, можно ли считать участников *молния*, *подошва*, *молоко*, *дрожжи* актантами предметного имени. Как кажется, они занимают промежуточное положение между актантами и «чистыми» модификаторами:

- мясо_У на углях_З ← готовить ⟨X (кто), Y (что)... Z (способ: циркулянт)⟩*
- самая чистая машина_У на бензине_З ← функционировать ⟨Y, с помощью средства Z⟩*

* * *

Сложность в интерпретации предложных оборотов как актантов или модификаторов имен существительных состоит в том, что здесь не применимы стандартные для глаголов синтаксические критерии (критерий «неопущения» актанта, критерий облигаторности в смысле работы Шмелев 1998). Рассматривая сочетания предметных и ре-

ляционных имен с предлогами ПО и НА, мы наблюдали континуум в актантных и сирконстантных употреблениях:

- выражение обязательного актанта (*коллега по работе*);
- выражение факультативного актанта (*сосед по парте*);
- экспликация константы в толковании (*дед по матери*);
- насыщение стандартно блокируемого семантического слота (*студент по классу вокала*);
- выражение актантов периферийных предикатов толкования (*каша на молоке*);
- модификация (введение нового предиката) (друг по переписке, женщина по ремонту радиоаппаратуры).

Актантные и сирконстантные употребления с предложными сочетаниями имеют общие механизмы интерпретации, однако в одном случае заполняются обязательные семантические слоты, в другом активизируются «спящие» валентности, в третьем случае — случае модификаторов — в интерпретации примера появляется новый предикат.

Литература

- Барулин 1996 — Барулин А. Н. К построению модели синтеза русских нумеративов (глубинное и поверхностно-семантическое представление) // Московский лингвистический журнал. № 2. 1996.
- Гловинская 1996 — Гловинская М. Я. Активные процессы в грамматике (на материале инноваций и массовых языковых ошибок) // Е. А. Земская (ред.). Русский язык конца XX столетия (1985—1995). М.: Языки русской культуры, 1996.
- Мельчук, Жолковский 1984 — Мельчук И. А., Жолковский А. К. Толково-комбинаторный словарь современного русского языка. Вена: Wiener Slawistischer Almanach, 1984.
- Рахилина 2000 — Рахилина Е. В. Когнитивный анализ предметных имен: семантика и сочетаемость. М: Русские словари, 2000.
- СИИ 32 — Семиотика и информатика. Вып. 32. 1991. М.: ВИНТИ.
- Шмелев 1998 — Шмелев А. Д. Типы «невывраженных валентностей» // Семиотика и информатика. Вып. 36. С. 167—176.
- Dowty 1979 — Dowty D. R. Word meaning and Montague grammar. Dordrecht: D. Reidel, 1979.
- Grimshaw 1992 — Grimshaw J. Argument structure. Cambridge (MA), London: MIT Press, 1992.
- Mel'čuk 2004 — Mel'čuk I. A. Actants in semantics and syntax, I: actants in semantics // Linguistics. 42.1. 2004. P. 1—66.
- Pustejovsky 1995 — Pustejovsky J. The generative lexicon. Cambridge (MA): MIT Press, 1995.

Meaning-Text Theory and the Conceptual Apparatus of Modern Linguistics: Towards a Better Description of Case and Voice in Maasai

I. Mel'čuk

Observatoire de linguistique Sens-Texte, Université de Montréal
Igor.Melcuk@umontreal.ca

1. One of the goals of the Meaning-Text theory is a conceptual system for linguistics; its first stage is a conceptual system for morphology, and this system is tested on the data of Maasai — cases and voices.

2. Linguistic concepts must satisfy several substantive and formal requirements (to aim at the most general class of phenomena possible, to cover the prototypical kernel of the concept, to be rigorous, etc.).

3. The Maasai cases: traditionally, the 'nominative' and the 'accusative'. It is shown that the traditional NOM is an oblique, while the traditional ACC is a nominative.

4. A corollary: the 'Subject + Main Verb' construction of Maasai is ergative — with all verbs (transitive and intransitive).

5. The SSynt-Subject of Maasai traditional passive in **-ki** is said to be in the ACC, which is typologically embarrassing. Based on a general calculus of voices, it is shown that the form in **-ki** is a partial demotional passive or a suppressive, so that the corresponding noun is its DirO, rather than its Subject (and it is of course in the 'new' nominative).

Теория «Смысл — Текст» и понятийный аппарат современной лингвистики: к проблеме описания падежей и залогов в масаи.

1. Одна из целей теории «Смысл — Текст» — создание понятийной системы для лингвистики; ее первый этап составляет понятийная система для морфологии, которая иллюстрируется в докладе применительно к описанию падежей и залогов языка масаи.

2. Лингвистические понятия должны удовлетворять ряду содержательных и формальных требований (строиться на основе наиболее общего класса описываемых явлений; покрывать наиболее прототипические явления, составляющие ядро понятия; быть строгими; и т. д.).

3. Падежи в масаи: традиционно различаются номинатив и аккузатив. В докладе показано, что традиционный номинатив — это на самом деле обликвус, а традиционный аккузатив — в действительности номинатив.

4. Важное следствие: основная глагольная конструкция масаи («Подлежащее + Сказуемое») является эргативной при любом глаголе, как переходном, так и непереходном.

5. Традиционно считается, что обозначение пациенса при так называемой пассивной форме на -*ki* является ее подлежащим и стоит в аккумулятиве; такое положение дел типологически сомнительно. На основе универсального исчисления залогов, показано, что данная глагольная форма — это частичный понижающий пассив или бессубъектный суп-прессив, так что существительное, обозначающее пациенса, является его прямым дополнением, а не подлежащим (и стоит в номинативе, что нормально для прямых дополнений в эргативной конструкции).

1. Conceptual System for Linguistics

One of the central goals of the Meaning-Text theory [= MTT] is to develop a sufficiently rich conceptual/terminological system for linguistics that will allow linguists to formulate unambiguous statements about linguistic facts. This system of concepts must be rigorous and formal enough, so that the consistency and accuracy of linguistic statements could be established by logical procedures. The ultimate aim is the elaboration of a formal linguistic (meta-)language — an undertaking comparable to what was done in mathematics by Nicolas Bourbaki,¹ who rewrote the whole of the 20th century mathematics in terms of a unified formal metalanguage. To put it in simpler terms, linguistics needs a well-organized terminology. The first stage in the creation of the language of linguistics has been the development of a conceptual system and appropriate terminology for linguistic morphology (Mel'čuk 1993—2000, where one finds 248 concepts and corresponding terms). This conceptual system is a set of definitions of morphological concepts — such as 'linguistic sign,' 'morph,' 'wordform,' 'grammeme,' 'case,' 'voice,' etc.

The topic of the present paper is precisely the importance of good terminology. I will try to demonstrate how the application of a few well-defined concepts to a complicated enough linguistic situation — the entangled relationship between grammatical case and grammatical voice in

¹ Nicolas Bourbaki, a fictitious French mathematician, undertook, in the 30s of the 20th century, a huge task of reformulating the whole body of modern mathematics in terms of a unified conceptual apparatus so as to provide a solid foundation for mathematics. He managed to publish over 20 volumes of a highly formalized treatise that deals with all major aspects of the field — algebra, set theory, number theory, calculus, topology, mathematical logic, and many other topics. In point of fact, 'Nicolas Bourbaki' is a pen name for a team of French mathematicians, led by Jean Dieudonné, Henri Cartan, Claude Chevalley, and André Weil. Their method of exposition is axiomatic and abstract, proceeding normally from the general to the particular.

Maasai — can lead to a better description of linguistic facts. Given the limitations of space, I will reduce the references and explanations to a bare minimum. (A fuller version of this presentation can be found in Mel'čuk 1997 and 2005.)

2. Characteristics of the Linguistic Definitions Proposed

The definitions put forward within the MTT framework satisfy two sets of requirements, concerning their substantive and formal aspects.

Substantively, a definition must be:

- AS LARGE AS POSSIBLE (a linguistic phenomenon is defined starting from the most general class to which it belongs);
- based on MAXIMAL SEPARATION OF DEFINING FEATURES (cluster concepts are admitted only as the last recourse);
- designed to account, in the first instance, for the PROTOTYPICAL CASES of the concept — in order to be generalized so as to cover all appropriate peripheral cases.

Formally, a definition must be:

- formal, i.e. applicable literally;
- rigorous, i.e. containing only previously defined concepts and/or else indefinable concepts, listed as such;
- sufficient and necessary, i.e. covering all the phenomena that are perceived as being subsumable under the concept, and nothing but such phenomena;
- universal, i.e. applicable to any relevant phenomena of any language.

When considering a class of linguistic phenomena $\{P_i\}$ — the pre-theoretical set of data — which are to be subsumed under concept C , the definition of C is constructed by taking the following six steps:

1) Establish a 'kernel' subclass $\{P_j\}$ of the class $\{P_i\}$ ($\{P_j\} \subset \{P_i\}$) — that is, isolate among all the P_i s the phenomena P_j s that we would like to have covered by our definition under any circumstances. These P_j s correspond to the most typical particular case of C — in other words, to a prototypical C , symbolized as c . They will constitute the empirical basis of the future definition and are chosen quite intuitively; this choice is taken as a postulate.

2) Analyze the concept c to find its constitutive components.

3) Develop a calculus of all logically possible cases of c , presumably covered by C . To do so, combine the constitutive components of c in all

logically possible ways; try to explain the unacceptability of the combinations banned by the language.

4) Formulate the definition of *C* by generalization of the concept *c*; extract all underlying concepts vital for this definition and make sure that these can be defined in their turn.

5) Review the whole field by applying the definition of *C* to less clearcut, fuzzy or dubious items in the set $\{P_i\}$, in order to see whether all relevant phenomena have been covered.

6) Discard similar but essentially different phenomena *C'*, delimiting them with respect to *C*; sketch out a definition for *C'*, to make sure that this can be done in a reasonable way.

Now the definition of *C* is ready. We have to check it to make sure that it:

- (i) covers all items which are intuitively sufficiently similar to P_j s;
- (ii) rejects all items which are intuitively sufficiently dissimilar to P_j s;
- (iii) produces results for all intermediate domains where our intuition balks — results that can be supported by further arguments elaborated especially for the solution in question.

The proposed system of linguistic concepts hopefully satisfies these requirements. Below, it is tested against the data of Maasai — as far as Maasai cases and voices are concerned.

3. Case in Maasai

The Maasai noun distinguishes (tonally) two case forms. Traditionally, one case (which is, among other things, the case of the Grammatical Subject) is called the nominative and the other (among other things, the case of the Direct Object), the accusative. The accusative is the basic case: it is the accusative form that is given in the dictionary. (“When asked for the name of anything, the Maasai will always give the word in its *accusative* form”, Tucker and Mpaayei 1955: 175.)

The traditional accusative marks the noun in one of the following nine syntactic roles:

- 1) noun uttered in isolation (e.g., used to name something);
- 2) Direct Object [= DirO] of a transitive verb;
- 3) Surface-Syntactic Subject [= SSynt-Subject] of a passive verb, denoting, as is to be expected with passives, the patient;
- 4) predicative nominal with the verb -RÁ ‘be’ (as in ‘be **the chief**’)²;

² The symbol ´ indicates the high tone, the symbol ` — the low tone, and the symbol ^ — the falling tone.

- 5) direct address without a vocative particle (= bare address);
- 6) Possessor in a possessive NP (if introduced by the possessive particle);
- 7) complement of the associative-conjunctive particle O 'and' (syntactically similar to the Eng. *plus*, as in *plus he and Mary*);
- 8) prepositionless Indirect and Oblique Objects [= IndirO and OblO];
- 9) "subjects that occur before the verb" [these are in fact Fronted Topics].

The traditional nominative marks the noun in one of the four syntactic roles:

- 1) Subject of any verb not in the passive (i.e., the Subjects of both transitive and intransitive verbs) and of an adjective or a numeral used predicatively;
- 2) Agentive Complement [= AgCo] of a passive verb;
- 3) complement of a preposition (such as TE 'at/in/to/with');
- 4) complement of a vocative particle (masculine LO, LI and feminine NA — roughly, 'oh...').

The use of the traditional case terminology in descriptions of Maasai and other Nilotic languages raises serious objections. Specifically, the following practices seem highly problematic:

From a syntactic viewpoint

1. Calling the 'accusative' the form of the noun uttered in isolation.
2. Calling the 'accusative' the noun form used as a bare address and as the Subject of a passive verb.
3. Calling the 'nominative' the form of the noun governed by a preposition or a particle.

From a morphological viewpoint

4. Calling the 'accusative' the morphologically basic, i.e. unmarked, form of a noun (presented in the dictionary), while the morphologically complex, or marked, form is called the 'nominative.'

Using the terms *accusative* and *nominative* in the indicated way disrupts the generally accepted doctrine on grammatical case and runs counter to many universalist theories of syntax. It is detrimental to linguistic typology and even more so to all attempts to develop a universal linguistic theory and a corresponding linguistic metalanguage.

My solution is simple and drastic: simple, because it boils down to exchanging the names of the cases, and drastic, since it has many repercussions in the description of Maasai. It derives from the following definition of the nominative case:

Definition 1: Nominative Case

|| In language **L**, which has the inflectional category of case, the particular case that marks the noun used to name an object or a

situation — i.e., a noun uttered in isolation — is the *nominative*, whatever its other syntactic roles in **L** and whatever its formal exponents.

It is thus proposed to restore to the nominative its etymological meaning: 'the case of nomination'.

Under this proposal, Maasai (like all related languages) has the *nominative*, the former 'accusative', and the *oblique*, the former 'nominative'³. The resulting two-case system is typologically highly plausible: it is similar, for instance, to that of Kabardian/Circassian, Kurdish and Old French. With this new terminology, the description of the use of nominal cases in Maasai ceases to be exotic.

The 'new' nominative marks nouns in one of the following eight SSynt-roles:

- 1) the basic (lexicographic) form of a noun — that is, a noun form uttered in isolation;
- 2) DirO of a transitive verb (as it typically happens in the ergative construction);
- 3) predicative nominal with the verb -RÁ 'be' (*Árá Sirónkà* 'I-am Si-ronka', where the name *Sirónkà* is in the nominative);
- 4) bare form of address (without a vocative particle);
- 5) Possessor (introduced by the possessive particle in a possessive NP);
- 6) complement of the associative-conjunctive particle O 'and' ($\approx$ 'plus');
- 7) prepositionless IndirO and OblO;
- 8) Fronted Topic.

The set of SSynt-roles of the 'new' nominative in Maasai does not include the noun denoting the patient of a passive form (which it should, so to speak, inherit from the 'former' accusative, see item 3 in "The traditional description of case marking," Section 3). The reason is that this sentence element is the DirO, not the Subject, of what is traditionally considered a passive form and thus corresponds to item 2 in the revised list above. This will be explained more fully in Section 4.

The OBLIQUE marks nouns in one of the following four SSynt-roles:

- 1) Subject of any verb (including 'be') and of any predicative adjective or numeral;
- 2) AgCo of a passive verb;

³ I do not insist on a specific name for this case. I call it *oblique*, because in a system with only two cases this terminology seems the most appropriate. In a system with more cases, such as in Turkana or in Wappo, I prefer for this case the term *subjective* (cf. the term *Subject Case*, proposed in Sasse 1984: 111—112).

- 3) complement of a preposition;
- 4) complement of a vocative particle.

The new nominative is the formally basic, unmarked form, stored in the lexicon; the oblique is obtained from it by a *tonal apophony* — i.e., by replacing the tonal scheme of the nominative form by another tonal scheme. It is well known that the nominative tends to be formally unmarked — in other words, to have a zero exponent (a zero suffix or a zero apophony). This means that if among **L**'s cases just one is unmarked, it will be, more often than not, the nominative. That is exactly what we see in Maasai and other Nilotic languages, provided of course that the proposed change of case names is accepted. For instance, in Kalenjin, which has case suffixes (in certain noun classes), the picture is quite clear: only the oblique can be expressed by a non-zero suffix, and the 'new' nominative is formally unmarked (Tucker and Bryan 1966: 468). Much more examples from Nilotic languages can be given. This seems sufficiently convincing in order to accept the proposed terminological change.

4. The Basic Verbal Construction in Maasai

If the case names are changed as suggested above, the only exoticism (if this is really an exoticism) that remains in the description of Maasai is the following:

The basic verbal construction of Maasai is an *ergative construction*, since the Subject of any verb, including the verb 'be', and of any predicative adjective/noun, is never in the nominative: it is in the oblique.

Of course, the truth of this statement hinges upon the definition of the ergative construction. I uphold the definition of this concept as proposed in Mel'čuk 1978 and then developed in Mel'čuk 1988: 182, 251, 258—259ff and Mel'čuk 1992: 105ff.

Definition 2: Ergative Construction

An *Ergative Construction* is a construction
Surface-Syntactic Subject ← **subjectival** – Main Verb [= MV],
in which the Subject

1) can potentially express the Causer (that is, with appropriate verbs in the same diathesis⁴ as the MV); and

⁴ The concept of diathesis is vital in this context, but cannot be properly introduced (see Mel'čuk 1993, 1997, 2004 and 2005). Suffice it to say here that the diathesis of a lexical unit L is the correspondence between the semantic and the deep-syntactic actants of L.

2) is marked by a case other than the nominative, such that the choice of this case does not depend on the meaning of the Subject itself (i.e., on its determinacy, partitivity, etc.).

In other words, in a prototypical Ergative Construction of **L** — with a semantically transitive verb like ‘kill’, ‘build’, ‘write’, etc. — the Subject expresses the Causer; this is sufficient to call by analogy any construction “Subject + MV” in **L** ergative, provided its Subject is not in the nominative.

Consider the following Maasai sentences (the Subject and the name of its case are boldfaced; contextual tonal assimilations are not shown):

(1) Maasai

a. With the verb -RÍK ‘cause nausea’

Á + *rík* + *Ø* ***nanó*** *Sirónkà*
1SG.SUB-3SG.OBJ cause.nausea ACT I-OBL S.-NOM
‘I nauseate Sironka’.

vs.

Áa + *rík* + *Ø* ***Sirònkà*** *nánó*
3SG.SUB-1SG.OBJ cause.nausea ACT S.-OBL I-NOM
‘Sironka nauseates me’.

Áá + *rík* + *Ø* ***nanó*** *iyíé*
1SG.SUB-2SG.OBJ cause.nausea ACT I-OBL you_{SG}-NOM
‘I nauseate you_{SG}’.

vs.

Kí + *rík* + *Ø* ***íyie*** *nánó*
2SG.SUB-1SG.OBJ cause.nausea ACT you_{SG}-OBL I-NOM
‘You_{SG} nauseate I’.

É + *rík* + *Ø* ***ɔlkítèŋ*** *Sirónkà*
3SG.SUB-3SG.OBJ cause.nausea ACT ox-SG-OBL S.-NOM
‘The ox nauseates Sironka’.

vs.

É + *rík* + *Ø* ***Sirònkà*** *ɔlkítèŋ*
3SG.SUB-3SG.OBJ cause.nausea ACT S.-OBL ox-SG.NOM
‘Sironka nauseates the ox’.

b. With the verb -DÓL ‘see’

Á + *dól* + *Ø* ***nanó*** *Sirónkà*
1SG.SUB-3.OBJ see ACT I-OBL S.-NOM
‘I see Sironka’.

vs.

<i>Áa</i>	+ <i>dʒl</i>	+ $\emptyset$	<i>Sírònkà</i>	<i>nánú</i>
3SG.SUB-1SG.OBJ	see	ACT	S.-OBL	I-NOM
'Sironka sees me'.				
<i>Áá</i>	+ <i>dʒl</i>	+ $\emptyset$	<i>nanú</i>	<i>iyié</i>
1SG.SUB-2SG.OBJ	see	ACT	I-OBL	you _{SG} -NOM
'I see you _{SG} '.				
vs.				
<i>Kɪ</i>	+ <i>dʒl</i>	+ $\emptyset$	<i>iyie</i>	<i>nánú</i>
2SG.SUB-1SG.OBJ	see	ACT	you _{SG} -OBL	I-NOM
'You _{SG} see me'.				
<i>É</i>	+ <i>dʒl</i>	+ $\emptyset$	<i>ɔlkítèŋ</i>	<i>Sirónkà</i>
3SG.SUB-3SG.OBJ	see	ACT	ox-SG.OBL	S.-NOM
'The ox sees Sironka'.				
vs.				
<i>É</i>	+ <i>dʒl</i>	+ $\emptyset$	<i>Sírònkà</i>	<i>ɔlkítèŋ</i>
3SG.Subj-3SG.Obj	see	ACT	S.-OBL	ox-SG.NOM
'Sironka sees the ox'.				

The SSynt-Subject of a Maasai clause is taken to be the NP that is traditionally said to be the Subject: the Subject immediately follows the MV (and thus precedes the Object) and, in case of an intransitive verb, determines MV's person and number. In a transitive verb, the DirO also participates in agreement, as seen in (1a-b): the MV in 'I see Sironka' vs. 'I see you_{SG}' has different forms, as in 'Sironka sees me' vs. 'Sironka sees the ox'. (Person-number agreement is manifested in megamorph, or portmanteau, verb prefixes.) Although in the sentences in (1b) the Subject does not express the Causer, in (1a) it does; therefore, the Subject + MV construction which is used in Maasai for all verbs can be called ergative according to Definition 2 above.

Thus, the price of changing case names as suggested in **3** is the acceptance of an ergative construction for all verbs in Maasai. This price is not too high: this type of Ergative Construction 1) is typologically plausible (it is widely found outside of the Nilotic family) and 2) is widespread in the same geographical area where Maasai is spoken.

1) An 'overall' ergative construction

In the prototypical ergative construction, as found in Basque, Chukchee, Koryak, Hindi, Nepali, Kurdish, Georgian, etc., the Subject is marked in a special way only with transitive verbs; the Subject of intransitive verbs is marked with the nominative as is the DirO. But in Maasai the oblique case marks the Subject with all verbs — not only with transitives, but also with intransitives and even with adjectives and numerals used predicatively:

(2) Maasai

- a. *N + é + dóú n + kíshu óo*
 SEQ⁵ 3PL.SUB descend FEM cow-PL.OBL POSS
óo Kéékonyókie te Kínopóp
 POSS to
 'And then the cows of Kekonyokie came down to Kinopop'.

- b. *É + púó iltójaná*
 3PL.SUB go-PL⁶ person-PL.OBL
 'The people go'.

- c. *Biyót iltójaná*
 healthy-PL.NOM person-PL.OBL
 'The people are healthy'.

- d. *Á + rá nanó sápòk*
 1SG be I-OBL big-SG.NOM
 'I am big'.

- e. *Ná + bo nanó*
 FEM one I-OBL
 lit. 'Am-one[FEM] I' = 'I [a woman] am alone'.

Yet languages in which the SSynt-Subject is always in a case other than the nominative (i.e., than the case of nomination) are quite well known and even not so rare. Let me mention three such languages:

— First, Megrelian, where the ergative case in **-k** (different from the nominative in **-i**) marks the Subjects of all verbs: e.g., *K'oč + k kumortu* 'The-man came' and *K'oč + k gaagibu ėqar + i* 'The-man heated the-water'.

— Second, Wappo, where the ergative case in **-i** (opposed to the nominative expressed by a zero suffix) marks all Subjects, even those with passive forms and adjectival verbs: *Chic + i ĩolkhe* 'The-bear got-caught'; *Chic + i tučākhi* 'The-bear is-big'.

— And third, Japanese, with its subjective case in **-ga** (the nominative having a zero suffix) used for all Subjects. (The thematic marker **-wa** replaces case markers of Japanese nouns, so that a Subject as a communicative Theme does not have **-ga** on the surface.)

⁵ SEQ = *sequential*, a grammeme meaning 'and then...'.
⁶ The verb -LÓ 'go' is one of a few verbs in Maasai which have special 'plural' stems (in this case, suppletive: *púó*) used when the Subject is in the plural (cf. Item 2) in 4.

2) An ‘overall’ ergative construction in Eastern Africa

Even more interestingly, many languages where the SSynt-Subject is always in a case different from the nominative are found in Eastern Africa, not only among Paraniotic languages, but also among Berber-Libyc, Cushitic, and Nilo-Saharan languages. Here is a telling example. In Berber the Subject, when in a neutral position (immediately following the MV), is marked by the oblique case with all verbs, and the DirO (if there is one), is in the nominative (= citation form):

(3) Berber

a. *Ičča weqžun amšiš*
eat-AOR dog-SG.OBL cat-SG.NOM
‘The **dog** ate the cat’.

vs.

Ičča wemšiš aqžun
eat-AOR cat-SG.OBL dog-SG.NOM
‘The **cat** ate the dog’.

b. *Yenza wemšiš / weqžun*
be.on.sale-PRES cat-SG.OBL / dog-SG.OBL
‘The **cat** / The **dog** is being sold’.

A more traditional name for the Berber cases is *free state* [= NOM] and *construct state* [= OBL], but actually these forms are quite normal cases. (Sasse 1984: 120–122 draws a convincing parallel between Berber and Cushitic case distribution and marking. Cf. also Aikhenvald 1986 for some data and further references concerning the non-nominative character of the syntactic system in Afroasiatic languages, with special attention to Berber.)

Moving to Cushitic, consider, for instance, (4):

(4) Oromo (= Galla)

Muk + ní gog + e
wood OBL dry PAST.3SG.MASC
‘The **wood** dried up’.

vs.

Terfaa + n muka + Ø gog + s + e
OBL wood NOM dry CAUS PAST.3SG.MASC
‘**Terfaa** dried up the wood’.

The Subject is always marked with the oblique case, while the DirO is always in the nominative: it is the same situation as in Maasai, Teso, Turkana, etc., on the one hand, and in Kabyle Berber, on the other.

A similar phenomenon is found in Somali: the Subject phrase has a special marker, attached to the rightmost element of the phrase, while the DirO and other complements are in the citation (= basic) form. It is enough to leaf through the descriptions of a few Cushitic and Nilo-Saharan languages in Tucker and Bryan 1966 to recognize the familiar pattern: the basic (i.e. unmarked) form of the noun — called, as a rule, the accusative — is used in isolation and for all syntactic objects and complements, while the Subject (either uniformly or only in the post-verbal position in the presence of a preverbal DirO, as in Mursi) has special case marking; this form is traditionally called the nominative (e.g., Beja: p. 108, Dasenech: p. 205, Burji and Darasa: p. 253, Mursi: p. 546).

In the light of the above discussion it is reasonable to insist that the names of both cases should be exchanged as proposed above. It is also clear that the ergative construction of the type described — all the SSynt-Subjects in the oblique, and all the DirOs in the nominative — is really typical of many languages of North-East Africa. In this respect, one could probably talk of a *Sprachbund*. Therefore, the idea that the basic verbal construction of Maasai is ergative should not be perceived as something overwhelming (cf. Frajzyngier 1984a, b and Andersen 1988, where this idea is explicitly discussed as applied to Chadic and Nilotic languages).

5. Voice in Maasai

Maasai has a special verbal form with the suffix **-ki** (the allomorphs **-ki**, **-ki**, **-i**, and **-i** are distributed phonemically). Traditionally, this form is called the passive; but is it? The key to the understanding of the nature of the **ki**-form in Maasai lies in establishing its SSynt-Subject. Greenberg (1959) showed that the Subject of the 'passive' form in Maasai is in fact not the NP in the nominative (= the traditional 'accusative'), but a zero 3rd person plural pronominal lexeme (historically meaning 'people'). This follows from two facts:

Agreement of the Main Verb

In Maasai, the MV agrees in number and person with the Subject, and a transitive MV agrees also with its DirO. Now, the passive Subject person-number prefixes for a given person (of the patient) are the same as the active-transitive Object person-number prefixes of that person combined with a 3rd person Subject:

(5) Maasai

passive		active (with the 3rd person Subject)
'I am nauseated'	<i>áa + rík + í</i>	'He/They nauseate(s) me' <i>áa + rík</i>
'You _{SG} are nauseated'	<i>kí + rík + í</i>	'He/They nauseate(s) you _{SG} ' <i>kí + rík</i>
'He is nauseated'	<i>é + rík + í</i>	'He/They nauseate(s) him' <i>é + rík</i>
'We are nauseated'	<i>é + rík + í</i>	'He/They nauseate(s) us' <i>é + rík</i>
'You _{PL} are nauseated'	<i>é + rík + í</i>	'He/They nauseate(s) you _{PL} ' <i>é + rík</i>
'They are nauseated'	<i>é + rík + í</i>	'He/They nauseate(s) them' <i>é + rík</i>

The choice of the agreeing prefix indicates that the 'invisible' Subject of the passive form is 3rd person.

Plural stem verbs

In Maasai, an infinitive governed by the MV agrees in number with the Subject of the MV, not with its own — explicit or understood — agent (the corresponding grammemes on the infinitive and on the MV are boxed):

(6) Maasai (Tucker and Mpaayei 1955: 65)

A + *tárétò* *ɔltɔŋáni* / *iltɔŋána*
 1[SG] .SUB-3.OBJ helped person-SG.NOM / person-PL.NOM
á + *múk* < **áá* + *múk* > *enáishó*
 INF.[SG] brew < INF.PL brew > beer-SG.NOM
 'I helped the person/the people brew beer'.

vs.

Kí + *tárétò* *ɔltɔŋáni* / *iltɔŋána*
 1[PL] .SUB-3.OBJ helped person-SG.NOM / person-PL.NOM
áá + *múk* < **á* + *múk* > *enáishó*
 INF.[PL] brew < INF.SG brew > beer-SG.NOM
 'We helped the person/the people brew beer'.

As one can see, the infinitive '[to] brew' agrees in number with the Subject of 'help' ('I' vs. 'we') rather than with its own understood agent ('person' vs. 'people').

A few Maasai verbs have two different stems, one used with the Subject in the singular, and the other with the Subject in the plural — for instance, *ló* [singular Subject] ~ *púò* [plural Subject] 'go', *lotú* [SG] ~ *puonú* [PL] 'come', *tón* [SG] ~ *tóní* [PL] 'sit', *nyokiíé* [SG] ~ *nyokíoo* [PL] 'be red', etc. Greenberg quotes three idiomatic constructions in which the 'passive' of the two-stem verbs 'go', 'come' and 'sit' is used as an auxiliary and governs the infinitive (Tucker and Mpaayei 1955: 88). The first two constructions are periphrastic future passives (something like "They"-are-gone/come me to-beat ≈ 'I will be beaten'), and the third is another

periphrastic passive with the meaning 'by someone who stayed for the purpose' (lit. "*They*"-are-sat me to beat $\approx$ 'They stayed to beat me'). In these constructions, "a twofold choice between singular and plural must be made, first in employing the singular or plural stem of the auxiliary verb..., and secondly, in regard to the form of the infinitive. In all instances the choice is unequivocally plural" (Greenberg 1959: 173). Thus, to say 'I will be followed' Maasai uses (7):

- (7) *Áa* + *púó* + *í* *áà* + *sùj*
 3PL.SUB-1SG.OBJ go-[PL] PASS INF.[PL] follow
 lit. "'They"-are-gone-me to-follow' = 'I will be followed'.

The choice of the plural agreeing prefix on the infinitive and of the plural stem of the auxiliary clearly indicate that the invisible SSynt-Subject of the presumed passive form is in the plural.

Based on this evidence, we are forced to analyze the sentence (7) as having a zero dummy SSynt-Subject $\emptyset_{(3PL)}^{EMPTY}$ —an indefinite-personal pronoun, roughly equivalent to Eng. IT, Fr. IL, Ger. ES, but in the plural (i.e., "'they'"). The Maasai dummy Subject most probably was etymologically also semantically animate, but today it is semantically empty; the 'humanness' is, so to speak, included in the meaning of the passive form—it is used only when the presumed agent is human.

As a result, what is often considered as the SSynt-Subject of the Maasai form called the passive (= the NP that denotes the patient), is by no means its Subject: this NP is the DirO of the verb (this treatment is explicitly proposed in Payne *et al.* 1993). This NP is invariably in the nominative, as are all DirOs in Maasai, both pronouns and nouns; cf.:

(8) Maasai	PASSIVE	ACTIVE (with the 3rd person Subject)
'I am seen'	<i>áa</i> + <i>dól</i> + <i>í</i> <i>nánó</i>	'He/They see(s) me' <i>áa</i> + <i>dól</i> <i>nánó</i>
'You _{SG} are seen'	<i>kí</i> + <i>dól</i> + <i>í</i> <i>iyié</i>	'He/They see(s) you _{SG} ' <i>kí</i> + <i>dól</i> <i>iyié</i>
'He is seen'	<i>é</i> + <i>dól</i> + <i>í</i> <i>ninyé</i>	'He/They see(s) him' <i>é</i> + <i>dól</i> <i>ninyé</i>
'We are seen'	<i>é</i> + <i>dól</i> + <i>í</i> <i>iyíóók</i>	'He/They see(s) us' <i>é</i> + <i>dól</i> <i>iyíóók</i>
'You _{PL} are seen'	<i>é</i> + <i>dól</i> + <i>í</i> <i>mtái</i>	'He/They see(s) you _{PL} ' <i>é</i> + <i>dól</i> <i>mtái</i>
'They are seen'	<i>é</i> + <i>dól</i> + <i>í</i> <i>nincé</i>	'He/They see(s) them' <i>é</i> + <i>dól</i> <i>nincé</i>

Still, the form under discussion cannot be simply called 'impersonal active,' as Dimmendaal (1983: 72 and *passim*) appropriately calls the corresponding form in Turkana. The crucial difference is that while in Turkana an Ag(entive) Co(mplement) (= 'by N') is impossible with this form, it is possible with what Tucker and Mpaayei call 'passive' in Maasai (the AgCo is boldfaced):

(9) Maasai

É +rík + í nkishú aainei Imorrân
 3PL.SUB-3PL.OBJ lead PASS cow-PL.NOM my (young)warriors-OBL
 'My cows are / will be lead by (young) warriors' (Tucker and Mpaayei 1955: 81, §94).

É +ípót + í éñkerái
 3PL.SUB-3SG.OBJ call PASS child-SG.OBL
 'He is called by the child' (Tucker and Mpaayei 1955: 176, (ii)).

É +irɔr + ɔkɔ + kí yíóók iltɔɔaná
 3PL.SUB-1PL.OBJ speak APPL.PAST PASS us-NOM person-PL.OBL
 'We were greeted by the people' (Tucker and Mpaayei 1955: 132, §172).

[The APPL(icative) form of the verb IRɔ 'speak' means 'greet'.]

É +ísís + í iltɔɔaná
 3PL.SUB-3SG.OBJ praise PASS person-PL.OBL
 'He is praised by the people' (Tucker and Mpaayei 1955: 176, (ii)).

It is true that in most examples of passives given by Tucker and Mpaayei (1955) the AgCo is absent (it is contextually understood or simply irrelevant); so it is possible that the passive without AgCo is preferred in Maasai. Yet some unambiguous examples with the AgCo in the oblique are present and the possibility of AgCo's is explicitly mentioned by the authors (p. 176: "when acting as ... agent of a passive verb, ... the noun..."). Therefore, the verbal form which I am discussing represents the DEMOTION of the SSynt-Subject to an AgCo, the real SSynt-Subject being replaced by a dummy — a zero empty lexeme of the 3rd person plural. The construction is superficially very similar to what is observed in the following French sentences:

(10) *Il a été raconté beaucoup d'histoires horribles par les survivants,*
 lit. 'It has been told many awful stories by the survivors'.

Il a été voté par ce parlement des lois qui nous semblent iniques,
 lit. 'It has been voted by this parliament laws which seem iniquitous to us'.

But in actual fact, the Maasai and the French constructions are essentially different. In French, the postverbal prepositionless NP is by no means a DirO (unlike a DirO, it cannot be cliticized: this is a Pseudo-Subject), while in Maasai it is a DirO. In addition, the dummy Subject IL in French is in the singular, while in Maasai, the dummy Subject $\emptyset_{(3PL)}$

is in the plural. Finally, the French IL is not phonologically zero while the Maasai dummy is.

From the viewpoint of the general calculus of voices proposed in Mel'čuk 1993, the Maasai situation exemplifies the following modification of the basic diathesis of the verb (Item 3, p. 14; see also Mel'čuk 2004: 295):

X	Y	⇒	X	Y
I	II		III	II

The phrase which in the basic diathesis of the lexical unit L is L's DSyntA I (= the SSynt-Subject) and corresponds to its Sem(antic) A(ctant) 'X' becomes its DSyntA III (= its AgCo) in the derived diathesis of L, while L's DSyntA II (= its DirO, corresponding to the SemA 'Y') remains in place. Moreover, on the Surface-Syntactic level, a dummy element (in Maasai, the zero pronoun $\emptyset_{(3PL)}^{EMPTY}$), which does not correspond to any SemA or DSyntA of L, is introduced at the Surface-Syntactic level and becomes the SSynt-Subject of the verb form under consideration.

This voice is called **Partial Demotional Passive**. Often it is referred to as Impersonal Passive (*impersonal* means here 'having an impersonal pronoun as a dummy SSynt-Subject'). Thus we come to roughly the same conclusion as Perlmutter and Postal (1984: 159–165), who argued — against Keenan (1976) — that the Maasai passive is, in point of fact, an impersonal passive — that is, a passive without a 'real' SSynt-Subject.

But this is still not all. Heine and Claudi (1986: 80) report that many Maasai speakers do not allow the AgCo to appear with a passive form: "For the majority of Maa [= a dialect of Maasai. — I. M.] speakers, the use of the passive suffix and agent coding are mutually exclusive." Some allow the AgCo, but only after an intonation break as a 'cleft' element; some allow it only under certain conditions. The same statement is found in Payne *et al.* 1993: 29–30. I am in no position to discover the truth about the use of the AgCo in Modern Maasai and, moreover, it is quite possible that there is no single truth: the Maasai passive construction is in a transitional stage (Heine and Claudi 1986: 82), so that numerous hesitations and disagreements among speakers are the norm. To account for this, we need two different descriptions of the construction: one with and one without an AgCo. For the passive form that potentially takes an AgCo, such a description has already been given — Partial Demotional Passive. However, the form that precludes an AgCo shows a different type of diathesis modification: there is no permutation of Deep-Syntactic actants, since the NP that is inherently DSyntA I, corresponding to X, is

suppressed rather than demoted, and the inherent DSyntA **II** retains its role. The addition of a dummy SSynt-Subject (= the zero pronoun $\emptyset_{(3PL)}^{EMPTY}$) does not change the nature of this operation. The corresponding modification of the basic diathesis can be represented as follows:

X	Y	⇒	X	Y
I	II		—	II

This is a Subjectless Suppressive (Mel'čuk 1993: 14, item 4). Such a description treats the Maasai **-ki/-i** form that may not take the agentive NP as being structurally similar to the Spanish *se*-form in (11):

- (11) Sp. *Se construye tres puentes*
 lit. '["It"] builds itself three bridges'. = 'Three bridges are being built'.

No real SSynt-Subject is possible nor is an AgCo; the dummy (= empty zero) SSynt-Subject is a 3rd person singular pronoun.

According to Dimmendaal's description (1983: 132—133), this is what Turkana has (as 'Impersonal Active,' in Dimmendaal's terminology).

As is to be expected, the Subjectless Suppressive can also be formed in Maasai from intransitive verbs. This fact, first established in Perlmutter and Postal 1984: 164, is illustrated in Payne *et al.* 1993: 10—11: $\mathcal{E} + kwet + i$ *metábaykí* 'There-will-be-running tomorrow' (where no NP denoting the agent can be added), $\mathcal{E} + rany + i$ 'There-is-singing', etc.

To sum up: the Maasai verbal form in **-ki/-i** can be described in two mutually exclusive ways. In the speech variety which allows an AgCo (in the oblique case), the **-ki/-i** form is a Partial Demotional Passive, while in those varieties where an AgCo is excluded, it is a Subjectless Suppressive. This difference parallels the difference between the Impersonalized full passive (12a) and the Impersonalized partial (= agentless) passive in French (12b):

- (12) a. *157 ordinateurs [= I] ont été vendus par nos représentants*
 [= **II**] ⇒
 '157 computers have been sold by our representatives'.
Il a été vendu par nos représentants [= II] 157 ordinateurs [= I]
 'It has been sold by our representatives 157 computers'.
- b. *Des ordinateurs personnels [= I] se vendent partout (*par nos représentants [= II])* ⇒
 lit. 'Personal computers sell themselves everywhere (by our representatives)'.

*Il se vend partout des ordinateurs personnels [= I] (*par nos représentants [= II])*

lit. 'It sells itself everywhere personal computers (by our representatives)'.

* * *

I would like to think that my arguments and examples have convincingly shown a good terminology to be extremely important. Just using the case and voice names properly allows for a logical and 'normal' description of Maasai facts, which otherwise undergo anti-intuitive and typologically implausible treatment. "Watch you concepts and your terms!" — such must be the present-day battle-cry of theoretical linguistics.

References

- Aikhenvald, Aleksandra. 1986. On the reconstruction of syntactic system in Berber-Lybic. *Zeitschrift für Phonetik und Kommunikationsforschung*, 39: 5. 1986. P. 527—539.
- Andersen, Torben. 1988. Ergativity in Pāri, a Nilotic OSV language. *Lingua*, 75. 1988. P. 289—324.
- Dimmendaal, Gerrit Jan. 1983. *The Turkana Language*. Dordrecht; Cinnaminson: Floris, 1983.
- Frajzyngier, Zygmunt. 1984a. On the Proto-Chadic syntactic pattern. In: J. Bynon (ed.), *Current Progress in Afro-Asiatic Linguistics (Papers of the Third International Hamito-Semitic Congress)*, Amsterdam/Philadelphia: Benjamins, 1984. P. 139—160.
- Frajzyngier, Zygmunt. 1984b. Ergative and nominative-accusative features in Mandara. *Journal of African Languages and Linguistics*, 6: 1. 1984. P. 35—45.
- Greenberg, Joseph H. 1959. The origin of the Maasai passive. *Africa*, 29: 2. 1959. P. 171—176.
- Heine, Berndt and Ulrike Claudi. 1986. *On the Rise of Grammatical Categories. Some Examples from Maa*. Berlin : D. Reimer [= *Kölner Beiträge zur Afrikanistik*, 13. Band], 1986.
- Keenan, Edward L. 1976. Towards a universal definition of Subject. In: Ch. Li (ed.), *Subject and Topic*, New York: Academic Press, 1976. P. 303—333.
- Mel'čuk, Igor. 1978. Towards a definition of the concept 'Ergative Construction'. In: *Proceedings of the XIIth Intern. Congress of Linguists, Vienna 1977*, Innsbruck: Universität Innsbruck, 1978. P. 384—387.
- Mel'čuk, Igor. 1988. *Dependency Syntax. Theory and Practice*. Albany, N.Y.: The State University of New York Press, 1988.
- Mel'čuk, Igor. 1992. Towards a logical analysis of the notion 'Ergative Construction'. *Studies in Language*, 16: 1. 1992. P. 91—138.

- Mel'čuk, Igor. 1993. The inflectional category of voice: Towards a more rigorous definition. In: B. Com-rie and M. Polinsky (eds), *Causativity and Transitivity*, Amsterdam/Philadelphia: Benjamins, 1993. P. 1—46.
- Mel'čuk, Igor. 1993—2000. *Cours de morphologie générale. Vol. 1—5*. Montréal; Paris: Presses de l'Université de Montréal/CNRS Éditions, 1993—2000.
- Mel'čuk, Igor. 1997. Grammatical cases, basic verbal construction, and voice in Maasai: Towards a better analysis of the concepts. In: W. Dressler, M. Prinzhorn & J. Rennison (eds). *Advances in Morphology*, Berlin; New York: Mouton de Gruyter, 1997. P. 131—170.
- Mel'čuk, Igor. 2004. Opredelenie kategorii zaloga i isčislenie vozmožnyh zalogov: 30 let spustja. In: V. Xrakovskij, A. Mal'čukov & S. Dmitrenko (red.). *40 let Sankt-Peterburgskoj tipologičeskoj škole*. Moskva, Znak, 2004.
- Mel'čuk, Igor. 2005. *Aspects of the Theory of Morphology*. Berlin; New York: Mouton de Gruyter, 2005.
- Payne, Doris, Mitsuyo Hamaya and Peter Jacobs. 1994. Active, inverse, and passive in Maasai. In: T. Givón (ed.), *Voice and Inversion*, Amsterdam/Philadelphia: Benjamins, 1994. P. 283—315.
- Perlmutter, David and Postal, Paul. 1984. Impersonal passives and some relational laws. In: D. Perlmutter, and C. Rosen (eds). *Studies in Relational Grammar 2*. Chicago; London: The University of Chicago Press, 1984. P. 126—170.
- Sasse, Hans-Jürgen. 1984. Case in Cushitic, Semitic and Berber. In: J. Bynon (ed.). *Current Progress in Afro-Asiatic Linguistics (Paper of the Third International Hamito-Semitic Congress)*, Amsterdam/Philadelphia: Benjamins, 1984. P. 111—126.
- Tucker, Archibald and Margaret Bryan. 1966. *Linguistic Analyses. The Non-Bantu Languages of North-Eastern Africa*. London: Oxford University Press, 1966.
- Tucker, Archibald and John Tompe Ole Mpaayei. 1955. *A Maasai Grammar with Vocabulary*. London: Longmans, Green and C°, 1955.

The Russian Conjunction *A* Revisited *

Irina Mikaelian

Pennsylvania State University,
ixm12@psu.edu

Anna A. Zalizniak

Russian Academy of Sciences
anna-zalizniak@mtu-net.ru

The authors refine the semantic description of the conjunction *a* and propose an explanation of both its crucial position within the Russian system of coordination and the absence of its direct equivalents in Western European languages. They suggest that *a* is a language-specific word, which implicitly carries the idea of ‘unpredictability’ — one of the key ideas in the Russian semantic universe, also reflected in a number of other words and expressions.

Еще раз о русском союзе *A*. В работе уточняется семантическое описание союза *a*. Предлагается объяснение его центральной роли в русской системе сочинительных союзов и, одновременно, отсутствия для него единого переводного эквивалента в языках Западной Европы. Утверждается, что союз *a* является лингвоспецифичным словом, имплицитно несущим в себе идею ‘непредсказуемости’, которая присутствует в ряде других лингвоспецифичных слов и выражений русского языка.

1. Preliminary remarks

The conjunction *a* is one of the most thoroughly investigated Russian words (see, in particular, Levin 1970/98, Kreidlin, Padučeva 1974a, b, Sannikov 1989, Yanko 1990, Yokoyama 1990, Padučeva 1997, Nikolaeva 2004, Uryson 2004 a, b, Fougeron 2004). The main types of use of this conjunction have been isolated and described. Correlation with semantically close conjunctions *i* and *no* has been addressed. Most authors have mentioned particular pragmatic and expressive functions of the conjunction *a*, as well as its high frequency in casual colloquial language.

E. V. Padučeva (1997) has demonstrated that the most significant feature of the conjunction *a* which, in particular, opposes it to *no* in some contexts, is that the conjunction *a* reflects the state of the mind of the

* This work was supported by RFBR (grant # 05-06-80215-a).

speaker: his or her expectations, opinions, evaluations or emotions, while the conjunction *no* makes reference to the state of mind of the acting person mentioned in the sentence. We will discuss this feature of the Russian *a* in more detail later.

In our paper, we will refine the semantic description of the conjunction *a* and try to propose an explanation of two interrelated facts: the crucial position of this conjunction within the Russian system of coordination and the absence of its direct equivalents in Western languages.

It is well known, in particular, to teachers of Russian as a foreign language and to translators, and it has been repeatedly mentioned by scholars of Russian language (see, for instance, Kreidlin, Padučeva 1974a: 32, Fougeron 2004: 57) that modern languages of the Western Europe do not have any conjunction that would correspond to the Russian *a*, while the two other main coordinative conjunctions, *i* and *no*, have standard correspondences in all these languages, cf. English *and* and *but*, German *und* and *aber*, French *et* and *mais*, Italian *e* and *ma*, etc.¹.

German has a conjunction *sondern* which only corresponds to Russian *a* in the construction *ne... a*, which is generally considered as a complex conjunction and will not be analyzed here. A similar function of “correction” is attested for the Spanish conjunction *sino* and the Swedish *utan* [Malchukov 2004: 103]. The Latin *at* seems to be a very close equivalent of the Russian *a*: *At tu, Catulle, destinatus, obdura* “But you, Catullus, be resolved and firm.” Cf. also the following example (from Trubačev 2004: 168): *Scipio est bellator, at Marcus Cato — orator*. The Latin-Russian dictionary by Dvoreckij distinguishes two main uses of *at*: “when it is a question of making distinction without opposition” or “with opposition”. These uses are typical of the Russian *a* as well.

Thus the conceptual domain served by three different items in Russian — *i*, *a*, and *no* — is shared by two items in Western languages, so that the Russian *a* can be rendered either by the conjunction ‘and’ or by the conjunction ‘but’, depending on the context. Obviously, each language has some additional lexical means, which allow expressing the meaning corresponding to different uses of *a* (cf. English *while*, *yet*, *instead*, *rather*; French *pourtant*, Italian *mentre*, *invece*). Thus, a Russian sentence *Ja rabotaju, a ty spiš* ‘can be rendered in Italian either as *Io lavoro mentre tu dormi* (“I am working **while** you are sleeping”), where *mentre* ‘while’ insists on the simultaneity of the two situations, or as *Io lavoro, tu, invece, dormi*, where *invece* ‘on the contrary, instead’ corre-

¹ This study does not take into account the data of Slavic languages other than Russian.

sponds to the contrastive semantic element of *a* (see also examples (15), (16), (18) below).

In any case, the specific semantic features of utterances with the conjunction *a* are either lost or distorted. The absence of a standard equivalent for *a* in Western languages creates a still greater problem because this conjunction is a highly frequent Russian word which serves as the basis for a great number of discursive strategies ².

2. The semantics of *A*: a general presentation

Following Krejdlin & Padučeva 1974a,b, we distinguish three main uses of the Russian *a* ³:

(i) juxtaposition of two situations ⁴:

- (1) *У нее отец ирландец, а мать французженка.* “Her father is Irish **and** her mother is French”;
- (2) *Тогда ты молодой был, красивый, а теперь постарел.* “You were handsome and young then, **and** now you are an old man” (Čekhov);

(ii) unexpected (abnormal) effect, or abnormal circumstance ⁵:

- (3) *Октябрь, а солнце что твой август.* “It’s October, **but** the sun is shining as in August” (Pasternak);

² According to the Frequency Dictionary of Russian (Zasorina 1977), the conjunction *a* is twice as frequent as the conjunction *no* (index of frequency for *a* is 10719, while it is 5176 for *no*). The data for the conjunction *i* (its index is 36266) cannot be taken into account directly because they include not only its uses as a clause coordinator, but also cases where *i* functions as an NP-coordinator. The latter is formally excluded for both *a* and *no*.

³ This classification is accepted, with some variations, by most later studies.

⁴ We think that the word *juxtaposition* corresponds better to this use of *a* than the Russian word *sopostavlenie* — the term generally applied by Russian scholars. The closest Russian correspondence of *juxtaposition* would be rather *sopoloženie*. While both Russian words represent perfect equivalents of the English (and French) *juxtaposition* from the formal (word-formational) point of view, *sopostavlenie* comprises an idea of comparison, which is absent in the English (and French) word. We choose the term *juxtaposition* because of its neutral semantics: as we try to show in what follows, “comparison” does not correspond to central uses of the Russian *a* described in section 3.

⁵ The term “abnormal circumstance” has been introduced in Uryson 2004a. Cf. also the term “denial of expectations” in Lakoff 1971; Malchukov 2004: 179.

(4) *Война гуляет по России, / А мы такие молодые!* “The war roams Russia, / **And** we are so young!” (Samojlov);

(iii) additional information and generalization ⁶:

(5) *За рекой был сад, а за садом начинался лес.* “There was a garden beyond the river, **and** beyond the garden, there was a forest”.

(6) *...учитель был для нее род слуги или мастерового, а слуга или мастеровой не казался ей мужчиною.* “...a tutor was to her mind a kind of servant or artisan, **and** a servant or an artisan was not a man in her eyes” (Puškin. “Dubrovskij”).

It is noteworthy that these three semantically distinct uses correlate with formal properties of the respective constructions. These properties can be presented as follows:

(i) *XP, a YQ*; where *X* and *Y* are topics and *P* and *Q* are comments within each clause in the compound sentence. There exists a major restriction on the semantics of these elements: *X* and *Y* should belong to the same set (either natural or situational), while *P* and *Q* must be understood as different values of the same parameter ⁷;

(ii) *P, a Q* where *P* and *Q* represents two different statements;

(iii) *XP, a PQ*, so that the topic of the second proposition reproduces the comment of the first one. This type of using *a* was analyzed in Uryson 2002. It has a marginal connection with the problem discussed in our paper and will not be considered in what follows.

3. Three types of juxtaposition

Let us analyze in more detail the case of juxtaposition. We will distinguish three types of contexts where this meaning manifests itself.

1. A juxtaposition which reveals an *o p p o s i t i o n* ⁸:

(7) *На голове у него была меховая шапка, а на ногах — кеды.* “He wore a fur hat (on the head) and gym-shoes (on the feet).”

⁶ E.Uryson (2002) calls the use of *a* illustrated by ex. (5) “turning of the narration”; Sannikov (1989) distinguishes “generalization” (ex. 6) as a separate meaning. We consider these two uses as context dependent variants.

⁷ For details, see Krejdlin & Padučeva 1974b.

⁸ Some scholars consider this use of *a* as a separate meaning or lexeme. Nikolaeva (2004) calls it “contrastive juxtaposition” (*kontrastnoe sopostavlenie*). According to Uryson (2004b), it is a lexeme carrying a “comparison *stricto sensu*” (*sobstvenno sravnenie*).

- (8) *Без них обедали всегда в первом часу, как везде у людей, а при них в седьмом.* “Before they came, we used to have dinner at noon, like everybody else, **but** now we have it at seven in the afternoon” (Čekhov).

2. A juxtaposition which reveals similarity; more exactly, it presents two situations as complementary. The best way to account for this use of *a* would be the Russian expression *Vsjo odno k одному* “Everything fits a certain pattern”):

- (9) *На голове у него была меховая шапка-ушанка, а на ногах — валенки.* “He wore a fur hat (on the head) and valenki (on the feet).”
- (10) *Но от сигары во рту стало горько и противно, а вино оказалось не такого вкуса, как в прошлом году.* “But the cigar left a bitter and disgusting taste in his mouth, **and** the wine had not the same flavour as it had had the year before.” (Čekhov)
- (11) *Соседы говорили о свадьбе, как о деле уже конченом, а добрая Прасковья Петровна радовалась, что дочь её наконец нашла себе достойного жениха.* “The neighbors talked of the wedding as though everything had been settled, **and** the good Praskovja Petrovna rejoiced that her daughter had at last found a suitor worthy of her.” (Puškin)
- (12) *Написать записку по-французски казалось ей неуместным, а в русской орфографии сама матушка не была сильна.* “It seemed to my mother inappropriate to write a note in French, **and** she wasn’t strong enough in Russian orthography.” (Turgenev) [Both solutions of the problem are unacceptable]

3. A juxtaposition that does not convey either similarity or opposition. This is the most common case: any simple enumeration of elements with their respective properties allows the use of the conjunction *a* in Russian and excludes the use of conjunctions *i* and *no*. At the same time, depending on circumstances, these sentences can express both meanings (similarity or opposition):

- (13) *У нее отец физик, а мать филолог.* “Her father is a physicist **and** her mother is a philologist”.
- (14) *Маша делает уроки, а Петя читает книгу.* “Masha is doing her homework **and** Petja is reading a book”.

In this way, *a* comprises the ideas of both distinction and similarity. But the distinction and the similarity do not have the same status. The

distinction constitutes a formal requirement such that a sentence like **Petja umnyj, a Vasja umnyj* (were P and Q formally coincide) is ungrammatical. Regarding the similarity (between juxtaposed items on the one hand and between their respective properties on the other hand), it constitutes the condition of a semantically correct usage of utterances with the conjunction *a*, as well as of its adequate interpretation by the recipient. The notion of similarity should be understood in a broad way, as a logical or situational link: sentences comprising *a* of juxtaposition answer an explicit or implicit question that has a generic meaning with respect to both clauses of a compound sentence with *a*, cf.: *Kto po porfessii ee roditeli?*; *Čem zanimajutsja deti?* The knowledge of this condition allows producing and understanding jokes like the famous pun by a Russian politician *U menja mat' russkaja, a otec jurist* ‘My mother is Russian, and my father is a lawyer’, where the name of a profession is reinterpreted as the name of a nationality.⁹ At the same time, the saying *V ogorode buzina, a v Kieve djad'ka* illustrates the absurdity of a grammatically correct utterance, which does not satisfy semantic restrictions on the construction with *a*: neither the topics (*v ogorode* and *v Kieve*) nor the comments (*buzina* and *djad'ka*) reveal sufficient logical link between them.

However, the conjunction *a* does not express either opposition or similarity directly. A clear contrast appears in cases where this meaning is already present in the semantics of the juxtaposed lexical units, cf.: *young — old, stupid — smart, to work hard — to do nothing*, etc. In other cases, similarity, contrast, or neutral interpretation of the utterance depend on the situation and on the intention of the speaker. Thus, example (13), depending on the intention of the speaker, may mean: ‘both parents of the person in question are intellectuals’ (and moreover, they are mutually complementary). It may also mean: ‘the father and the mother of this person have quite different and, in a sense, opposite professions’ (a contrast). Finally, it may be neutral and carry plain information about the professions of the person’s parents.

4. Russian conjunction *A* as a language-specific word

The translation of constructions with the conjunction *a* into languages that do not have its direct counterpart requires a rather subtle semantic analysis of both the context and the situation in which the utterance is

⁹ The full understanding of this pun also requires a specific extra-linguistic knowledge.

made. As a general rule, one can translate them using the conjunctions ‘and’ (see examples (1), (2), (5), (6), (10)-(14)) or ‘but’ (see examples (3), (8)), although translation by the conjunction ‘but’ is in most cases unsatisfactory (see below). The third possibility is to use another lexical unit or whatever appropriate linguistic device (e.g. intonation, word order), cf.:

- (15) *Коврин говорил ласково и убедительно, а она продолжала плакать.* “Kovrin talked affectionately and persuasively, **while** she went on crying” (Čekhov).
 (16) *Надобно признать, что я несчастлив: играю мирандолом, никогда не горячусь, ничем меня с толку не собьешь, а все проигрываюсь.* “I must confess that I am unlucky: I play mirandole, I always keep cool, I never allow anything to put me out, **and yet** I always lose!” (Puškin. “The dame of spades”)

Cf. also the French translation for (16):

“Vraiment, je n’ai pas de veine. Je ne double jamais ma mise. Rien ne me démonte. **Pourtant**, je perds toujours”

The semantics of the Russian *a* manifests itself most clearly in contexts where it competes with the conjunction *no*, and this is the case in the contexts of unexpected effect. Let us consider this variation in more detail.

As mentioned earlier, the conjunction *a* expresses attitudes and evaluations by the speaker, while *no* reflects the state of mind of a person mentioned in the sentence. Consider examples (17) and (18)¹⁰.

- (17) *День был дождливый, но Коля не взял зонт.* “The day was rainy, **but** Kolja did not take his umbrella”.
 (18) *День был дождливый, а Коля не взял зонт.* “The day was rainy, **and** (that was especially bad, since) Kolja **had not taken** his umbrella.”

The meanings of (17) and (18) are considerably different. The conjunction *no* in (17) links two contradictory events — the rain and Kolja’s action — in one complex situation. The contradiction is presented as internal to this situation, such that the events of the two clauses should be understood as consecutive: the rain had started before Kolja’s action, and Kolja was aware of it when he did not take the umbrella, cf.: *Kolja uvidel, čto idet dožd’, no (nesmotrja na eto) ne vzjal zont* — “Kolja saw that it was rainy, but (nevertheless) did not take his umbrella”. Thus, Kolja’s action is presented as contradicting the normal cause-effect implication of

¹⁰ These examples have been analyzed in [Sannikov 1989], [Uryson 2004a,b].

the first clause¹¹. The conjunction *no* not only reveals this contradiction, but also implies an explanation that would restore the normal logic of events, for instance: Kolja did not take the umbrella because he had a raincoat¹².

By contrast, the conjunction *a* simply juxtaposes two contradictory facts, but does not point to any logical link between them. At the same time, the speaker is more directly involved: it doesn't matter whether Kolja knew or not that the day was rainy; it is the speaker who confronts the two facts: the rain and the fact that Kolja does not have an umbrella, and it is the speaker who presents them as an unfortunate happenstance.

One of the possible translations into English proposed for example (18) (see above) reflects the same idea: The past perfect tense (*Kolja had not taken his umbrella*) focuses on the resulting situation ("Kolja does not have the umbrella"). It also suggests that Kolja did not know that it would be raining and the issue whether or not to take an umbrella did not arise for him. At the same time, the English translation is more explicit than the Russian sentence, which is vague on why it so happened that Kolja did not take an umbrella. If the context clearly suggests that the rain had started before Kolja left so that Kolja knew about it and acted consciously, the conjunction *a* indicates that the speaker considers Kolja's action as unexplainable and inconsistent. In this case, the sentence (18) could be translated as *The day was rainy, and Kolja still did not take the umbrella*.

The difference between the conjunctions *a* and *no* can be formulated in a more general way. The conjunction *no* cannot link two independent events [Levin 1970/98: 802, Sannikov 1989: 170]. Thus, the use of *no* is inappropriate in *Idet dožd', a / *no Kolja ne vzjal zont* "It rains and / *but Kolja has not taken an umbrella": the second clause introduces an action preceding the rain and thus independent of it. In contrast, the conjunction *a* presents two events as independent facts, even if they follow each other in time, cf.:

- (19) а. Петя обещал прийти, **но** не пришел. "Petja promised to come but didn't come."

¹¹ If two consecutive events are linked by a normal cause-effect relation, only the conjunction *i* can be used: (*Kolja uvidel, čto idet dožd', i vzjal zont* — "Kolja saw that it was rainy and took the umbrella").

¹² The study [Uryson 2004a] gives sufficient evidences proving that the semantic configuration of the conjunction *no* of "abnormal effect" always comprises a semantic actant corresponding to Cause.

- b. Петя обещал прийти, *a* не пришел. “Petja promised to come and hasn’t come”.

Example (19a) presents two consecutive actions of the same person. The second action contradicts the first one: the Agent acted, for some reasons, not as expected. Example (19b) presents the same situation differently: the speaker considers two facts — Petja’s earlier promise and his absence in the place where the speaker expected to see him — and expresses his or her disappointment. It is remarkable with this respect that the conjunction *a* cannot link two consecutive events if they follow each other immediately. Consider example (20):

- (20) “...Ни скучных песен, ни пиров — / Не стану есть, не буду слушать, / Умру среди твоих садов!” / Подумала — и стала кушать. (Пушкин. Руслан и Людмила) “... No boring songs, no feasts — / I won’t eat, I won’t listen, / I would rather die/In your gardens / That’s how she thought — and (immediately) began to eat.

The action by the Agent contradicts her decision made immediately before, and thus appears as inconsistent. The conjunction *no* does not fit here because it would present the action as resulting from a conscious decision of the Agent, which is not the case. The conjunction *a* cannot be used in this context because it would destroy the dynamic of the narration. In these situations, the only conjunction to be used in Russian is the conjunction *i*¹³. Consider also examples (21a) and (21b):

- (21) a. Она такая толстая, *a* ест пирожные. “She is so fat and (still) eats cakes!”
 b. *Она подумала “Какая же я толстая!” — *a* съела пирожное. “She thought: How fat I am! — and (immediately) ate a cake”.

Example (21a) presents a static opposition of facts by the speaker (the appreciation “she is fat” also belongs to the speaker and not to the Agent), and the use of *a* is quite appropriate here. In contrast, *a* is impossible in example (21b) where two consecutive actions are related.

Let us consider one more example taken from Uryson 2004a:

- (22) a. Иван развелся с женой, *a* на Кате не женился. “Ivan divorced his wife and (still) did not marry Katja.”

¹³ This use of *i* can be defined as “*i* of anti-consequence”. It is characterized by a specific intonation (which is marked by a dash preceding *i* in written texts).

- b. *Иван развелся с женой, но на Кате не женился.* “Ivan divorced his wife, **but** he didn’t marry Katja.”

The meaning of both sentences includes the presupposition by the speaker that Ivan should have married Katja after his divorce. Both sentences assert that this did not happen. The sentence (22b) with *no* implies that Ivan had planned to marry Katja, but did not for some reasons. By contrast, the sentence (22a) is vague as to whether Ivan had planned to marry Katja or not: the idea that Ivan should have married her and thus the idea of his contradictory behavior may exclusively belong to the speaker.

Consider also example (4) reproduced below where the contradiction is entirely a result of speaker’s subjective view of the world, so that only the conjunction *a* can be used:

- (4) *Война гуляет по России, / А мы такие молодые!* “The war roams Russia, / And we are so young!” (Samojlov);

Using the conjunction *a*, the speaker presents the situation as unhappy, hardly explainable and, probably, not needing an explanation¹⁴. In other words, it just happened.

To sum up, *a* of unexpected/abnormal effect carries the following ideas: (i) the speaker considers two contradictory facts; (ii) the speaker does not know how to explain this contradiction; (iii) the speaker feels bad/strange about it; (iv) the speaker does not look for an explanation: he or she accepts the contradiction as a normal state of affairs. In other words, the speaker accepts the unpredictability of the world. It should be underlined that all these ideas remain implicit.

The conjunction *a* in all its uses appears as the most “vague” coordinative conjunction in Russian. It does not imply any particular objective relation (temporal or logical) between two confronted facts. It rather functions as a meta-textual operator and can express, in an implicit way, a range of specific attitudes. We suggest that Russian conjunction *a* is a language-specific grammatical word: it reflects a specific conceptual configuration which is also present in a number of other language-specific words and constructions analyzed by A. Wierzbicka and others authors. This configuration can be explicitly rendered by the formula “The world is unpredictable”.

¹⁴ Uryson proposed to introduce a semantic element ‘the speaker does not know the cause [of the situation]’ into the semantic configuration of the conjunction *a* of ‘abnormal effect’ [Uryson 2004a]. We consider that, by using *a*, the speaker is not at all concerned with a cause (or a rational explanation) of the situation.

We think that the idea of “unpredictability” can be applied not only to *a* of unexpected effect, but also to *a* of juxtaposition, though it usually functions on a different semiotic level. Thus, as *a* of unexpected effect, *a* of juxtaposition also confronts two facts. The result of this confrontation is *unpredictable* and belongs to the speaker: the speaker can present two facts as contradictory or complementary, he can also restrain himself from drawing any conclusion. On the other hand, using *a* of juxtaposition, the speaker can also express his emotional attitude towards events that happen, cf. *Doč' blondinka, a syn brjunet! (Kak stranno! Kak interesno! Vot ved' kak byvaet!)* “The daughter is blond, and the sun is dark-haired! (How strange! How interesting! It happens!)”¹⁵.

The lack of space does not allow us to consider other uses of *a*, such as *a* of additional information, initial *a*, or the particle *a*, but we think that all these uses have important common semantic features, one of which is the idea of ‘unpredictability’.

References:

- Kreidlin, G. E., Padučeva E. V. (1974a) ‘Značenie i sintaksičeskie svojstva sojuza *a*’ In: NTI. Ser.2, №9, 31—37.
- Kreidlin, G. E., Padučeva E. V. (1974b) ‘Vzaimodejstvie asociativnyx svjazej i aktual'nogo členenija v predloženijs s sojuzom *a*’ In: NTI. Ser. 2, № 10, 32—37.
- Lakoff, R. (1971) ‘If's, and's and but's about conjunctions’ In: Ch. Fillmore & T. Langendoen (eds), *Studies in Linguistic Semantics*. New York, 114—149.
- Levin, Ju. I. (1970/98) ‘Ob odnoj gruppe sojuzov russkogo jazyka’ In: Ju. I. Levin *Izbrannye trudy*. Moskva, 789—808.
- Malchukov, A. (2004) ‘Towards a Semantic Typology of Adversative and Contrast Marking’ In: *Journal of Semantics*, 21: 177—198.
- Nikolaeva, T. M. (2004) ‘Sočinitel'nye sojuzy *a*, *no*, *i*: istorija, sxodstva i različija’ In: *Verbal'naja i neverbal'naja opory prostranstva mežfrazovyx svjazej*. Moskva, 13—46.
- Padučeva, E. V. (1997) ‘Egocentričeskaja semantika sojuzov *A* i *NO*’ In: *Slavyanskije sočinitel'nye sojuzy*. Moskva, 36—47.
- Sannikov, V. Z. (1989) *Russkie sočinitel'nye konstrukcii: semantika, pragmatika, sintaksis*. Moskva.

¹⁵ It should be mentioned that the frontier between the two uses of *a* (*a* of juxtaposition and *a* of unexpected effect) is not clear-cut. The “pure” case of juxtaposition reveals itself in static, atemporal statements (*Mat' russkaja, a otec francuz*). But it is more difficult to classify a sentence like *Tom prosnulsja, a Tim ešče spal* “Tom woke up, but/and Tim was still asleep”, which presents features of both uses.

- Trubačev, O. N. (2004) 'K sravnitel'no-etimologičeskoj xarakteristike sojuza *a* i sočetanij s nim v praslavjanskom' In: O. N. Trubačev. *Trudy po etimologii*. T. 2. Moskva, 166—170.
- Uryson, E. V. (2002) 'Sojuz *A* kak signal "povorota povestvovanija"' In: *Logičeskij analiz jazyka. Semantika načala i konca*. Moskva, 348—357.
- Uryson, E. V. (2004a) 'Sojuzy *A* i *NO* i figura govoryaščego' In: *Voprosy jazykoznanija*, № 6, 2004, 64—83.
- Uryson, E. V. (2004b) 'Nekotorye značenija sojuza *A* v svete sovremennoj semantičeskoj teorii' In: *Ruskij jazyk v naučnom osveščanii*, N 2 (8).
- Yanko, T. E. (1990) 'Ešče raz o sojuzax *A* i *NO*' In: *Logičeskij analiz jazyka. Protivorečivost' i anomalnost' teksta*. Moskva, 246—258.
- Yokoyama (1990) — O. Yokoyama. 'K analify russkix sočinitel'nyx sojuzov *A* i *NO*' In: *Logičeskij analiz jazyka. Protivorečivost' i anomalnost' teksta*. Moskva, 190—193.
- Zasorina L. N. (1977) *Častotnyj slovar' russkogo jazyka*. Moskva.

Semantic Paraphrasing in Meaning-Text Linguistic Models

Jasmina Milićević

French Department, Dalhousie University & OLST, Université de Montréal
jmilicev@dal.ca

The paper deals with paraphrasing rules [= rules allowing for production of paraphrases, i.e. semantically equivalent sentences] in Meaning-Text linguistic models. It focuses on semantic paraphrasing rules, a new type of rules based on semantic decomposition of lexical meanings, necessary to treat certain approximate paraphrases which cannot be adequately treated by the existing lexical-syntactic paraphrasing rules.

Семантическое перифразирование в лингвистических моделях типа «Смысл — Текст». Рассматриваются правила перифразирования [= правила производства перифраз, т. е. синонимичных фраз] в рамках модели Смысл — Текст. В центре внимания — семантические правила перифразирования; это правила нового типа, основанные на семантических разложениях лексических смыслов и необходимые для описания некоторых прилительных перифраз, которые не могут быть адекватно заданы с помощью существующих лексико-синтаксических правил перифразирования.

1. Paraphrase and its modeling in linguistics

Paraphrase is a relation between synonymous sentences, i.e., sentences having (nearly) the same meaning, like sentences in (1):

- (1) a. *Alberta crop crisis causes a sharp increase in wheat prices.*
b. *Wheat prices increase sharply in the wake of Alberta crop crisis.*

Sentences between which this relation holds are called *paraphrases* (of each other).

Studying paraphrases means studying either their production or their recognition. Production of paraphrases [= paraphrasing] is an operation consisting in generating paraphrases from a common representation or

explicitly indicating paraphrastic links between their representations. Recognition of paraphrases is an operation which consists in ‘reducing’ paraphrases to a common representation or in ‘uncovering’ paraphrastic links between their representations. This paper deals exclusively with paraphrasing.

Paraphrasing is involved in an essential way in everyday linguistic activity. On the one hand, paraphrasing is needed in order to be able to speak fluently. This is so because of numerous pitfalls — in the form of restricted lexical co-occurrence, word-order constraints, lexical holes, defective paradigms, etc., awaiting anyone who is trying to utter a sentence. In order to be able to avoid a dead-end in sentence production, the speaker vitally needs spare paraphrastic variants. Thus, speaking can be thought of as *virtual paraphrasing* — a process involving multiple choices between possible paraphrases¹. On the other hand, paraphrasing is needed in order to speak/write better: to make oneself clear, to avoid unwanted repetitions, to produce a particular artistic effect, etc. This conscious paraphrasing, by which the speaker modifies an existing sentence, can be termed *reformulative paraphrasing*; it is exploited namely in complex linguistic tasks such as writing, revision and translation ($\approx$ interlinguistic paraphrasing).

A corollary of this is that a linguistic model of a language L should adequately reflect the paraphrasing competence of native speakers of L . That is, a model of L should include as an integral part a model of paraphrasing.

|| To model paraphrasing means to simulate/approximate the speaker’s paraphrasing abilities by rules stated in a formal language).

Assuming a stratificational model of language, there can be two types of linguistic rules serving this purpose. On the one hand, we can use *correspondence rules*, which are ‘normal’ synthesis rules, not specifically designed for paraphrasing, and get paraphrases by multiple synthesis out of an underlying meaning representation. On the other hand, we can use rules specifically designed for paraphrasing, called *equivalence rules*, which state equivalencies between sentences at a given level of representation. Rules of the first type model virtual paraphrasing and are paraphrasing rules only in a broad sense. Rules of the second type model re-

¹ It seems indeed that a latent, unconscious paraphrasing activity is constantly going on during speech production. It is sometimes revealed by performance errors, such as, for instance, ‘double generation’ (*to enable them to be able* = a mix of *to enable* and its semantic decomposition into *to cause to be able*) and contamination (**do an agreement* = the use of light verb DO, a collocative of a quasi-synonymous DEAL, with AGREEMENT).

formulative paraphrasing and are paraphrasing rules proper. They constitute a *paraphrasing system* (of *L*).

In the remainder of the paper, I will present the Meaning-Text paraphrasing system.

2. Paraphrasing in Meaning-Text Models

A model of a language *L* based on the Meaning-Text theory (Mel'čuk 1974), called *Meaning-Text Model* [= MTM] simulates speech-production capabilities of native speakers of *L*. An MTM presupposes four major levels of representation of utterances: semantic, syntactic, morphological, and phonological. Each representation level except semantic level is further split into deep- and surface- sublevels, the deep sublevel being oriented towards the meaning and the surface one towards the form of expression. A representation at each level is a set of structures, comprising a basic structure, which characterizes the main linguistic entity described at that level (semantemes, lexemes/syntactic constructions, wordforms, phonemes), and a number of peripheral structures, supplying additional characterizations (communicative, prosodic, etc.) of the basic structure. The levels of representation particularly interesting for paraphrasing are the semantic level and the deep-syntactic level. The basic formalism used at the semantic level of representation is the semantic network (see, for instance, Polguère 1998), and the one used at the deep-syntactic level is the dependency tree (see Mel'čuk 1988).

An MTM includes a built-in paraphrasing component: a set of equivalence rules called *lexical-syntactic paraphrasing rules* (introduced in Žolkovskij & Mel'čuk 1967 and Mel'čuk 1974, revised and augmented in Apresjan 1974: 324—334, Iomdin 1972, Mel'čuk 1992 and Milićević 2003: 177—247). These rules state equivalencies between lexical-syntactic configurations at the deep-syntactic level of representation. Two types of rules are distinguished. Lexical rules carry out (quasi-)synonymous lexical substitutions in the deep-syntactic tree. They are based on lexical relations of (exact and quasi-) synonymy, antonymy, conversion and semantic derivation (situation name, agent name, patient name, etc.) and are formulated in terms of *lexical functions* (see, for instance, Žolkovskij & Mel'čuk 1967 and Wanner 1996), a cross-linguistically valid formal tools for describing lexical relations. Syntactic (restructuring) rules perform structural rearrangements in the tree, triggered by lexical substitutions. They make use of a few elementary syntactic operations (fission of a tree node in two/fusion of two tree nodes in one, transfer of a tree branch, relabeling of branches) and combinations thereof.

Paraphrases in (1) illustrate a kind of paraphrases that can be treated by lexical-syntactic rules. They manifest a particular case of the equivalence 'X causes Y' ~ '[there is] Y because of X'. Both are all-rhematic sentences, communicatively equivalent for an underlying question of type *What's new on financial markets?* Their Deep-Syntactic Structures [= DSyntS] are shown in Figures 1 and 2, respectively.

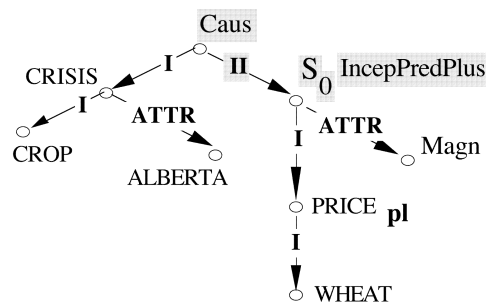


Fig. 1. DSyntS of (1a)

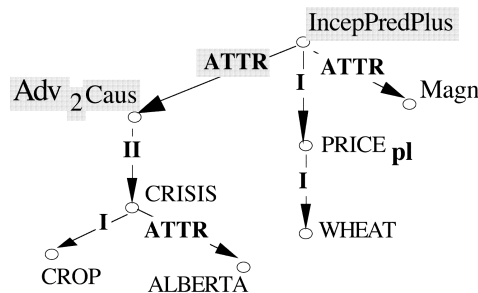


Fig. 2. DSyntS of (1b)

Lexicographic information needed to link the above DSyntSs (encoded in terms of lexical functions, which will not be explained here) is as follows.

Lexical functions of PRICE:

- 1) $S_0\text{IncepPredPlus}$: *increase*_(N); 2) IncepPredPlus : *go up, increase*_(Vintr), *rise*; 3) $\text{Magn}(S_0\text{IncepPredPlus})$: *large, substantial, sharp*; 4) $\text{Magn}(\text{IncepPredPlus})$: *substantially, sharply*².

² Lexical functions Nos 3 and 4 are second-order functions, i.e. lexical functions of lexical functions (of the keyword). The way of encoding such functions hasn't been well worked out as yet, but the issue is beyond the scope of this paper.

Two more generally applicable lexical functions:

5) Caus(L_[N]): *cause, set off, spur*; 6) Adv₂Caus: *because (of), due (to), owing (to), in the wake (of)*.

Based on the above information, the lexical-syntactic rule $L \equiv \text{Adv}_2(L)$ carries out the substitution $\text{Caus} \equiv \text{Adv}_2(\text{Caus})$. This lexical replacement triggers a syntactic restructuring rule which performs head-switching (inversion of subordination):

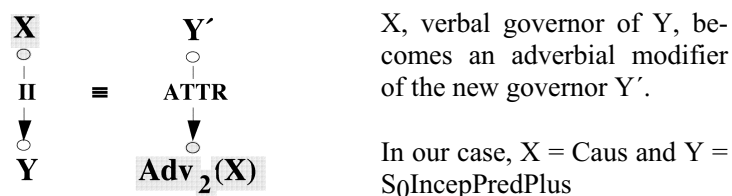


Fig. 3. Deep-Syntactic Restructuring Rule-1

The head-switching rule is accompanied by an automatic transposition $V \equiv S_0(V)$, which turns S₀IncepPredPlus [Y] into IncepPredPlus [Y'], as we need a verb in the position of the top node of the DSynt-tree. (The rule is applied from right to left.)

3. Limits of Deep-Syntactic Paraphrasing

Lexical-syntactic paraphrasing rules are not capable of treating all types of paraphrases. Consider, for instance, the paraphrases in (2):

- (2) a. *Smoking increases the risk of cancer.*
 b. *For smokers, the risk of cancer is higher.*
 c. *In smokers, incidence of cancer is higher.*

Sentence (2a) is about probable causation of a state of affairs (i.e. cancer), sentence (2b) states something about the probability of this state of affairs, while sentence (2c) speaks of the actual existence of this state of affairs. Thus, they differ in two respects: presence vs. absence of causation [(2a) vs. (2b/c)] and presenting the situation as potential vs. actual [(2a/b) vs. (2c)]. In all sentences, the verb is in the habitual present and they are all generic statements, predicating something that is true of a class of individuals (smokers). Sentences in (2) are communicatively equivalent: all can answer an underlying question such as *What are the consequences of smoking?*

These sentences are fairly close paraphrases of each other. Yet, paraphrastic links between them cannot be described in lexical terms, i.e., as lexical equivalencies; cf. their respective DSyntSs in Figures 4—6.

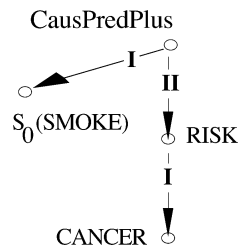


Fig. 4. DSyntS of (2a)

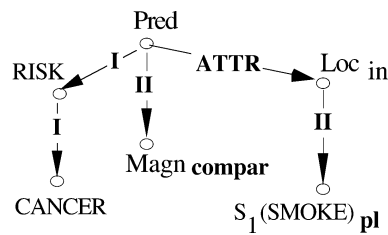


Fig. 5. DSyntS of (2b)

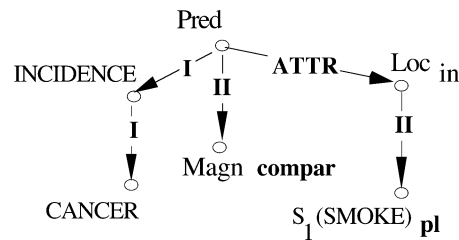


Fig. 6. DSyntS of (2c)

Sentence (2a) features the causative verb *increase* and (2b) the construction “copula + adjective in the comparative degree”, *be higher*, resulting from the omission of causation. In (2a), we have the action noun *smoking* and in (2b/c) the agentive noun *smokers* (= ‘people who smoke’). Finally, (2a/b) contain *risk* ($\approx$ ‘(high) probability of something bad happening to someone’) and (2c) *incidence* ($\approx$ ‘frequency of happening (of something)’).

Stating paraphrastic links between these sentences requires semantic decomposition of lexical meanings they contain and restructuring of their semantic content. Thus, they have to be related at the semantic level of representation. To accomplish this, a new type of paraphrasing rules is needed — *semantic paraphrasing rules* (Milićević 2003: 127—177 and to appear)³.

³ Semantic paraphrasing rules were foreseen already in Mel’čuk (1974: 64 and 73), but no attempt has been made since to actually write them, due to a high degree of difficulty involved in the task.

4. Semantic Approach to Paraphrasing

Semantic paraphrasing rules state equivalencies between configurations of semantemes (at the semantic level of representation of utterances). They are of two major types. Propositional rules carry out meaning-preserving reductions/expansions of the semantic network, perform additions/omissions of semantemes and restructuring of semantic links. Communicative rules are responsible for meaning-preserving reorganizations of the communicative structure.

Let me show how semantic equivalence rules are used to relate sentences in (2). The full-fledged Semantic Structure of (2a) and (2b), i.e., showing all redundant links not expressed in actual sentences, is given in Figure 7. Semantic configurations needed for paraphrasing (in this particular case) are boldfaced.

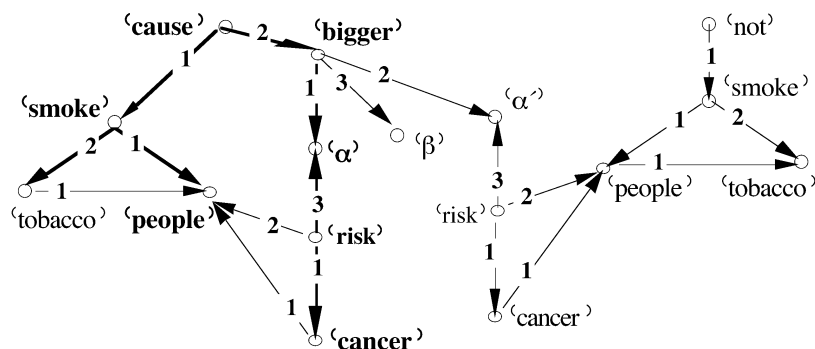


Fig. 7. SemS of (2a) and (2b)

This SemS can be read verbatim as follows: 'smoking tobacco by people causes α , the risk of cancer in people who smoke tobacco, to be bigger by β than α ', the risk of cancer in people who do not smoke tobacco.'

Partial SemRs [SemS + Sem-CommS] of (2a) and (2b) are given in Figures 8 and 9 (a shaded area indicates a lexicalization zone: a semantic configuration to be expressed as a single lexical unit; T1/R1 stand for primary Theme/Rheme, and T2/R2 for secondary Theme/Rheme. The underlined semanteme within a Thematic-Rhematic area is the communicatively dominant node of this area). Their respective literal readings are 'smoking [tobacco by people] increases the risk of cancer [for people who smoke tobacco]' and 'for smokers [of tobacco] the risk of cancer is higher [because of their smoking of tobacco]'.

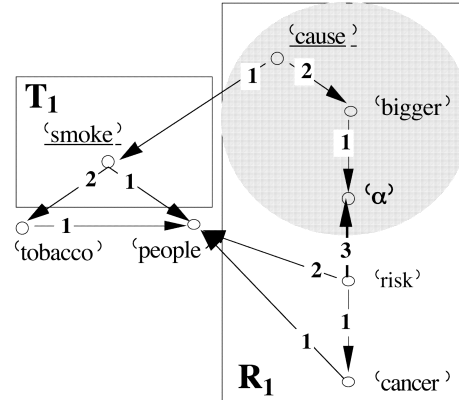


Fig. 8. SemR of (2a)

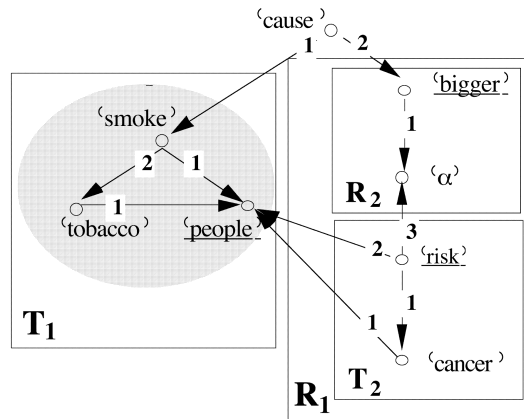


Fig. 9. SemR of (2b)

Different (but equivalent) communicative structures that we see in Figures 8 and 9 impose different lexicalizations of the identical semantic material, including different ellipses of redundant links. (This cannot be done at the DSynt-level because we do not keep redundant links in the DSynt-Structure.)

Instead of keeping the unused semantic links, we can postulate an implication rule which will remove the semanteme 'cause' from the SemS of (2a); here is the rule, stated informally:

Semantic quasi-equivalence rule-1
 'X causes Y (which is Z)' $\cong$ 'there is Y (which is Z)'

Sentences (2a/b) and (2c) cannot be traced back to the same SemS. Compare the SemR of (2b) with that of (2c), given in Figure 10 below.

The two SemRs have minimally different SemSs ('risk' vs. 'frequency') and identical Sem-CommSs. The SemR of (2c) reads literally as follows: 'in smokers [of tobacco] the incidence of cancer is higher [because of their smoking of tobacco].' These SemSs can be related by the following semantic quasi-equivalence rule (stated informally):

Semantic quasi-equivalence rule-2

'[there is] a high probability of X' $\cong$ 'X [happens] often'

The application of this rule requires a prior decomposition of the semanteme 'risk' (whose dominant component is '(high) probability').

The above rule is an inference, valid in habitual context only; cf. *He may come tonight* $\neq$ *He comes often*. It involves some common-sense reasoning, based on the speaker's knowledge of the world (*reality postulates* of Žolkovskij 1964 and *naive physics* of Apresjan 1974 and 2000). Thus, a rule of this type is partly outside of the linguistic system. This is only to be expected, since it links sentences that cannot be assigned identical SemSs⁴.

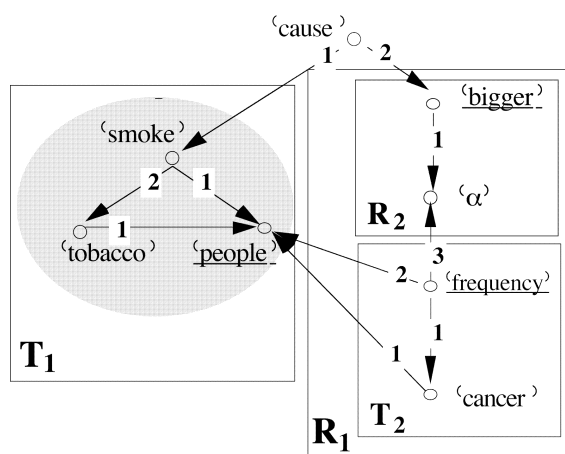


Fig. 10. SemR of (2c)

⁴ Deeper-level representations of these sentences, which we might call *conceptual*, are, presumably, identical. Since the problem of writing conceptual representations of sentences is a serious one and has yet to be tackled within the Meaning-Text theory, I will refrain from speculations on the topic. Note, however, that a step in this direction is vital if we are to account for paraphrases such as those in (2), which are fairly frequent.

Semantic equivalence rules are still pretty much an uncharted territory. To begin with, we do not even know their possible number, although my guess is that it will be in the hundreds. (What we do know, though, is that they are universal, i.e., cross-linguistically valid, just as deep-syntactic equivalencies are.) Their elaboration will require substantial efforts going well beyond what has been done so far, because of the extreme complexity of the phenomena involved and perhaps also because of the insufficient degree of development of our conceptual apparatus.

To show the magnitude of the problem, here is another example of paraphrases which look innocuous enough but whose description requires a number of subtle semantic equivalence rules.

- (3) a. *John's comfortable income enables him to travel a lot.*
 b. *John has a comfortable income so he can travel a lot.*
 c. *With his income, John can afford to travel a lot.*

These paraphrases display the following propositional differences.

(3a/c) vs. (3b).

— In (3a/c), we have the causation of a state of affairs, namely, having a comfortable income causes John to be able to travel, while in (3b) John's being able to travel is presented as a consequence of his having a comfortable income. The causation is expressed in (3a) as a part of the meaning of the verb ENABLE ('cause [someone] to be able to P'), while in (3c) it is expressed by the preposition WITH ('having [N]', 'since [someone] has [N]'); these expressions of causation are communicatively conditioned (see below). As for the consequence, it is expressed in (3b) as SO, which is just one of its possible lexicalizations, along with THEREFORE, THUS, CONSEQUENTLY, etc.⁵

(3a/b) vs. (3c).

— While (3a/b) state explicitly that the amount of money talked about is big, (3c) does not: *with his income* can also be used to refer to a small amount; cf. *With his income, John cannot afford to travel a lot.* The ellipsis of the qualifier *comfortable* is possible in (3c) because the expression being qualified, i.e. *income*, is communicatively given (see below), and is within the scope of causation (it is the ASem 1 of 'cause'). This kind of ellipsis is possible with (3a) as well, where *comfortable income* also meets both of the above conditions (*John's income enables him to travel a lot*), but not with (3b), where it fails to meet the first one (*John has an*

⁵ Almost any semanteme can be expressed in many different ways, but I cannot develop this point here. Note only that semantic paraphrasing rules need to be backed up by elaborate lexicalization rules. Two such rules, involved in the production of sentences in (3), will be briefly mentioned at the very end of the paper.

income so he is able to travel a lot). Cf. an example where the qualifier cannot be omitted without the loss of information because the qualifying expression is not in the scope of causation (even though it is communicatively given): *John's comfortable income comes from an inheritance (is a mystery to me)* $\neq$ *John's income comes from an inheritance (is a mystery to me)*. Note that both *John's comfortable income* (3a) and *John has a comfortable income* (3b) entail 'John has an income'; thus, omitting the qualifier *comfortable* in (3c) amounts to expressing an entailment of (3a/b). (3b) vs. (3c).

— Expression *John can travel* (3b) and *John can afford to travel* (3c) are equivalent in this particular context, where the cause of John's capacity to travel is his having (enough) money, but they are not mutually substitutable in all contexts; cf., for instance, *With the advances in technology, people can write computer programs faster* $\neq$ *With the advances in technology, people can afford to write computer programs faster*.

Sentences (3a-c) are communicatively equivalent for an underlying question such as *What about John's lifestyle?*, but they have different Sem-CommSs.

— Sentence (3a) has only a Primary Rheme, with the semanteme 'cause' as the communicatively dominant node. Sentences (3b) and (3c) have secondary Themes/Rhemes within the primary Rheme; the dominant nodes of these secondary communicative areas are 'big' and 'consequence' for (3b) and 'cause' and 'able' for (3c).

— Expressions *John's comfortable income* (3a) and *with his income* (3c) are given (i.e., convey the information taken to be present in the mind of the Addressee at the moment of utterance), while *John has a comfortable income* (3b) is new (not taken to be present in the mind of the Addressee).

The (approximate) SemRs of sentences in (3) are shown, respectively, in Figures 11, 12 and (13).

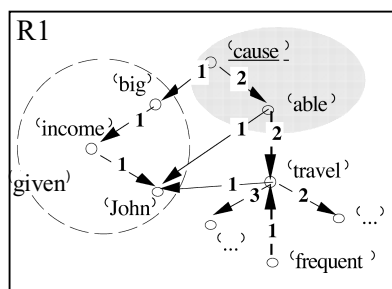


Fig. 11. SemR of (3a)

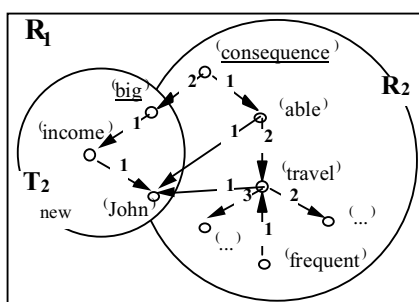


Fig. 12. SemR of (3b)

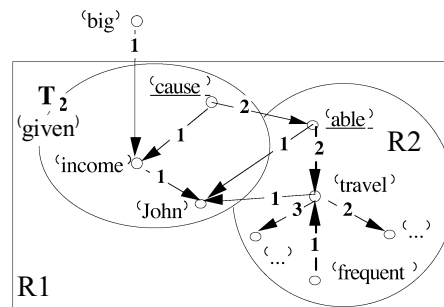


Fig. 13. SemR of (3c)

Let me sketch two propositional quasi-equivalencies needed to link these representations.

To get SemR of (3b) starting from that of (3a), we need a rule stating the quasi-equivalence between the semantemes 'cause' and 'consequence', conversives of each other.

Semantic quasi-equivalence rule-3:

'P causes Q' $\cong$ 'Q is a consequence of P'.

To get the RSem of (3c) from that of (3a) we need an ellipsis rule which will remove the semanteme 'big' (underlying the lexeme COMFORTABLE).

Semantic quasi-equivalence rule-4:

'P which is W' $\cong$ '[there is] P'

| 1) 'P' is given; 2) 'P' is the ASem 1 of 'cause'.

This rule allows for the paraphrases such as *John's comfortable income* $\sim$ *John's income*; *with his [= John's] comfortable income* $\sim$ *with his income*, etc.

Note that another rule, similar to but less 'radical' than the present one, can be written, which would allow us not to delete 'W' but to implement it using less specific expressions SUCH and KIND OF, which are generic terms of sorts for the qualifiers like *comfortable*. With such a rule, it is possible to get paraphrases of type *with [his] comfortable income* $\sim$ *with such an income* (*with this kind of income*), etc.

I will conclude by mentioning two lexicalization rules needed to produce sentences in (3).

The first lexicalization rule takes care of the implementation of the semanteme 'cause' as WITH, the implementation that we see in *with his*

income (3c). This is an adverbial variant of CAUSE and it can be used in contexts where the cause is known (or given, in communicative terms). Such a cause can also be expressed using the lexeme SINCE (*Since John's income is comfortable*; *Since John has a comfortable income*) and, in a more syntactic vein, by means of a present participle trivially verbalizing the lexeme expressing the ASem 1 of 'cause' (in our case, light verbs *have* and *be* used with INCOME; cf.: *Having a comfortable income*, *John ...*; *His income being comfortable*, *John ...*).

The second lexicalization rule accounts for the difference *can travel* (3b) ~ *can afford to travel* (3c). It stipulates that the semanteme 'able' can be implemented in two different ways: as CAN or, optionally, as CAN AFFORD, the second lexicalization being allowed in the context where the ability to do something depends on the existence of (adequate) financial resources.

Acknowledgments

Many thanks to Marie-Josée Hamel, Leonid Iomdin, Igor Mel'čuk and Alain Polguère for their comments on a prefinal version of this paper.

References

- Apresjan, Ju. (1974). *Leksičeskaja semantika. Sinonimičeskie sredstva jazyka*. Moskva: Nauka.
- Apresjan, Ju. (2000). *Systematic Lexicography*. Oxford, NY: Oxford University Press, 1974.
- Iomdin, L. (1972). Ob odnom klasse anglijskix kauzativnyx glagolov, obrazovannyx sposobom konversii. *Mašinnyj perevod i prikladnaja lingvistika*, n° 15. 1972. P. 99—142.
- Mel'čuk, I. (1974). *Opyt teorii lingvističeskix modelej Smysl ⇔ Tekst. Semantika, sintaksis*. Moskva: Nauka. [Reprinted in 1999, Moskva: Jazyki russkoj kul'tury], 1974.
- Mel'čuk, I. (1988). *Dependency Syntax: Theory and Practice*. Albany, N. Y.: State University of New York Press, 1988.
- Mel'čuk, I. (1992). Paraphrase et lexique: la théorie Sens-Texte et le *Dictionnaire explicatif et combinatoire*. In: Mel'čuk et al. 1992. P. 9—59.
- Milićević, J. (2003). *Modélisation sémantique, lexicale et syntaxique de la paraphrase*. [Ph. D. Thesis]. Montréal: Université de Montréal, 2003.
- Milićević, J. (to appear). *Semantic equivalence rules in Meaning-Text Paraphrasing*. In: Wanner, L., (ed.). *Selected Lexical and Grammatical Topics in*

- Meaning-Text Theory. In Honor of Igor Mel'čuk.* Amsterdam/Philadelphia: Benjamins.
- Polguère, A. (1998). Meaning-Text Semantic Networks as a Formal Language. In: Wanner, L. ed., *Recent Trends in Meaning-Text Theory*. Amsterdam/Philadelphia: Benjamins, 1998. P. 1—20.
- Wanner, L. ed., (1996). *Lexical Functions in Lexicography and Natural Language Processing*, Amsterdam/Philadelphia: Benjamins, 1996.
- Žolkovskij, A. (1964). Predislovie. In: Rozenzweig, V. Ju. (ed.). *Mašinnyj perevod i prikladnaja lingvistika*, n° 8. 1964. P. 3—16.
- Žolkovskij, A. and Mel'čuk, I. (1967). O semantičeskom sinteze. *Problemy kibernetiki*, n° 19. 1967. P. 177—238.

Соответствие «смысл $\Leftrightarrow$ текст» в исторической перспективе

Е. В. Падучева

Всероссийский институт научной и технической информации РАН
avenul@atom.ru

Современная семантика немыслима без тех понятий и приемов семантического описания, которые были разработаны в рамках модели «Смысл $\Leftrightarrow$ Текст» — таких как валентность, синтаксический и семантический актанта, диатеза. В рамках МСТ возникло понятие сентенциальной формы и толкования, имеющее аргументно-предикатную структуру. Из МСТ выросла теория интегрального описания языка и системная лексикография.

Meaning $\Leftrightarrow$ Text correspondence in a historical perspective. Modern semantics is unthinkable without those notions and devices of semantic description that originated in the frame of «Meaning $\Leftrightarrow$ Text» model — such as valency, semantic and syntactic argument, diathesis. «Meaning $\Leftrightarrow$ Text» model gave rise to the notion of sentential form, and lexical decompositions having predicate-argument structure. This theory is responsible for the idea of integral description of language and systemic lexicography.

Что было. Представим себе лингвистику середины 60-х годов прошлого века. Господство победившего (т. е. разрешенного в СССР) структурализма: главные авторитеты — Л. Блумфильд, З. Харрис; популярный в переводе на русский язык Г. Глисон и, конечно, Хомский (который на нашем черно-белом горизонте попадал в структуралисты). И на этом сугубо асемантичном фоне, почти из ничего, возникает модель «Смысл $\Leftrightarrow$ Текст». Возникает — семантика. Базой для нее служит синонимия: смысл понимается (по Куайну, Черчу, Якобсону) как **инвариант синонимических преобразований**. Миллион синонимических перефразировок предложения

Д-р Смит не может перевести этот текст из-за обилия технических терминов,

предъявленный в Мельчук, Жолковский 1965, 1967, доказывает, что смысл есть. Сейчас трудно даже себе представить, какой революцией было это введение смысла в разряд существующих категорий.

Разумеется, появляются и в других странах работы, созвучные этому важному семантическому прозрению. Есть работы У. Вейнрейха (Weinreich 1966; между прочим, это одна из первых попыток сформулировать принцип композиционности в семантике); есть Lakoff 1968; есть известные работы Филлмора 60-х и начала 70-х годов; есть Wierzbicka 1969 и 1972 и др. Но мощностъ заряда, который идет от работ И. А. Мельчука, А. К. Жолковского, Ю. Д. Апресяна и их коллег колоссальна: ведутся широкомасштабные семантические исследования, работает коллектив талантливых людей, увлеченных открывшейся перед ними невиданной перспективой, и они сразу добиваются огромного успеха.

Итак, формируется некая совокупность идей, которая чуть позже получит название модель «Смысл $\Leftrightarrow$ Текст» (МСТ). Какая же? Я намечу некоторые штрихи, делая основной упор на семантику.

1. **Смысл.** До середины 60-х годов слова «смысл», как термина, в лингвистике не было. Говорили о значении слова (например, английского); о функции грамматической формы (в не-структурной лингвистике); была даже семасиология. Но не смысл. Перечитывая сейчас, когда слово стало привычным, работы 60—70-х годов, удивляешься той осторожности, с которой говорили о смысле пионеры семантики.

2. **Смысл предложения.** МСТ предложила семантическое представление для предложения как целого. Это отдельное достижение творческой мысли: на всех уровнях (фонологическом, морфологическом, синтаксическом, семантическом) предлагалось представление не для слова или грамматической формы, а для предложения в целом. Что говорить, если даже Роман Якобсон, яростный борец с асемантизмом американских дескриптивистов, говоря о смысле, имел в виду, в основном, смысл слова, ср. его известное — опирающееся на Ч. Пёрса — определение: «Значение знака — это его перевод в другой знак, в котором его значение выявлено более полно (is more fully developed)». (Отсюда эксплицитная перефразировка, которая играла принципиальную методологическую роль для обоснования понятия смысла в МСТ.) Так или иначе, в середине 70-х уже можно было сказать: «В настоящее время это понятие смысла предложения все шире входит в лингвистический обиход» (Падучева 1975).

3. **Толкование слова.** Одно из открытий МСТ состояло в том, что слово должно толковаться не само по себе, а в составе сен-

тенциальной формы — с переменными по актантам. В то время, как Дж. Катц и П. Постал еще говорили на языке семантических признаков, в Москве смысл слова уже складывался из смысловых компонентов, имеющих аргументно-предикатную структуру. Толкование слова как сентенциальной формы, впервые предложенное в МСТ, полностью определяет нынешнюю концепцию лексикологии и лексикографии.

4. Модель управления: валентности; семантические и синтаксические актанты, диатеза. Именно эти понятия лежат в основе семантических метаязыков, достаточно тонких и мощных, чтобы обеспечить естественное взаимодействие семантики с синтаксисом. На базе понятия валентности (она же — семантическая роль, см. Апресян 1974: 25) возникает ролевая классификация предикатных имен — отдельное достижение МСТ, значение которого еще не осознано до конца.

5. Лексические функции. Собственно говоря, лексические функции (ЛФ) — это первое, что связывалось с моделью «Смысл $\Leftrightarrow$ Текст» в сознании лингвистического сообщества. Повторяющееся семантическое отношение между словами в сочетаниях *бурные аплодисменты, жгучий брюнет, отъявленный негодяй* и сотнях других, производило ошеломляющее впечатление (другой ряд — *оказывать влияние, вести борьбу, производить впечатление*, и таких много). В Мельчук 1974: 79 отмечены два момента, которые придают некоторому смыслу статус лексической функции: широкая сочетаемость и разнообразие выражений. Отсюда двоякая роль ЛФ. Одно их назначение — систематизация фразеологии; другое — то, что на лексические функции опирались синонимические преобразования, а на синонимических преобразованиях, по крайней мере на первых порах, держался смысл¹.

Лексические функции были основным ресурсом синонимии: массовость синонимии была открыта именно благодаря выходу за границы слова. Разница между *влиять* и *оказывать влияние*; *владеть* и *являться владельцем*; *вызывать* и *быть причиной* ничтожна — в основном она касается актуального членения. А разные слова, даже близкие синонимы, обычно различаются по многим параметрам —

¹ Сейчас, пожалуй, на первый план выходит третий аспект: ЛФ понимаются как строевые смыслы некоего универсального семантического языка (см. Перцова 2004) — при том, что в Мельчук 1974: 78 ясно говорится про ЛФ, что они «вводятся как средство описания лексической сочетаемости, а вовсе не семантики».

пример можно найти практически в любой словарной статье Нового объяснительного словаря синонимов (НОСС). Полного тождества смыслов нет даже для отдельных значений слова, не говоря о словах в целом (пары типа *глядеть* — *смотреть*, которые совпадают почти во всех значениях, — скорее исключение, чем правило).

Пафос создателей МСТ во многом состоял в том, что это действующая модель — описание, которое может быть подвергнуто эмпирической проверке. Если говорить о моделировании языковой способности к осуществлению синонимических преобразований, то это достижение МСТ, никем впоследствии не повторенное.

Что сейчас. МСТ заложила основы современной семантики и надолго определила направление дальнейших исследований. Как бы далеко мы ни продвинулись, связь с тем импульсом, который возник в середине 60-х, не перестает ощущаться. Лингвисты, работавшие в лексической семантике в русле МСТ, имели огромную фору: работы американских ученых — Д. Даути, Р. Джекендоффа, Б. Левин, — которые играют важную роль в мировой семантике, в том числе, в семантике лексики, появились на десятилетия позже. Так как же развиваются семантические идеи МСТ в современной лингвистике? Наметим несколько основных направлений².

1) В МСТ классического периода соответствие «смысл $\Leftrightarrow$ текст» мыслилось как переход «смысл $\Rightarrow$ текст» — от смысла к тексту. Все-таки, порождающая грамматика Хомского оказывала влияние на всю мировую лингвистику — в том числе и на Россию. Да и смысл, внешне обретший право на существование, требовал для себя центральной позиции. Сейчас это соответствие мыслится, скорее, как переход в обратную сторону — **от текста к смыслу**. Если МСТ ориентировалась на порождение текста, то сейчас ключевое слово — **интерпретация**. Взять хотя бы статью Богуславский 2004 о слове *большинство* и способах выражения семантических валентностей этого трудного слова: описание вкладывается в рамки интерпретационного, а не порождающего подхода; нет речи о глубинной структуре, в которой семантическим валентностям соответствовали бы синтаксические, и трансформациях, приводящих к данной поверхностной. Надо сказать, что отказ от трансформаций в порождающей грамматике означает, что и там произошел аналогичный сдвиг от порождения к интерпретации.

² Я исхожу из того, что развитие идей МСТ не ограничено рамками Московской семантической школы — как бы ни были заданы эти рамки в научном плане или географическом.

Соответствие между смыслами и текстами было и остается задачей на будущее. Для полного описания этого соответствия нужны как минимум два компонента: множество смыслов (семантический язык — порождающая семантика) и множество текстов (порождающая грамматика). И в том и в другом плане остается над чем подумать. Но при этой неполноте порождающий подход оказывается в тупике, а установка на интерпретацию позволяет устанавливать соответствие «текст $\Leftrightarrow$ смысл» в локальном плане — на уровне отдельных слов или конструкций, но так или иначе двигаясь вперед.

2) Другой важнейший сдвиг: в игру входит третья вершина треугольника Фреге — **действительность**. Смыслы соотносятся теперь не только с текстами, но и с внеязыковой реальностью. Но для лингвиста, в отличие от логика, слово — это не обозначение какого-то фрагмента действительности, а его концептуализация. Она может быть разной для разных говорящих и для разных языков. Тем самым обретает новое содержание термин языковая картина мира (Апресян 1999). Первый мощный стимул в этом направлении содержался, конечно, в Wierzbicka 1992. Сейчас это широкий фронт исследований (см. Зализняк, Левонтина, Шмелев 2005), которые служат, в частности, семантическим обоснованием лингвистической типологии (Кибрик 2003).

Новый подход к соотношению языка и действительности создает плодотворную основу для сопоставления языков, в том числе — родственных. В том месте, где в газете «Правда» было в свое время написано «Пролетарии всех стран, соединяйтесь!», в «Rude pravo» фигурировало *Proletáři všech zemí, spojte se!*, букв., «...соединитесь!». Скорее всего, дело тут именно в различии концептуализаций, привычных для той или иной языковой модели мира: там, где русский язык ставит в фокус внимания деятельность, чешский следует европейскому стандарту и упрощает дело, сосредотачиваясь на результате. Языковая модель мира определяется набором свойственных данному языку концептов, и концепты, можно думать, различаются в первую очередь акцентами, поставленными на разных сторонах того или иного явления «объективной действительности».

3) Еще один сдвиг — от синонимии к омонимии (точнее, к полисемии). **Многозначность** осознается теперь как центральная проблема семантики лексики и грамматики. В самом деле, так много в языковом поведении слова зависит от значения, а значение меняется от контекста к контексту! Один способ справиться с многозначностью состоит в том, чтобы представить значения многозначного слова как полученные в результате семантических переходов, сдвигов

(отчасти контекстных); говорят о семантической деривации, о мотивированности одного значения слова в словаре другим. В каком-то смысле сейчас объектом семантики становятся в большей степени модели семантической деривации, чем отдельные значения.

Продemonстрирую массовость и в то же время регулярность многозначности на семантических переходах в классе глаголов восприятия (Падучева 2004: 199).

Семантический переход от восприятия к ментальному значению. Производное ментальное значение развивается, например, у глаголов *видеть*, *смотреть*, *замечать*, *рассматривать* (ср. *рассматривать* как намек и мы *рассматриваем* ваше предложение), *чувствовать*, *казаться*, *обнаружить*, *слышать*, *воображать*, *столкнуться*, *следить*, *показаться*; *представляться*, *видеться* (та же неоднозначность у слова *взгляд*):

- (1) а. От прилавка ему хорошо *видалось* клубное крылечко [зрительное значение];
- б. Мне *видится* это так [ментальное значение].

Глагол *свидетельствовать* этимологически предполагает видение; но в контексте *Это свидетельствует о его незаурядном таланте* он имеет ментальное значение; *пролить свет* — значит ‘сделать более понятным’ — хотя свет нужен, вообще говоря, для того, чтобы видеть.

Производное ментальное значение возникает и у каузативных глаголов; так, *показать* — глагол восприятия, но может иметь и значение ‘доказать’, ментальное.

Производное значение речи развивает глагол *заметить*; оно проявляет себя в сочетаемости с наречиями:

- (2) Это ты *верно* заметил [= ‘верно сказал’]; ср. **верно* увидел.

Семантический дериват в классе глаголов речи есть и у ряда других глаголов восприятия, ср. *отметить*, *подметить*, *выделить*. Нечто похожее — в примере

- (3) *Обращаю* ваше внимание на следующую неточность,

где у *обращать* (внимание) речевое и даже, пожалуй, перформативное значение, производное от значения восприятия.

В русском языке эта деривационная модель — переход от восприятия к речи — не продуктивна. Между тем, как отмечается в статье Dobrovol'skij 2001, глаголы восприятия в немецком языке (такие как *sehen*, *finden*, *wahrnehmen*, *empfinden*) весьма охотно употребляются в значении речи, например:

- (4) In seiner Rede *empfand er* die Vorwürfe als gerechtfertigt ‘В своей речи он сказал, что *ощущает* (букв. «ощутил») упреки как справедливые’.

Производное значение воображаемого восприятия (зрительного или слухового), когда в ситуации нет участника Стимул, так что образ возникает в сознании независимо от стимула извне, — у глагола *видеть*, ср. *вижу, как сейчас*; это значение есть также у глаголов *послышаться, представиться* <взору>.

Глаголы *послушать, слушать, послушаться, внять* демонстрируют переход ‘воспринять’ → ‘подчиниться’: *не слушай* может значить ‘не слушайся, не подчиняйся, не относись всерьез’; *прислушаться* <к мнению> = ‘отнестись с вниманием’; *их голос не был услышан* = ‘их мнение не было принято во внимание’. То же развитие у *оглядываться* (*Даже в играх он оглядывался на взрослых*).

Регулярным, т. е. повторяющимся, является также семантический переход ‘смотреть’ → ‘относиться’:

- (5) *Я на это смотрю просто* = ‘отношусь просто’; *смотреть сквозь пальцы* = ‘потворствовать’; *несмотря на* ≈ ‘безотносительно к’;
- (6) *коситься* = 1) ‘смотреть искоса, сбоку’; 2) ‘смотреть косо, относиться с подозрением’, ‘выражать взглядом подозрительное отношение’.

Переход ‘видеть’ → ‘иметь’ представлен примерами *найти, потерять*; ср. также *не видать как своих ушей* = ‘не иметь’ или *Вам не видать таких сражений*.

Переход от восприятия к межличностному контакту отмечается у глаголов *встречаться, заглянуть* <на огонек>, *увидеться*.

Смысл ‘видеть’ может выветриваться до идеи простого контакта с объектом, т. е. пребывания в том же месте. Значение соположения — тривиальное следствие перцептивного компонента; чтобы Y мог видеть X, они должны находиться в одном месте; отсюда употребления:

- (а) Эти стены *видели* многое; (б) Крым всегда будет рад вас *видеть*.

В самом деле, смысл (а) — ‘в этих стенах многое происходило’; смысл (б) сложнее — из-за *рад*; но ясно, что речь идет о приезде в Крым.

Для восприятия необходимо, чтобы Экспертиент находился в определенном месте, и потому восприятие часто является следствием

перемещения — либо Эксперимента, либо Перцепта; это отражено в семантике глаголов, которые обозначают восприятие как следствие контакта, возникшего при перемещении:

столкнуться, наткнуться, напороться, нарваться; попать 1 (как в *Мне попался белый гриб*) [попать 2 — это как в *Я попался*].

Отсюда неоднозначность у *наткнуться* (а также *напороться*, *нарваться*): 1) ‘двигаясь, прийти в контакт’; 2) ‘случайно увидеть’.

У глагола *явиться* есть два значения, которые различаются только фокусом внимания: ‘представиться взору Наблюдателя, переместившись’, как в *Гамлету явилась тень отца*, и ‘переместиться (представ перед Наблюдателем)’, как в *явился по первому требованию*. Более сложная деривация связывает два значения англ. глагола *to glimpse*: 1) ‘мельком взглянуть’; 2) ‘промелькнуть’, ср. подобное соотношение в русской паре *глядеть* — *выглядеть*: объект становится подлежащим.

Следствием перемещения может быть, наоборот, выход из поля зрения, как у *скрыться*, *деться*, *задеваться*. У *скрыть* компонент «перемещение» факультативный; а у *скрыться* фактически обязательный. У *выступать*, *высовываться* в определенных контекстах появляется Наблюдатель. У глагола *проступить* значение ‘начать находиться в поле зрения’ стало основным.

Другой пример — многозначность ‘сказать — сделать’. Она свойственна многим глаголам речевого действия (Падучева 2004: 371):

- (7) а. *подтвердить* приказ [= ‘сказать, что Р’];
б. *подтвердить* право [= ‘сделать нечто, подтверждающее Р’].
- (8) а. *ответить* [= ‘сказать’];
б. *ответить* [= ‘понести наказание за содеянное’].
а. *повторить* сказанное [= ‘сказать еще раз’];
б. *повторить* действие [= ‘сделать еще раз’].

Такое же различие (с поправкой на дополнительную разницу, обусловленную акциональным классом) между двумя значениями глаголов *доказать*, *решить*, *убедить*:

- (10) он убеждает / убедил $\supset$ ‘говорит’ / ‘говорил’;
это убеждает / убедил $\supset$ ‘каузировало изменение состояния Эксперимента’.

Глагол *угovarивать* не допускает этого семантического сдвига (*Он уговаривает*, но **Это уговаривает*) — у него неподавляемый речевой компонент.

В Подлесская 2004 описана многозначность глагола *дать*, которую можно представить как противопоставление ‘каузация словом’ — ‘каузация делом’:

- (11) а. *Дайте* мне задание;
 б. Ребенок *дал* мне поспать.

Та же многозначность слова и дела у глагола *ошибиться*. Глагол СВ *ошибаться* имеет, согласно словарям, два значения: оцениваться как ошибочное может и мнение, и поступок (обычно выбор решения). В первом значении глагол имеет квазисинонимичный НСВ (*ошибаться* — *ошибиться* — перфектная видовая пара); а во втором (ср. *ошибиться* 2 в Mel'chuk 2004; *ошибиться адресом* = ‘пойти по неправильному адресу’, действие) у НСВ только тривиальное значение.

Неоднозначность возникает при наличии в слове нескольких семантических компонентов, каждый из которых может по-своему взаимодействовать с контекстом. Так, фраза *Напрасно он ее утешал* может означать: 1) ‘утешал, но не утешил’ (*напрасно* взаимодействует с речевым компонентом глагола *утешать* и имеет значение неэффективности речевого действия); 2) ‘высказал мнение, что все не так плохо, но ошибся’ (*напрасно* взаимодействует с компонентом мнение); 3) ‘лучше было не утешать’ (*напрасно* является модификатором предложения и выражает отрицательную оценку действия в целом), см. Зализняк 1992: 155—166.

При наличии у той или иной формы нескольких значений возникает задача понять, как они связаны между собой: чаще всего, производное значение бывает мотивировано исходным значением и контекстом. В работе В. Апресян (1991) фигурируют два семантических компонента глагола *гореть*: грубо говоря, ‘испускать огонь, а значит, свет и тепло’ и ‘уничтожаться под действием огня’ (ср. толкование в Апресян 1974: *гореть* = ‘выделяя огонь, изменяться под воздействием огня’). При этом вторичные, производные значения *гореть* могут быть представлены как полученные в результате сдвига фокуса внимания на один из компонентов и подавление другого. Так, в контексте (12), в фокусе второй компонент, а в (13) — первый:

- (12) а. На нем одежда *горит* = ‘быстро изнашивается, как будто уничтожается под действием огня’;
 б. Путевка *горит* = ‘срок годности скоро истечет’.
 (13) а. Лампа *горит* = ‘лампа включена’;
 б. Крыши *горят* на солнце = ‘сверкают, блестят, испускают свет’.

В некоторых случаях можно говорить о парадигмах семантической деривации, общих для классов слов, см. Падучева 2004: 401 о парадигме глаголов звука; Урысон (в печати) о семантической парадигме слов типа *дождь*.

Выявление связей между значениями многозначного слова привело к осознанию частичной мотивированности лексико-функциональных значений, см. Reuther 1998, Mel'čuk, Wanner 1996; о связи между лексико-функциональным и обычным значением глагола см. Апресян 2004.

4) **Таксономия.** Одно из ценнейших приобретений лингвистики последних лет — это онтология: таксономические классы глаголов и имен. В новаторских работах Мельчук 1974 и Апресян 1974 не отводилось существенного места традиционному аспекту словарных определений — таксономическому. Внимание к таксономии было привлечено в Wierzbicka 1980 и Dowty 1979, где была намечена связь между таксономической категорией слова и его форматом толкования. Таксономический бум произвела книга Lakoff, Johnson 1980, обратившая внимание лингвистов на тот факт, что категории играют важную роль в метафорических, а также и метонимических переносах (на которых, в основном, и стоит вся динамика лексической семантики, Падучева 2004: 155), и что для словарного толкования важно родовое понятие слова — *genus proximum* аристотелевской логики.

Сколь важен таксономический класс участника для выявления актантной структуры слова, можно видеть на примере глагола *changer* 'изменить', см. Mel'čuk 1992: изменяется обычно не предмет, а ПАРАМЕТР, ср. *изменился направление, адрес, объем*. В Iordanskaja, Raperno 1996 подробно описан важный таксономический класс имен — ЧАСТИ ТЕЛА. Надо сказать, что одним из первых таксономических классов, описанных в лингвистической литературе, были ЭМОЦИИ, см. Иорданская 1970.

Другая классификация, независимая от первой, — аспектуальная (= акциональная). Теперь в лексическую семантику инкорпорированы вендлеровские глагольные классы (Wierzbicka 1980: 177, Булыгина 1982), которые до того имели хождение только в аспектологии. Акциональный класс задает схему толкования слова, а из схемы толкования выводится парадигма видовых значений глагола и масса других важных семантических следствий (Падучева 1996: 107). Ср., однако, в Апресян 2003 о фундаментальной классификации, которая не различает таксономических и акциональных классов.

5) **Компоненты толкования и их акцентный статус.** Толкование слова в МСТ (70-е и 80-е годы) было предложением. Теперь

толкование скорее мыслится как набор компонентов — каждый со своим коммуникативным (акцентным) статусом: ассерция, пресуппозиция, импликация, атрибутивный компонент и др. См. пример (1). Глаголы *скрыть* и *спрятать* близки по смыслу; оба значат примерно ‘переместить, чтобы не было видно’. То, что *долго* имеет сферой действия разные компоненты, означает, что у этих компонентов разный статус:

- (1) а. *долго скрывал* *X* = ‘об *X*-е долго не знали’;
- б. *долго прятал* *X* = (в одном из пониманий) ‘долго думал, куда поместить, и перемещал’.

Интересен пример из Wickboldt 2000:

- (2) а. Since he entered the room, he’s been looking for a seat;
- б. Since he entered *quietly*, he’s been looking for a seat.

Союз *since*, который имеет два значения, причинное и временное, в (2а) может быть понят в обоих значениях, а в (2б) — только в причинном. Дело в том, что наречие образа действия подавляет временную валентность глагола действия. Русское *когда* имеет только временное значение; поэтому (3) может быть осмысленным разве что в контексте противопоставления (...а в тот раз, когда она приехала с ордером на арест, я был в гостиной):

- (3) Когда полиция ворвалась в дом *незаконно*, я был на кухне.

Блокировать временной параметр события могут и другие обстоятельства. Пример из Падучева 2004: 188:

- (4) а. В полдень он начал свою речь;
- б. [?]В полдень он начал свою речь *с приветствия*.

МСТ предусматривала только один вид взаимодействия слова с контекстом — зачеркивание компонентов. На самом деле, чаще имеет место не зачеркивание, а сдвиг фокуса внимания. В перфектной паре *огорчить* — *огорчат* видовое противопоставление сводится к сдвигу фокуса внимания с событийного компонента на стательный. В предельной видовой паре соотношение между СВ и НСВ определяется сдвигом фокуса внимания с результата на деятельность: в *сварить* акцент на результате, в *варить* — на деятельности. Отсюда прагматическая потребность в общефактическом акциональном употреблении глагола НСВ, как в контексте *Кто варил этот борщ?*: только употребив несов. вид, можно поставить в фокус деятельность — форма сов. вида акцентирует результат, т. е. перфектное со-

стояние, а деятельность отходит на задний план — на периферию зоны внимания.

Одной из важных для МСТ идей был изоморфизм предложения и слова: толкования слов устроены точно так же, как и толкования предложений (Weinreich 1966, Мельчук 1974: 61). Эта линия получила продолжение. Пример (по мотивам Булыгина, Шмелев 1988). У предложения (5) имеются два прочтения, различающиеся просодией:

- (5) У меня осталось *немного* песен (которые я могу спеть).
 - а. ‘существует множество песен, которые остались [пресуппозиция]’ &
‘это множество невелико [ассерция]’;
 - б. ‘существует множество песен, которые остались [ассерция]’ &
‘это множество невелико [дополнительная ассерция]’.

При отрицании (5б) второй компонент выпадает:

НЕ (5б) = У меня не осталось песен.

В (5) имеется смысловая конфигурация, состоящая из ассертивного и атрибутивного компонента. Атрибутивный компонент с семантикой малой степени при отрицании выпадает. Но та же конфигурация может возникать на смысловых компонентах в толковании слова; так, компонент ‘контакт слабый’ у *касаться* (Богуславский 1985: 30) ведет себя так же, как безударное *немного*:

- (6) *Ветка касается окна* =
 - 1) ‘между окном и веткой имеется контакт’ [ассерция];
 - 2) ‘контакт слабый’ [дополнительная ассерция].

При отрицании второй компонент исчезает:

- (6') *Ветка не касается окна* = ‘между окном и веткой нет контакта’.

Получается, что правило взаимодействия с отрицанием для предложения и слова одно и то же.

Похожий семантический изоморфизм слова и предложения можно продемонстрировать на базе Богуславский 1985: 65. Пример (7) показывает, что некоторые валентности невыразимы в контексте отрицания:

- (7) а. Добыча угля повысилась на 50 тыс. тонн;
- б. *Добыча угля не повысилась на 50 тыс. тонн.

Но то же явление имеет место и на уровне слов. Начнем с известного примера из Мельчук, Холодович 1970:

попасть в зайца — *промахнуться в зайца.

У *попасть* есть валентность Мишень, а у *промахнуться* = ‘не попасть’ она отсутствует. Аналогичное различие между глаголом с «положительным» значением и его коррелятом-антонимом или негативом можно продемонстрировать на большом количестве примеров:

подтвердить кому-то — *опровергнуть кому-то;
сказать, что — *промолчать, что;
присутствовать при — ?отсутствовать при;
считать, что — ?ошибаться, что.

Ср. также пары антонимов (курсивом выделен участник, который у отрицательного коррелята отсутствует; некоторые примеры — из Апресян 1974):

осудили за хулиганство — оправдали,
выбрать делегатом — забаллотировать,
одеть в пальто — раздеть,
назначить заседание на среду — отменить заседание,
появился где-то — пропал.

6) От предложения к тексту. На первых порах отношение «смысл — текст» рассматривалось преимущественно на текстах, состоящих из одного предложения (см., однако, работу Iordanskaja 1992). С течением времени возникла возможность проникнуть в суть языковых явлений, свойственных именно тексту. Здесь можно упомянуть, в частности, понятие режима интерпретации (Падучева 1986). В Мельчук 1985 у наречия *сейчас* было выявлено значение *сейчас-4*, которое не фиксируется в словарях. Оно замечательно тем, что, в отличие от обычного *сейчас*, сочетается с прош. временем несов. вида:

- (8) а. К Мише нас не пустили. *Он был *сейчас* с дамой;
б. Этот мир был ему *сейчас* [= *сейчас-4*] ненавистнее, чем когда-либо. (Пастернак)

Для примера (8б) было предложено объяснение (в Падучева 1996: 271): *сейчас-4* — это нарративное *сейчас*; в (8а) речевой режим, и *сейчас* недопустимо; а в (8б), в нарративном режиме, оно вполне уместно: нарративная интерпретация становится возможной после того, как эгоцентрические элементы языка «апроприированы» (по Бенвенисту) героем нарратива.

Понятие режима интерпретации опирается на фигуру Наблюдателя по Ю. Д. Апресяну и на связанное с ней понятие вторичного дейксиса (Апресян 1986).

* * *

В заключение я рассмотрю одну частную задачу, которую можно поставить и решить в рамках МСТ. Одно из важнейших понятий МСТ — модель управления, родственником которой является диатеза, соответствие между участниками (ролями) и их синтаксическими позициями, иначе — между семантическими и синтаксическими актантами (Мельчук, Холодович 1970). Диатеза занимает центральное место в современных исследованиях по синтаксису и лексике. Естественное развитие теории семантических и синтаксических актантов приводит нас к понятию *инкорпорированных участников* (Jackendoff 1990, Падучева 2004: 57).

Инкорпорированный участник возникает в связи с так наз. *linking problem* — проблемой установления связи между семантическими и синтаксическими актантами глагола. Это участник, который входит в концепт ситуации (т. е. упоминается в толковании), но не имеет выхода на поверхность и соответствующего синтаксического актанта.

Инкорпорированный участник есть у многих глаголов; это, например, *глаза* у глагола *видеть*, *моргать*, *уши* у *слышать*, *голос* у *говорить*, *огонь* у *гореть*; *губы* у *целовать* (Апресян 1974: 26), *ноги* у *ходить*, *руки* у *держать*, *принести*, *взять* (в МАС *брать* определяется как ‘принимать на руки, схватывать руками’); прочие части тела тоже легко инкорпорируются. Участник класса ЗВУК инкорпорирован в семантике *звенеть* и прочих глаголов звука; соответствующий инструмент инкорпорирован в *пылесосить*, *утюжить* и под. Инкорпорирование в классе возвратных глаголов: *одеться* (одежда), *защититься* (диссертация), *нахмуриться* (брови), *оскалиться* (зубы) и под.

При употреблении глагола в исходной диатезе инкорпорированному участнику не соответствует синтаксического актанта. Синтаксическое выражение инкорпорированного участника становится возможным лишь в том случае, если в предложении в концепт участника, заданный толкованием, добавляется атрибут или квантификатор:

- (1) *Простым глазом* мы можем увидеть около 6 тысяч звезд (Зощенко);
- (2) видел *своими глазами*; Я *своими полуслепыми глазами* и то вижу, что рубашка не глаженная; видел *одним глазом* (ср. **вижу глазами*);
- (3) поднял мой рюкзак *одной рукой* (ср. **поднял рукой*; **поднял руками*).

Можно сказать, что в этом случае инкорпорированный участник «экскорпорируется» — покидает свою позицию За кадром и входит в перспективу.

Инкорпорированный участник дал о себе знать в семантике наречий. Наречия часто семантически связаны с прилагательными, но направление семантической производности в разных случаях разное. Наречия намеренности (которые описывают состояние сознания субъекта при выборе действия), такие как *намеренно*, *сознательно*, *машинально*, *нечаянно*, — семантически первичны: соответствующие прилагательные сочетаются только с предикатными именами (*машинальное действие*). Но есть и обратная ситуация. А именно, наречие в семантической структуре предложения может быть атрибутом к предметному аргументу — инкорпорированному участнику³.

Наречие *босиком* (см. о нем Филипенко 2003) является атрибутом к инкорпорированному участнику *ноги* в семантике некоторых глаголов движения:

шел / бегал босиком $\supset$ ‘ноги босые’.

Наречие *коротко* в контексте *коротко постричь* — это определение к *волосы*, которые являются инкорпорированным участником одного из значений глагола *постричь*:

Ее *коротко* постригли = ‘ее постригли; результат: волосы короткие’.

Наречия *устно*, *письменно* (см. об этих наречиях в Филипенко 2003: 128) — это определения к инкорпорированному участнику Текст глаголов речи:

ответил *устно* — значит ‘устным текстом’.

О том, что в ситуацию речи входит участник Текст (т. е. речевое произведение), быть может, инкорпорированный, см. Зализняк 1991. В дериватах он выходит на поверхность. Так, у глагола *гласить*, семантического конверсива от глагола речи, участник Текст — субъект: *Надпись гласит*. Ср. англ. *The label reads* ‘наклейка гласит’.

Имеется целая серия наречий, смысл которых можно представить как прилагательное размера при инкорпорированном участнике Текст глаголов речи —

сказал *вкратце*, *кратко*, *коротко* = ‘текст короткий’;
говорил *долго* = ‘говорил: текст длинный’.

³ Я благодарю М. Ф. Филипенко за предоставленную в мое распоряжение базу данных по наречиям.

Наречия *детально, лаконично, лапидарно, односложно, сжато* — это более специальные определения к тому же инкорпорированному участнику Текст.

Наречия *мелко, крупно, тонко, толсто* — это предикаты к инкорпорированному участнику Части у глаголов разделения на части:

порезал ветчину *тонко* = ‘порезал ветчину: части тонкие’.

Предикаты, характеризующие инкорпорированного участника Результат:

открыл дверь *широко* = ‘открыл дверь; проход широкий’;

закрыв дверь *плотно* $\supset$ ‘привел дверь в контакт с проемом; контакт с плотный’.

Наречие *прямо*; в контексте глаголов движения это предикат к предметному инкорпорированному участнику Траектория:

идет *прямо* = ‘идет; траектория прямая’.

В Geuder 2002 рассматриваются два значения наречия *элегантно*. В контексте глагола *одеться* его сферой действия могут быть два разных компонента глагола:

(4) она оделась *элегантно* =

а) одежда *элегантная* [предметный аргумент; адъективное значение];

б) движения *элегантные* [непредметный аргумент; адвербиальное значение].

А в контексте глагола *раздеться* возможно только одно из этих значений — адвербиальное, поскольку инкорпорированный участник *одежда* не входит в концепт результата этого действия.

Представляют интерес также наречия степени — *полностью, частично, наполовину*. Их смысл может быть представлен словами *весь, часть, половина*, тоже в применении к предметному участнику — который может быть либо выраженным на поверхности, как в (5), либо инкорпорированным, как в (6); но в любом случае он является предметным:

(5) *запасы наполовину* израсходовались [израсходовалась половина запасов];

солнце наполовину скрылось [скрылась половина солнца].

(6) Пятно света *наполовину* сползло со стены на пол [половина траектории];

Моя деятельность *наполовину* проходит вообще вне завода [половина времени].

Наречие — область менее изученная, чем глагол. То, что хотя бы часть наречий можно описать такими простыми средствами, кажется вкладом и в семантику лексики и в теорию семантических актантов.

* * *

Подведем итоги. Модель «Смысл $\Leftrightarrow$ Текст» лежит у истоков современной семантики; из нее выросла теория интегрального описания языка и системная лексикография. Семантика в рамках МСТ возникла и оформилась раньше, чем в других направлениях; в короткий срок она создала огромную научную продукцию в виде словарей нового типа, содержащих сведения о словах, которые нельзя было почерпнуть ни из каких лингвистических источников.

Без тех понятий и приемов семантического описания, которые возникли в свое время в МСТ, немыслима современная семантика. Взять хотя бы понятия валентности, семантического и синтаксического актанта, диатезы и др. На базе этих понятий возник аппарат, позволяющий решить проблему соединения семантики с синтаксисом. Понятие семантического представления предложения позволило объединить усилия лингвистов и логиков. В рамках МСТ были предложены понятие сентенциальной формы и толкования, имеющие аргументно-предикатную структуру. Именно модель «Смысл $\Leftrightarrow$ Текст» создала необходимые предпосылки для интегрального описания языка.

Многое было переосмыслено за прошедшие почти 40 лет; кое-где сдвинуты акценты. Перестало быть актуальным четкое деление на языковые и неязыковые знания. Была осознана необходимость выявления связей между значениями — как у слов, так и других языковых элементов. Подготовлена база для сопоставления языковых моделей мира, для выхода за пределы одного языка и изучения языка в типологической перспективе.

Есть две возможности: рассматривать семантику МСТ как теорию, замкнутую в себе, или как естественно развивающуюся в контексте мировой науки. В докладе делается попытка исследовать эту вторую возможность и показать, как модель «Смысл $\Leftrightarrow$ Текст» и выросшая из нее семантическая школа интегрирована в мировую лингвистику.

Литература

- Апресян 1974 — *Апресян Ю. Д.* Лексическая семантика: Синонимические средства языка. М.: Наука, 1974.
- Апресян 1986 — *Апресян Ю. Д.* Дейксис в лексике и грамматике и наивная модель мира // Семиотика и информатика. Вып. 28. М., 1986. С. 5—33.
- Апресян 1999 — *Апресян Ю. Д.* Теоретическая семантика в конце XX столетия // Изв. РАН. Сер. лит. и яз. 1999. Т. 56. № 4.
- Апресян 2003 — *Апресян Ю. Д.* Фундаментальная классификация предикатов и системная лексикография // Грамматические категории: иерархии, связи, взаимодействие: Материалы международной научной конференции. СПб: Наука, 2003.
- Апресян 2004 — *Апресян Ю. Д.* О семантической непустоте и мотивированности глагольных лексических функций // Вопр. языкознания. 2004. № 4. С. 3—18.
- Апресян В. 1991 — *Апресян В. Ю.* Словарная статья глагола *гореть* // Семиотика и информатика. Вып. 32. М.: ВИНТИ, 1991.
- Богуславский 1985 — *Богуславский И. М.* Исследования по синтаксической семантике. М.: Наука, 1985.
- Богуславский 2004 — *Богуславский И. М.* Целое — Часть — Признак // Сокровенные смыслы: Слово. Текст. Культура: Сб. статей в честь Н. Д. Арутюновой / Отв. ред. Ю. Д. Апресян. М.: 2004. С. 44—53.
- Булыгина 1982 — *Булыгина Т. В.* К построению типологии предикатов в русском языке // Отв. ред. О. Н. Селиверстова. Семантические типы предикатов. М.: Наука, 1982. С. 7—85.
- Булыгина, Шмелев 1988 — *Булыгина Т. В., Шмелев А. Д.* Несколько замечаний о словах типа *несколько* (к описанию квантификации в русском языке) // Язык: система и функционирование. М.: Наука, 1988. С. 44—54.
- Жолковский, Мельчук 1965 — *Жолковский А. К., Мельчук И. А.* О возможном методе и инструментах семантического синтеза. НТИ. Сер. 2. 1965.
- Зализняк 1991 — *Зализняк Анна А.* Словарная статья глагола *ГОВОРИТЬ* // Семиотика и информатика. Вып. 32. М.: ВИНТИ, 1991. С. 71—83.
- Зализняк 1992 — *Зализняк Анна А.* Исследования по семантике предикатов внутреннего состояния. Slavistische Beiträge, B. 298. Otto Sagner Verlag. München, 1992.
- Анна А. Зализняк, И. Б. Левонтина, А. Д. Шмелев.* Ключевые идеи русской языковой картины мира. М., 2005 (в печати).
- Иорданская 1970 — *Иорданская Л. Н.* Попытка лексикографического толкования группы русских слов со значением чувства // Машинный перевод и прикладная лингвистика. Вып. 13. М.: МГПИИЯ, 1970.
- Кибрик 2003 — *Кибрик А. Е.* Константы и переменные языка. СПб.: Алетейя, 2003.
- Мельчук 1974 — *Мельчук И. А.* Опыт теории лингвистических моделей «Смысл ⇔ Текст». Ч. 1. Семантика, синтаксис. М.: Наука, 1974.

- Мельчук 1985 — Мельчук И. А. Семантические этюды. I. «Сейчас» и «теперь» в русском языке // *Russian linguistics*. Vol. 9. Nos. 2—3. 1985.
- Мельчук, Холодович 1970 — Мельчук И. А., Холодович А. А. К теории грамматического залога // *Народы Азии и Африки*. 1970. № 4. С. 111—124.
- НОСС 2004 — Апресян Ю. Д. и др. Новый объяснительный словарь синонимов русского языка. 2-е изд. Москва; Вена, 2004.
- Падучева 1975 — Падучева Е. В. Некоторые проблемы моделирования соответствия между текстом и его смыслом // *Изв. АН СССР. Сер. лит. и яз.* 1975. № 6.
- Падучева 1986 — Падучева Е. В. Семантика вида и точка отсчета // *Изв. АН СССР. Сер. лит. и яз.* 1986. Т. 45. № 5. С. 413—424.
- Падучева 1996 — Падучева Е. В. Семантические исследования: Семантика времени и вида в русском языке. Семантика нарратива. М.: Языки русской культуры, 1996.
- Падучева 2004 — Падучева Е. В. Динамические модели в семантике лексики. М.: Языки славянской культуры, 2004.
- Перцова 2004 — Перцова Н. Н. Об аналоге лексических функций в XVII веке // *Сокровенные смыслы: Слово. Текст. Культура: Сб. статей в честь Н. Д. Арутюновой / Отв. ред. Ю. Д. Апресян*. М.: 2004. С. 214—222.
- Подлесская 2004 — Подлесская В. И. Русские глаголы *дать / давать*: от прямых употреблений к грамматикализированным // *Русский язык в научном освещении*. 2004. Вып. 2(8).
- Урысон (в печати) — Урысон Е. В. Логическая структура полисемии и ее реализации (секрет слова СЛЯКОТЬ) // *Русский язык в научном освещении*. № 1. 2006. (В печати).
- Филипенко 2003 — Филипенко М. В. Семантика наречий и адвербиальных выражений. М.: Азбуковник, 2003.
- Dobrovol'skij 2001 — Dobrovol'skij D. Lexical semantics and pragmatic conventions // E. Németh (ed.). *PRAGMATICS IN 2000: Selected Papers from the 7th International Pragmatics Conference*. Vol. 2. Antwerpen: International Pragmatics Association, 2001. P. 165—173.
- Dowty 1979 — Dowty D. R. *Word Meaning and Montague Grammar. The Semantics of Verbs and Times in Generative Semantics and in Montague's PTQ*. Dordrecht (Holland): Reidel, 1979.
- Geuder msc — Geuder W. *Oriented Adverbs: Issues in the Lexical Semantics of Event Adverbs (msc)*.
- Iordanskaja 1992 — Iordanskaja L. Communicative Structure and its Use during Text Generation // *Int. Forum Inf. and Docum.* Vol. 17. №. 2. April 1992. P. 15—27.
- Iordanskaja, Paperno 1996 — Iordanskaja L., Paperno S. *A Russian-English Collocational Dictionary of the Human Body*. Columbus, Ohio: Slavica Publishers, 1996.
- Jackendoff 1990 — Jackendoff R. S. *Semantic Structures*. Cambridge etc.: MIT Press, 1990.

- Lakoff 1968 — *Lakoff G.* Instrumental adverbs and the concept of deep structure // *Foundations of language*. 1968. Vol. 4. № 1.
- Lakoff, Johnson 1980 — *Lakoff G., Johnson M.* *Metaphors We Live by*. Chicago; L.: Univ. of Chicago Press, 1980.
- Mel'čuk 1992 — *Mel'čuk I. A.* *Changer et changement en français contemporain (étude sémantico-lexicographique)* // *Bulletin de la société de linguistique de Paris*. T. LXXXVII.
- Mel'čuk 2004 — *Mel'čuk I. A.* Actants in semantics and syntax I: actants in semantics. *Linguistics*. 42—1. 2004. P. 1—66.
- Mel'čuk, Wanner 1996 — *Mel'čuk I., Wanner L.* Lexical Functions and lexical Inheritance for Emotion Lexemes in German // L. Wanner (ed.). *Lexical Functions in Lexicography and Natural Language Processing*. Studies in Language Companion Series. Vol. 31. Amsterdam, 1996.
- Reuther 1998 — *Reuther T.* О сочетаемости фазовых глаголов с существительными в чешском и русском языках // Е. Hajičová (ed.). *Issues in Valency and Meaning: Studies in honour of Jarmila Panevová*. Prague: Charles Univ. Press, 1998. P. 62—74.
- Weinreich 1966 — *Weinreich U.* Explorations in Semantic Theory // T. A. Sebeok (ed.). *Current Trends in Linguistics*. Vol. 3. London; The Hague; P.: Mouton, 1966. P. 395—477.
- Wickboldt 2000 — *Wickboldt J. M.* Some effects of manner adverbials on meaning. C. Tenny, J. Pustejovsky (eds.) *Events as grammatical objects: The Converging Perspectives of Lexical Semantics and Syntax*. Stanford: CSLI Publications. CSLI publications, 2000.
- Wierzbicka 1969 — *Wierzbicka A.* *Dociekania semantyczne*. Wrocław etc.: Ossolineum, 1969.
- Wierzbicka 1972 — *Wierzbicka A.* *Semantic Primitives*. Frankfurt am. M.: Athenäum, 1972.
- Wierzbicka 1980 — *Wierzbicka A.* *Lingua mentalis*. Sydney etc.: Acad. Press, 1980.

К описанию валентностных свойств ЛФ семейства Oper_i, Func_i, Labor_{ij} (на японском материале) *

А. С. Панина

Институт востоковедения РАН
panina-anna@yandex.ru

Японский язык дает удобную возможность для того, чтобы изучить перенос валентностей от существительных к ЛФ-глаголам семейства Oper, Func, Labor, поскольку по оформлению актанта падежными показателями можно с достаточной уверенностью определить, подчинен он существительному или глаголу.

Предварительное рассмотрение трех японских глаголов — Oper₁ *оконану* «проводить, осуществлять», Oper₂ *ужэру* «получать» и Labor₁₃ *мотицуру* «использовать» — в сочетаниях с существительными позволило выявить в их поведении некоторые тенденции. Та валентность существительного, которая, будучи передана глаголу, становится его подлежащим (1-я для Oper₁ и 2-я для Oper₂), почти всегда реализуется при глаголе. Напротив, валентности, для реализации которых при глаголе нет подходящей синтаксической позиции, поскольку она занята существительным, как правило, остаются при существительном.

Towards a description of valency transfer to Oper_i, Func_i and Labor_{ij}-type verbs in Japanese. The Japanese language presents a good opportunity to study transfer of valencies between nouns and Oper, Func and Labor-type verbs, because an actant's marking is a reliable indication of whether the valency is instantiated by the noun or transferred to the verb. Preliminary study of three Japanese verbs — Oper₁ *okonau* «to carry out», Oper₂ *ukeru* «to receive», and Labor₁₃ *mochiuru* «to use» — in collocation with nouns shows some tendencies in their behavior. The valency of the noun which, if transferred to the verb, becomes the verb's subject (1st for Oper₁ and 2nd for Oper₂), is almost exclusively realized by the verb. On the other hand, some valencies of the noun which, if transferred to the verb, would interfere with the syntactic position already occupied by the noun itself, tend to be realized by the noun.

* Исследование выполняется при поддержке РГНФ, грант № 05-04-04388а, и РФФИ, грант № 04-01-00623а

Семейство лексических функций (ЛФ) $Oper_i$, $Func_i$, $Labor_{ij}$ — это глаголы, модифицирующие синтаксическое поведение предикатных существительных. Подчиняя существительное (C_0), они не привносят в сочетание собственный смысл, а как бы образуют с C_0 единый глагол. Различаются члены семейства тем, в каком качестве к ним присоединяется C_0 — как прямое дополнение, подлежащее или косвенное дополнение соответственно.

Интересным представляется такой малоизученный аспект их синтаксического поведения как взаимодействие с актантами C_0 .

Некоторые актанты переходят от существительного к ЛФ-глаголу. Внутри каждого из трех упомянутых типов ЛФ выделяются подтипы, различающиеся именно тем, как соответствующие глаголы распределяют актантов C_0 по своим валентностям — например, $Oper_1$ делает своим подлежащим первого участника ситуации C_0 , $Oper_2$ — второго.

Но некоторые зависимые могут оставаться при C_0 . Так, в примере *Она дает детям уроки музыки* две валентности существительного *урок* — кто учит и кого учат — реализуются при $Oper_1$ *давать*, но третья, содержание урока, не переносится к нему, а остается при существительном.

В русском языке, где глаголы и существительные способны управлять одними и теми же предложениями, зачастую бывает с первого взгляда сложно сказать, C_0 или ЛФ-глаголу подчинено то или иное дополнение. Иногда возможны обе интерпретации: скажем, в примере *Мы проводим опыты на свиньях* предложную группу *на свиньях* можно понять и как относящуюся к существительному (ср. *Опыты на свиньях оказались успешными*), и как подчиненную глаголу (ср. *На свиньях мы подобных опытов не проводим*).

С этой точки зрения для изучения переноса валентностей лучше подходит японский язык, поскольку зависимые глаголы и существительных в нем оформляются по-разному. В оформлении актантов существительных обычно участвует генитивный показатель *но*, а глаголы этим показателем не управляют:

<u>фан</u>	<u>кара</u>	<u>но</u>	нурэдзэнтэ	vs.
фанат	от	Gen	подарок	
'подарки от поклонников' ¹				
<u>фан</u>	<u>кара</u>	нурэдзэнтэ	о	морау
фанат	от	подарок	DObj	получать
'получать подарки от поклонников'				

¹ В японском языке отсутствует категория числа.

Конечно, есть исключения. Существует класс существительных, которые в ряде контекстов, в частности, в заголовках и рекламе, способны иметь управление глагольного типа:

эйга но ко:кай vs.
 фильм Gen премьера
 ‘премьера фильма’
«Та:минэ:та: 3» га нэммацу ни ко:кай
«Терминатор 3» Subj конец года в премьера
 ‘В конце года выходит «Терминатор 3»’.

При глаголах показателем *но* может оформляться подлежащее определительного придаточного:

амэ но фуру ёру vs. амэ га фуру
дождь Gen падать ночь дождь Subj падать
 ‘ночь, когда идет дождь’ Дождь идет’.

Однако в целом оформление зависимых в этом языке является надежным показателем того, подчинены они глаголу или существительному.

Что касается лексических функций, японский язык вполне подтверждает гипотезу об их универсальности, высказанную в [МСТ, 109], демонстрируя большой репертуар лексики, подходящей под определение *Oper_i*, *Func_i* и *Labo_{ij}*.

Однако прежде чем приступить к рассмотрению конкретного материала, следует остановиться на одной важной теоретической проблеме, возникающей при использовании аппарата ЛФ.

Данная проблема не является специфической для японского языка, а коренится в самом определении семейства функций *Oper_i*, *Func_i* и *Labo_{ij}*: это критерий, по которому глаголы в конкретных словосочетаниях относятся к тому или иному типу.

Речь не идет о том, что один и тот же глагол по отношению к разным существительным может выступать в разных качествах. Если русский глагол *получать* — *Oper₁*(распространение), но *Oper₂*(совет), а японский глагол *какару* ‘приходить в контакт с чем-либо’ — *Func₂*(*утагаи* ‘подозрение’), но *Oper₂*(*вана* ‘ловушка’), это ничему не мешает, поскольку ЛФ-глаголы описываются не сами по себе, а всегда в паре с аргументом. Хуже, что тип функции зависит не только от лексемы аргумента, но и от того, как описана семантика этого аргумента, конкретно от того, в каком порядке пронумерованы валентности существительного: *получать* — *Oper₂*,

а не $\text{Орег}_3(\text{совет})$ только потому, что получатель совета — второй актанта, а не третий.

В самой модели «Смысл — Текст» принцип нумерации валентностей не эксплицирован. Однако понятно, что номера не могут приписываться в произвольном порядке: тогда не было бы смысла идентифицировать группы ЛФ-глаголов, схожим образом взаимодействующих с C_0 , именно по номерам валентностей.

В лексикографической практике, опирающейся на идеи модели «Смысл — Текст», существует традиция считать первым актанта тот, который реализуется синтаксически как подлежащее, а вторым — реализующийся как прямое дополнение. Здесь можно усмотреть связь с коммуникативным рангом актантов в понимании [Падучева 2004, 58], то есть с их положением в центре или на периферии зоны внимания говорящего. Синтаксические позиции подлежащего и первого дополнения являются центральными, остальные — периферическими. Соответственно, когда, скажем, у трехвалентного глагола типа *арестовать* валентность на подлежащее получает номер 1, валентность на прямое дополнение — 2, а на косвенное дополнение — 3, актанты в модели управления оказываются упорядочены по своему коммуникативному рангу, от центральных к периферическим.

Но хотя номера актантов и коррелируют с их коммуникативной значимостью, соответствие номеров синтаксическим позициям не является однозначным. Номера идут подряд без пропусков, третьей валентности не бывает, если нет второй; соответственно, валентность контрагента у непереходного глагола типа *сражаться* может получить «центральный» номер 2, хотя синтаксическая позиция соответствующего актанта, предложная группа, не является центральной.

И, конечно, при описании существительных проблема нумерации валентностей стоит особенно остро, поскольку у них не бывает не только прямых дополнений, но и подлежащих, и все актанты имеют примерно одинаково низкий коммуникативный ранг.

Так, в качестве примера редкой функции Labor_{21} в [МСТ, 93] приводится сочетание *ставить целью*: *Иван ставит себе целью вычислить интеграл*, вычисление интеграла трактуется как первый актанта слова *цель*, а Иван — как второй. Но достаточно описать *цель* по образцу, скажем, семантически близкого глагола *стремиться*, и содержание окажется вторым актанта, субъект — первым, а глагол *ставить*, соответственно, попадет в более обширный класс Labor_{12} .

В задачи данной статьи не входит решение этой проблемы. Однако указать на нее необходимо, поскольку конечной целью при исследовании переноса валентностей от C_0 к ЛФ-глаголам являются

утверждения о валентностях, а валентности идентифицируются по номерам.

Материалом данного исследования послужил экспериментальный корпус научных текстов — сборник «Цу:яку кэнкю:» («Исследования перевода»), № 2, 2000 г., доступный в виде электронного конкорданса по адресу http://ysomeya.hp.infoseek.co.jp/index_j.html

В этом корпусе были рассмотрены все вхождения трех ЛФ-глаголов: *оконау* ‘проводить, осуществлять’ (более 120 примеров), *укэру* ‘получать’ (более 40 примеров) и *мотиуру* ‘использовать’ (девять примеров).

Кроме того, для двух из этих глаголов, *оконау* и *укэру*, сочетаемость с несколькими предикатными существительными была рассмотрена с помощью поисковой системы Google в сети Интернет (www.google.co.jp).

ОКОНАУ ‘проводить, осуществлять’

Данный глагол подчиняет C_0 как прямое дополнение, а в качестве подлежащего всегда реализует его первую валентность, то есть принадлежит к типу $Oper_1$:

тё:ся га оконатта тё:са
 автор Subj проводить:пв исследование
 ‘исследование, которое провел автор’

В 24 из рассмотренных примеров в корпусе глагол стоит в форме пассива; при этом C_0 из прямого дополнения превращается в подлежащее, то есть форма функционирует как $Func_1$:

dialogue condensation га оконаварэтэ_иру
dialogue condensation Subj проводить:пасс.ДВ
 ‘Проводится **сжатие** диалога’

В 17 случаях C_0 выступает для глагола не как зависимое слово, а как синтаксический хозяин — вершина определительного придаточного:

сювацу:якуся га дзиссай_ни цу:яку гэмба дэ
 жест переводчик Subj реально перевод место в
оконаттэ_иру сагё:
 проводить: ДВ деятельность
 ‘**деятельность**, которую реально осуществляет сурдопереводчик в ситуации перевода’

Первая валентность C₀, 29 примеров, реализована почти исключительно подлежащим глагола:

Blakemore *ва* *but* *ни_цуйтэ* *до:ё:* *но* *бунсэки* *о*
Blakemore *Top* *but* *насчет* *такой_же* *Gen* *анализ* *DObj*
оконаттэ_иру
 проводить:ДВ
 ‘Блэйкмор аналогичным образом **анализирует** [слово] *but*’

Сюда же можно, вероятно, отнести большую часть случаев, где эта валентность не реализована вовсе, поскольку для японского языка характерен эллипсис тематического подлежащего. Встретилось два примера, где первый актант выступает как вершина определительного придаточного:

юби *модзи* *я* *кандзи* *но* *сакуё:* *о*
 палец буква и иероглиф *Gen* **заимствование** *DObj*
оконаттэ_иру *моно*
 проводить:ДВ *лицо*
 ‘лица, которые осуществляют **заимствование** жестов-букв и иероглифов’

Примеров, где бы первая валентность осталась при C₀, не обнаружено.

Вторая валентность C₀, 24 примера, в основном реализована при C₀ как генитивное дополнение:

кё:дзай *но* *тикудзи* *цу:яку* *о* *оконатта*
учебный материал *Gen* *следование* **перевод** *DObj* *проводить:пв*
 ‘...осуществляли последовательный **перевод** учебного материала’

При глаголе этот актант встретился всего пять раз, причем оформленный сложными отглагольными показателями типа *ни_цуйтэ*:

цу:якуся *гото* *но* *токутё:* *ни_цуйтэ* *бунсэки* *о*
 переводчик каждый *Gen* *особенность* *насчет* **анализ** *DObj*
оконатта
 проводить:пв
 ‘[Мы] провели **анализ** особенностей каждого переводчика’.

Третья валентность C₀, когда она есть (10 примеров), также реализуется обычно при самом C₀:

со:_йу *моно* *то* *но* *маттингу* *о* *оконаттэ*
 такой *вещь* *с* *Gen* **сопоставление** *DObj* *проводить:дс*
 ‘проведя **сопоставление** с такими объектами...’

УКЭРУ ‘получать’

Данный глагол подчиняет C_0 как прямое дополнение, в качестве подлежащего реализует преимущественно адресатную валентность C_0 ; если считать ее второй, относится к типу Op_{eg2} :

дзюко:сэй В га ... ко:си кара **фи:добакку** о
 студент В Subj преподаватель от **отзыв** DObj
 укэта дзирэй
 получать:пв пример
 ‘пример, где студент В ... получил от преподавателя **отзыв**’

Примеров, где бы глагол стоял в пассиве, не встретилось.

Первая валентность C_0 реализована в 17 случаях, которые разделились примерно поровну — восемь при существительном, девять при глаголе; характерны примеры, которые в остальных отношениях дословно совпадают:

дзюко:сэй кара (но) **фи:добакку** о укэта ато
 студент от (Gen) **отзыв** DObj получать:пв после
 ‘после того, как [он] получил **отзывы** от студентов’

Вторая валентность C_0 , 14 примеров, ни разу не была реализована при существительном — в пяти случаях этот актант выступает в качестве вершины определительного придаточного, в остальных подчинен глаголу как подлежащее:

нихонго кё:ику о укэтэ_иру гайкокудзин
 японский_язык обучение DObj получать:ДВ иностранец
 ‘иностранцы, которые проходят **обучение** японскому языку’

дзибун га цу:яку кунрэн о укэта токи
 сам Subj перевод обучение DObj получать:пв время
 ‘когда я сам проходил **обучение** переводу’

Третья валентность C_0 , 13 примеров, реализуется преимущественно при существительном примыканием, как именное сцепление, см. выше цу:яку кунрэн «обучение переводу»; встретился также один пример генитивного дополнения при C_0 :

сюппанбүцу но со:фу о укэру кото га дэкиру
 издание Gen присылка DObj получить факт Subj возможный
 ‘есть возможность **подписаться** на издания’ (букв. ‘получить присылку изданий’)

МОТИИРУ ‘использовать’

Данный глагол подчиняет S_0 как косвенное дополнение, в качестве подлежащего реализует его первую валентность, в качестве прямого дополнения — преимущественно валентность на инструмент; если считать ее третьей, то принадлежит к типу Labor₁₃.

Встретилось только девять примеров *мотиуру* в качестве ЛФ-глагола. Во всех оказалась не реализована **первая валентность** S_0 — произошел эллипсис тематического подлежащего, — и реализована **третья**: в семи случаях как прямое дополнение глагола и в двух как вершина определительного придаточного:

хо:тэй бидэо о *мотиита* *кунрэн*
суд видео DObj использовать:пв **обучение**
‘**обучение**, в котором использованы съемки судебных заседаний’

хон *кэнкю:* *дэ* *мотиита* дайдазай
этот **исследование** в использовать:пв данные
‘данные, которые [мы] использовали в настоящем **исследовании**’

Для двух из ЛФ-глаголов, *оконау* и *укэру*, была также рассмотрена с помощью поисковой системы Google сочетаемость с несколькими трехвалентными существительными. Сеть Интернет нельзя использовать в качестве корпуса для точного статистического исследования, и есть опасность встретить отдельные неграмматичные примеры, однако поиск в Интернете дает достаточно наглядную общую картину.

Оба глагола в целом показали то же синтаксическое поведение, что на материале экспериментального корпуса.

1-я валентность S_0 для *оконау* и 3-я для *укэру* реализуется почти исключительно при глаголе — подлежащим или вершиной определительного придаточного.

2-я валентность чаще реализуется при S_0 , примыканием (в виде именного сцепления) или как генитивное дополнение. Встречается только один глагольный способ реализации этой валентности — вершина определительного придаточного:

бэй *кайгун* *кара* *тайё* *о* *укэру* кансэн
США ВМФ от **заём** DObj получить воен. корабль
‘военные корабли, получаемые **в займы** от ВМФ США’

Обнаружились также единичные, но весьма интересные примеры, аналогов которым не было в экспериментальном корпусе. В них ЛФ-глагол управляет сразу двумя прямыми дополнениями — S_0 и его вторым актантом:

хои:ру но хаба надо о рэнраку о укэ
 фольга Gen ширина прочее DObj сообщение DObj получить:сс
 ‘получаем сообщение о ширине фольги и прочем, и...’

Такие употребления, возможно, говорят о том, что сочетания «С₀ + о + ЛФ-глагол» начинают рассматриваться некоторыми носителями как целостная единица с глагольным поведением.

Оставшаяся валентность С₀, 3-я для *оконау* и 1-я для *укэру*, реализуется как при С₀, так и при глаголе; для большинства глагольных способов реализации есть аналогичные именные:

до:куцу га ару то омоварэру басё дэ **тансаку** о
 пещера Subj быть Quot думать:пз место в **поиск** DObj
оконау *то* **vs.**
 проводить когда
 ‘когда [мы] провели **поиск в местах**, где можно предположить
 наличие пещер’

кимпэн дэ но **тансаку** о *оконаи*
окрестности в Gen **поиск** DObj проводить:сс
 ‘провели **поиск в окрестностях**, и...’

Именные варианты реализации кажутся несколько менее частотными, но точно установить это не позволяют свойства сети Интернет — одни и те же тексты в ней могут встретиться в нескольких экземплярах, искажая статистику.

Уже из этих предварительных данных можно извлечь некоторые закономерности.

При существительном не остается та валентность, которая соответствует подлежащему ЛФ-глагола. Вероятно, дело в том, что позиция подлежащего имеет более высокий коммуникативный ранг, чем варианты реализации валентности, доступные существительному: генитивная группа или член именного сцепления. С точки зрения глагола реализовать соответствующую валентность при существительном значило бы «понизить» участника с центрального положения до периферийного, и тогда оказалось бы неоправданным использование такого ЛФ-глагола, который ориентирован именно на этого участника.

Заслуживает особого упоминания тот факт, что подлежащее *укэру*, Орег₂, реализуется несколько чаще, чем подлежащее *оконау*, Орег₁ (в корпусе приблизительно 2/5 всех примеров против 1/4). Орег₂ действует аналогично пассивному залогу, «понижая» в ранге

первого участника ситуации C_0 и одновременно «повышая» второго участника до позиции подлежащего — не исключено, что дополнительную долю примеров составляют как раз те случаи, когда второй актант коммуникативно важен, а выбор $Oper_2$ вместо $Oper_1$ призван поместить его в центр внимания.

Известно, что употребительность разных типов ЛФ в рассматриваемом семействе не одинакова: «если, скажем, $Oper_1$ представлен (в частности, в русском языке) очень широко, то $Oper_2$ имеет выражение для сравнительно небольшого числа слов» [МСТ, 93]. Втрое большая частотность *оконау* в корпусе по сравнению с *укэру* подтверждает это и для японского. Вероятно, тут тоже играет роль коммуникативный ранг актантов: аналогичным образом, активный залог встречается значительно чаще пассивного.

Видно также, что некоторые валентности, напротив, имеют тенденцию оставаться при существительном. Глагол *оконау* редко занимает вторую валентность C_0 , *укэру* — третью. Причину можно усмотреть в том, что глагол не может предоставить соответствующему участнику ситуации подходящую синтаксическую позицию.

Позиция подлежащего у обоих глаголов уже занята — первым актантом существительного у *оконау* и вторым у *укэру*. В позиции прямого дополнения находится само C_0 ; примеры, в которых глагол присоединяет второе прямое дополнение наряду с C_0 , судя по их малому количеству, находятся на грани допустимого. Остается позиция косвенного дополнения, но в случае *укэру* на нее претендует первый актант C_0 («от кого получают»), а у *оконау* она может оказаться зарезервирована для третьего актанта C_0 — например, для инструментальной валентности:

хонъякуся ва оридзинару но сукуруипуто дакэ дэ
переводчик Тор оригинал Ген текст только с

хонъяку о оконану кото ва най

перевод DObj выполнять факт Тор нет

‘Не бывает так, чтобы переводчик выполнял перевод только по исходному тексту’

В результате получается, что вторую валентность проще оставить при самом существительном, что и имеет место в большинстве примеров.

Разумеется, это лишь предварительные результаты, которые нуждаются в уточнении и проверке. Необходимо рассмотреть другие ЛФ-глаголы, в особенности принадлежащие к типу Func₁, и сравнить

их синтаксическое поведение с поведением глаголов типа Opr_1 в форме пассивного залога.

Однако уже сейчас можно заключить, что японский язык действительно предоставляет подходящий материал для изучения переноса валентностей от существительных к ЛФ-глаголам. Можно также ориентировочно говорить о двух тенденциях: та валентность существительного S_0 , которая для глагола становится подлежащим, реализуется при глаголе; та валентность существительного S_0 , которая иначе претендовала бы на синтаксическую позицию, занятую при глаголе самим S_0 , реализуется при существительном.

Использованные грамматические сокращения

DObj — показатель прямого дополнения	ДВ — длительный вид
Gen — показатель генитива	дс — деепричастный статус
Subj — показатель подлежащего	сс — срединный статус
Top — показатель топика	пв — прошедшее время
Quot — показатель цитации	пз — пассивный залог

Литература

- [МСТ] — Мельчук И. А. Опыт теории лингвистических моделей «Смысл — Текст». М.: Наука, 1974.
- [Падучева 2004] — Падучева Е. В. Динамические модели в семантике лексики. М.: Языки славянской культуры, 2004.

Diathesis Alternations and NP Semantics *

Barbara H. Partee

University of Massachusetts, Amherst;
Russian State Humanitarian University;
and Moscow State University
partee@linguist.umass.edu

In this paper we examine the relationships among diathesis alternations, the semantics of verbs, and the referential status of NPs. Some diathesis shifts are argued to involve changes in the semantic type of NP arguments, including possible alternations between ‘referential’ NPs and ‘property-type’ NPs. We explore applications of this approach to alternations of Genitive and Accusative, both with intensional verbs like *ždat* ‘wait’ and with the Genitive of Negation.

Диатезные чередования и семантика именных групп. Аннотация.

В работе исследуются связи между диатезными чередованиями, семантикой глаголов и референциальным статусом именных групп. Показано, что некоторые диатетические сдвиги влекут за собой изменения в семантических типах именных групп, являющихся аргументами данных предикатов, в том числе возможные флуктуации между референциальными именованными группами и именными группами, обозначающими свойства. Мы исследуем возможности применения этого подхода к чередованиям родительного и винительного падежей при интенциональных глаголах типа *ждать* и к родительному падежу при отрицании.

1. Examples and issues

For the purposes of this paper, I take the notion of diathesis alternation in a broad sense, including both familiar alternations of syntactic patterns

* This material is based upon work supported by the National Science Foundation under Grant No. BCS-0418311 to Barbara H. Partee and Vladimir Borschev for a project entitled, “The Russian Genitive of Negation: Integration of Lexical and Compositional Semantics.” I thank Vladimir Borschev, Elena Paducheva, Ekaterina Rakhilina, Yakov Testele, and the students in my classes at RGGU and MGU during my Fulbright fellowship in Spring semester 2005 for valuable discussion. All inadequacies are my own.

as in examples (1a-b) and alternations of case assignment, as in examples (2a-b) and (3a-b).

- (1) a. The farmers loaded the truck with (*?the/*?some) hay.
 b. The farmers loaded (the/some) hay on the truck.
- (2) a. *On ždet podругu.* (Neidle 1988, p. 31)
 He waits girlfriend-ACC
 ‘He’s waiting for his girlfriend.’
 b. *On ždet otveta na vopros.*
 He waits answer-GEN to question
 ‘He’s waiting for an answer to the question.’
- (3) a. *On ne polučil pis'mo.*
 he NEG received letter-ACC.N.SG
 ‘He didn’t receive the letter.’
 b. *On ne polučil pis'ma.*
 he NEG received letter-GEN.N.SG
 ‘He didn’t receive any letter.’

In all these cases, my interest is in the interaction between the lexical semantic interpretation of the verb in its various diathesis frames and the semantic interpretation of noun phrase (NP) arguments that appear in those frames. I see this question as part of the larger issue of the interaction of lexical semantics with compositional semantics, a concern that has been at the heart of my joint work with Vladimir Borschev (Borschev and Partee 1999, 2002, Partee and Borschev 2003).

In (1a-b) we see a preference for the direct object in either diathesis frame to be definite, but the difference is most pronounced for the role which is ‘Means’ in (1a) (*sredstvo* in Apresjan (1974, 126–129)) and ‘Patient’ in (1b), which is almost obligatorily non-referential in (1a) but may be indefinite or definite in (1b). In (2a–b) there is a clear difference in referential status of the direct object argument in the two cases, and a correlated difference in the sense of the verb. The difference in referential status of the object in the two sentences (3a–b) is similar to the difference in (2a–b); in this case the verb itself does not seem different in the two sentences, but we can raise the question of whether and how the negation together with the verb might have a semantic effect analogous to the effect of the ‘intensional’ reading of the verb in (2a–b), a possibility that was raised by Neidle (1988) and by Partee and Borschev (2004). The questions of concern in this paper are to what extent the semantic interpretation of the NP is part of, or is affected by, the semantics connected with the diathesis alternations seen in such pairs, and how the semantics of the diatheses are connected with the lexical semantics of the verb.

There is not a sharp line between diathesis alternation and instances of a correlation between different senses of a verb and different interpretations of an NP argument, as in (4a—b), where the verb in (4a) denotes a relation between two entities and in (4b) a relation between an entity (the agent) and some intensional object, perhaps an intensional generalized quantifier as proposed by Montague (1973), perhaps a property as argued by Zimmermann (1993), but in either case not simply an entity.

- (4) John is looking for a blue Volkswagen.
 a. ... He can't remember where he parked it.
 b. ... If he finds one, he will buy it for his wife.

While examples like (4a—b) in English are not normally classed as diathesis alternation, the interaction of the interpretations of the verb and of the direct object argument is similar in that case to the interaction seen in (2a—b), where the difference is overtly marked in a choice of case on the object.

Filip has explored in (Filip In press) some cases that were first discussed in Partee (1991, 1995), in which verbal affixes together with diathesis shifts can be used to express various kinds of quantificational and closely related meanings like measure, distributivity, totality, exclusivity or exhaustiveness. One illustration is the use of the Czech prefix *po-*, discussed in Partee (1991, 1995) and illustrated here by examples (5a—b) from Filip (In press):

- (5) a. Maloval hesla (na stěnu). Czech
 paint.PAST.IMPF.3SG slogan.PL.ACC (on wall)
 ‘He painted (the/some) slogans (on the wall).’
 b. PO-maloval stěnu hesly.
 TOT-paint.PAST.PF.3SG wall.SG.ACC slogan.PL.INST
 ‘He covered the wall with slogans.’
 c. *PO-maloval hesla na stěnu.
 TOT-paint.PAST.PF.3SG slogan.PL.ACC on wall.SG.ACC
 *‘He covered (the/some) slogans on the wall.’

When the prefix *po-* is applied to an imperfective verb with a meaning like ‘write’, ‘draw’, etc., which takes an object of creation (5a), the resulting perfective verb in (5b) takes as its direct object the optional locative complement of the base verb, and the direct object of the base verb is demoted to an optional instrumental complement of the prefixed verb. The meaning of the perfective verb *pomaloval* in (5b) is ‘he painted all over X’ or ‘he covered X with painting’. Hence, the prefix *po-* is in a certain sense quantificational but at a lexical rather than a syntactic level.

Of particular relevance to the issues in this paper is the fact that while the bare NP *hesla* ‘the/some slogans’ can be interpreted as either definite or indefinite when it occurs as direct object of the imperfective verb in (5a), the same NP in its instrumental form in (5b) can only be indefinite, and in a sense is even more indefinite than the indefinite interpretation it may have in (5a).

In most studies by formal semanticists of connections between verb semantics and the semantics of NP arguments of verbs, the emphasis has been on quantificational properties, especially on the relation between aspectual properties of verbs (particularly the telic/atelic distinction) and the count/mass distinction, or more accurately a “quantized/non-quantized” distinction (Krifka), as studied in the work of Bach (1986), Dowty (1991), Krifka (1992), Filip (1999), and Partee (1997, 1999). Other kinds of connections have been made in the work of Padučeva, who has drawn correlations between decreased assertiveness (*snjataja utverditel’nost’*) in sentences, decreased referential status in NPs (Padučeva 2004, 328–329), and possibilities of diatheses in non-assertive moods that are not possible for a given verb in assertive mood. Yanko (2001) has studied a number of cases in which diathesis alternations correlate with differences in Theme-Rheme structure, a point of view which has recently gained some prominence in the work of Levin and Rappaport Hovav (2002, In press). Formal semanticists have recently begun seeking arguments to distinguish between diathesis alternations that carry genuine semantic distinctions and those that may carry only Theme-Rheme structure distinctions (Krifka 2004).

More generally, as formal semanticists try to formulate more adequate theories integrating lexical meaning and compositional meaning, one of the problems we face, undoubtedly already well known among those who have studied lexical semantics for longer than I have, is the difficulty of sorting out the contributions of specific lexical meanings, of specific syntactic configurations, and of general principles such as the preference for subjects to be topics and for topics to be definite. It is noteworthy that the properties of subjects tend to cluster together more strongly than the properties of direct objects and other arguments. Subjects tend to be topic, given, definite, referential, and well individuated (quantized); there is probably no other identifiable argument position that has such strongly clustered properties, perhaps because there is no other argument or adjunct which nearly obligatorily appears in every sentence. Direct objects are less uniform in their clustering of properties, probably because in some respects they are very similar to subjects — those are the two “structural arguments”, higher ranking in significant respects than all of the oblique arguments, which in turn are higher ranking than adjuncts.

A good example of the difficulty of sorting out such factors can be seen in recent debates concerning the semantics and pragmatics of the Dative Alternation in English (6a—b). Just as the choice of passive is often motivated by an intention to make the underlying direct object the Theme or Topic of the sentence, the choice of the double object construction as in (6b) has been argued by Levin and Rappaport Hovav (2002) and by Bresnan (In press) to be motivated in large part by an intention to make the Experiencer (the “dative” argument) the Theme.

- (6) a. Ann sold the car to Beth
 NP₀ V NP₂ to NP₁
 b. Ann sold Beth the car
 NP₀ V NP₁ NP₂

Krifka (1999) originally argued for a semantic account of the Dative Alternation in (6a-b). In his most recent work (Krifka 2004), he still argues that there is a basic semantic difference between the two diatheses, but acknowledges that pragmatic factors can sometimes override the semantic factors. Bresnan’s account (Bresnan et al. In press), in an Optimality Theoretic framework, is in some ways similar, allowing for a number of different factors with ranked weights to play a role in the choice between the two diatheses. The claim that the semantic difference between the two constructions is weak or non-existent may find support in the fact that in the Slavic languages there are not two diatheses with these double-object verbs but simply a choice of word order presumably reflecting a difference in Theme-Rheme structure.

Krifka’s analysis provides an interesting attempt to factor apart the contributions of the lexical item and of its diathesis frame. Omitting many important details, we can say that on Krifka’s analysis, the lexical semantics of many of the verbs that occur in both frames of (6a-b) is a manner of action; the construction in (6a) contributes an entailment that NP₀ caused NP₂ to go to NP₁, while the construction in (6b) contributes an entailment that NP₀ caused NP₁ to have NP₂. Some verbs, like *give* and *sell*, have so much information in their lexical semantics that the constructions contribute nothing new, and the sentences in the two frames end up truth-conditionally equivalent; in this case, it is especially likely that differences in Theme-Rheme structure will be the most salient differences. But in other cases, as with different sorts of verbs of causation of motion, like *throw* vs. *push*, the semantic difference between the two frames may be quite noticeable. And most relevant to the topic of this paper, we can see the grounds for a prediction of a difference in referential status to arise in certain cases. Consider the two subparts of the interpre-

tations, “NP₂ to go to NP₁” vs. “NP₁ to have NP₂”. In the first of these, NP₂ must evidently exist throughout the action; but in the second, NP₂ could in principle come into existence as a result of the action. And Krifka indeed discovers some cases where this difference results in only one of the two diatheses being possible with some NPs.

- (7) a. The explanation gave Susan a headache.
 b. #The explanation gave a headache to Susan.

The NP *a headache* is not referential and may have a “property interpretation” as discussed in the next section. (The referential status of NPs whose referents come into existence as a result of a stated action is a difficult question that has not received enough study so far.) In the case of the Dative Alternation, we see the situation that the semantics of the two diatheses permits cases in which the referential status of NP₂ differs, but does not require them to differ; fully referential NP₂ is possible in both frames, while non-referential NP₂ is possible only in construction (6b).

In the second half of the paper, we look more closely at the semantics of indefiniteness before returning to its interaction with diathesis alternations and verbal semantics.

2. Definiteness, indefiniteness, and referentiality.

Within formal semantics there has been a great deal of work on NP interpretations, including much work on quantification, on definiteness and indefiniteness, on kind-denoting NPs, on property-denoting NPs, on type-shifting among different possible interpretations of NPs, and recently on semantic typology and different kinds of indefinite and quantificational NPs.

Classic formal semantics adopted Montague’s proposal (Montague 1973) for the semantics of Noun Phrases (NPs). Every NP was interpreted as denoting a Generalized Quantifier, i.e. a set of sets, type $\langle\langle e, t \rangle, t \rangle$ (strictly, a set of properties, type $\langle\langle\langle s, \langle e, t \rangle \rangle, t \rangle\rangle$.) Some NP interpretations on Montague’s analysis are as shown in (8).

(8) <i>John</i>	$\lambda P[P(j)]$	(the set of all of John’s properties)
<i>every student</i>	$\lambda P \forall x[\text{student}(x) \rightarrow P(x)]$	(the set of all properties that every student has)
<i>a student</i>	$\lambda P \exists x[\text{student}(x) \ \& \ P(x)]$	(the set of all properties that at least one student has)

the king $\lambda P [\exists x[\mathbf{king}(x) \ \& \ \forall y (\mathbf{king}(y) \rightarrow y = x)] \ \& \ P(x)]$
 (the set of properties which the one and only king has)

What could “indefinite” mean in such a framework? The need for such a distinction first arose in discussions of the prohibition of “definite” NPs in English existential *there*-sentences: see the contrast between the acceptable sentences in (9) and the sentences in (10), which are anomalous without special contexts.

- (9) a. There is a new problem.
 b. There are three/many/several/few/no semantics textbooks.
 (10) a. #There is every/neither/the linguistics student.
 b. #There are most/both/the three democratic governments.

No intuitive notion of “definite” vs. “indefinite” explains why some quantifiers, like *three* and *many*, pattern with indefinites while others, like *every*, *most*, and *both*, pattern with definites. Milsark (1974, 1977) provided the kernel of a semantic explanation, dividing determiners into “weak” and “strong”, which was further developed by Barwise and Cooper (1981) and by Keenan (1987).

Definition (Keenan 1987): A determiner D is a *basic existential determiner* if for all models **M** and all $A, B \subseteq E$, $D(A)(B) = D(A \cap B)(E)$.

The corresponding natural language test is as follows: “Det CN VP” is true iff “Det CN which VP exist(s)” is true. A determiner D is *existential* if it is a basic existential determiner or it is built up from basic existential determiners by Boolean combinations (*and*, *or*, *not*). By this definition, it turns out that the existential, or weak, determiners are exactly the symmetric determiners. The determiners *three*, *a*, *some*, *no*, *at least three*, *exactly three*, *at most three* are all weak. The determiners *the*, *the three*, *every*, *both*, *most*, *neither* are all strong. These semantic definitions laid a successful groundwork for a great deal of further research on the semantics of weak and strong NPs and their distribution (Büring 1996, de Hoop 1995, Partee in press, Rullmann 1989).

While some differences in the possible occurrence of ‘weak’ and ‘strong’ NPs can be accounted for by drawing semantic distinctions within the theory of generalized quantifiers, as in the account above, it has been argued that in some cases, weak NPs are really of “property type” (Montague’s type $\langle s, \langle e, t \rangle \rangle$, functions from possible situations to sets of entities), rather than generalized quantifiers. Property-type analyses of various “weak NPs” are becoming increasingly common in Western formal semantics, and they may have an important role to play in ac-

counting for non-referential readings of NPs in various diatheses, possibly including the Russian Genitive of Negation.

Zimmermann (1993) argued that Montague's analysis of verbs like *seek* ("intensional transitive verbs", or "opaque verbs") as taking arguments of type "intension of Generalized Quantifier", or $\langle s, \langle s, \langle e, t \rangle \rangle, t \rangle$ is incorrect. He argues that the NP objects of opaque verbs should be semantically interpreted as *properties* (or type $\langle s, \langle e, t \rangle \rangle$.) He starts by comparing the sentences in (11) and (12).

- (11) Caroline found a unicorn. (extensional, unambiguous)
- (12) Caroline sought a unicorn. (intensional, ambiguous)

He notes that sentences with *seek* are ambiguous between a specific and a non-specific reading (or transparent vs. opaque reading). Sentence (11) is unambiguous, (12) is ambiguous. Furthermore, on the opaque reading of (12), the existence of a unicorn is not entailed. And a most important property is that substitution of extensionally equivalent expressions in an intensional context (on the opaque reading) does not always preserve truth-value. E.g., the extension of *unicorn* is the same as the extension of *13-leaf clover* (both are the empty set in the actual world). Substituting *a thirteen-leaf clover* for *a unicorn* in (11) preserves truth-value. The same substitution in (12) might not. Other intensional verbs like *seek* are *owe*, *need*, *lack*, *prevent*, *resemble*, *want*, *look for*, *request*, *demand*.

Quine (1960) had argued that *seek* should be decomposed into *try to find*. He argued that intensionality is (in general) the result of embedding under an attitude verb like *try*. Within *Caroline try [Caroline find x]*, there are then two places a quantifier phrase could take its scope: the higher clause, giving the transparent reading, and the lower clause, giving the opaque reading.

Montague (1973) achieved the same semantic effect with a simpler syntax, letting NPs express Generalized Quantifiers and functions apply to the *intensions* of their arguments. Then the transitive verb *seek* can apply directly to its NP argument without decomposition. So Montague treats a verb like *seek* as denoting a relation between an individual and an intensional generalized quantifier. The transparent reading results from "quantifying in".

But Zimmermann (1993) points out problems with both Quine's and Montague's classical analyses, including an *overgeneration* problem, in that true quantifier phrases are unambiguously "transparent" after inten-

sional transitive verbs like *seek*, although they are ambiguous in constructions like *try to find*. Simple indefinites with *a*, on the other hand, are ambiguous with intensional verbs. Compare:

- (13) a. Alain is seeking a comic book. (ambiguous)
 b. Alain is seeking each comic book. (unambiguous; lacks ambiguity of (c))
 c. Alain is trying to find each comic book. (ambiguous).

Zimmermann argues that we can capture the relevant generalizations once we note that definites and indefinites, which do receive opaque readings with intensional verbs, correspond, in a way he makes precise, to properties, type $\langle s, \langle e, t \rangle \rangle$. Zimmermann's proposal is that a verb like *seek* denotes a relation between an individual and a property. So *seek a unicorn* would be interpreted as (14)

- (14) *seek'* ($\wedge$ 'unicorn') (where $\wedge$ is Montague's 'intension operator')

This is a case of NP type-shifting by coercion: *seek* on its intensional reading demands a property-type argument. There is independent evidence that indefinite NPs easily shift into predicative property readings of this type, as was shown for predicate nominals and the PRED-argument of *consider* in Partee (1986). For the transparent, or specific, or de re, reading, Zimmermann gives an analysis (details omitted here) involving "quantifying in", similar to the analysis in Partee (1986) for Edwin Williams' example "This house has been every color". Zimmermann thus has a solution to the overgeneration problem, and to other problems which he raises in his article.

3. NP interpretations and diatheses

If both Generalized Quantifier readings of indefinites and property-type readings are available in principle, how can we identify which is the right analysis in a given instance? Following Zimmermann's lead, one criterion should be whether true quantifiers like *each student*, *most students* can occur in the given construction: if so, then we are dealing with an e-type argument position into which we can quantify; if not, then probably with a property-type argument.

On this criterion, the existential construction has a property-type NP position, as independently argued by McNally (1998). And following Chvany (1975), Freeze (1992), and others, we can argue that there is a diathesis shift for *be* and other 'existential' verbs, which can occur both in

ordinary predication sentences with a referential, type e subject (15a), and in existential sentences, where the NP is in a non-referential property-type argument position (15b).

- (15) a. The cat is in the kitchen.
b. There is a cat in the kitchen.

As for the other constructions we have looked at, there is not perfect consensus about whether they permit quantifier arguments or not. In the case of the *load* construction in English (1a-b), in English they seem to allow quantifiers to occur in either argument position (but the Russian near-equivalent *zagruzit*, discussed in Padučeva (2004, p. 59, p. 65) apparently does not (E. V. Padučeva, p. c).) But on closer examination, even English may allow only ‘measure’ quantifiers with the *sredstvo* argument when it is oblique, suggesting that that argument in that construction may have a property reading. The demoted object argument of *po-malovat* in (5b), which also becomes something like *sredstvo*, also strongly resists quantificational objects.¹

In the case of the potentially intensional verb *ždat* in (2a-b), we see that its intensional genitive-taking variant in (2b) has all the properties of English *seek*. Its extensional accusative-taking variant, (2a), allows referential NPs and quantificational NPs. We predict that genitive should be disallowed with essentially quantificational NPs such as those formed with *každyj*; the question is open.

If the Genitive of Negation on objects is associated with a difference in semantic interpretation as discussed in Borschev and Partee (2002), we would predict that it should also exclude essentially quantificational NPs; but this prediction is false, as shown by (16).

- (16) On ne rešil vsech zadač. ‘He didn’t solve all the problems-GEN.’

This result does not by itself eliminate the hypothesis that Genitive of Negation induces property-type readings on affected NPs, because of the notorious multiplicity of factors affecting its distribution. But it does sug-

¹ One must exercise caution in testing for the possibility of ‘essentially quantificational’ arguments; many determiners have both quantificational and non-quantificational interpretations, and even unambiguously quantificational determiners like *each* may give ‘false positives’ if combined with kind-denoting nouns. Thus the possibility of (i) is not a real counterexample to the claim that the *sredstvo* argument in (i) cannot be quantificational.

- (i) John covered the wall with each color of paint. (OK if he painted it many times.)

gest that the decreased referentiality often remarked with Genitive of Negation is not as extreme as that observed with intensional verbs like *ždat'*, and casts some doubt on the hypothesis.

In conclusion, it seems clear that certain diathetic alternations are naturally associated with changes in the referential status of certain NP arguments or adjuncts. Further investigation is needed to find clearer criteria to distinguish among different hypotheses about the exact nature of "NP weakening" as a result of different kinds of diathetic alternations. We have made some slight progress in examining the domain of "property-type NPs", but much more needs to be done.

References

- Apresjan, Jurij D. 1974. *Leksicheskaia Semantika. Sinonimicheskie Sredstva Jazyka*. Moscow: Nauka, 1974.
- Bach, Emmon. 1986. The algebra of events. *Linguistics and Philosophy* 9:5—16. Reprinted in Paul Portner and Barbara H. Partee, eds., *Formal Semantics: The Essential Readings*, Oxford: Blackwell, 1986. P. 324—333.
- Barwise, Jon, and Cooper, Robin. 1981. Generalized quantifiers and natural language. *Linguistics and Philosophy* 4. 1981. P. 159—219. Reprinted in Portner and Partee, eds., 2002. P. 75—126.
- Borschev, Vladimir, and Partee, Barbara H. 1999. Semantika genitivnoj konstrukcii: raznye podxody k formalizacii. In *Tipologija i teorija jazyka: Ot opisanija k ob'jasneniju. K 60-letiju Aleksandra Evgen'evia Kibrika*, eds. Ekaterina V. Rakhilina and Yakov G. Testelefs. 1999. P. 159—172. Moscow: Jazyki Russkoj Kul'tury.
- Borschev, Vladimir, and Partee, Barbara H. 2002. The Russian genitive of negation in existential sentences: the role of Theme-Rheme structure reconsidered. In *Travaux du Cercle Linguistique de Prague (nouvelle série)*, eds. Eva Hajičová, Petr Sgall, Jirí Hana and Tomáš Hoskovec, 185-250. Amsterdam: John Benjamins Pub. Co., 2002.
- Bresnan, Joan, Cueni, Anna, Nikitina, Tatiana, and Baayen, Harald. In press. Predicting the Dative Alternation. In *Proceedings of the Royal Netherlands Academy of Science Workshop on Foundations of Interpretation*. Amsterdam.
- Büring, Daniel. 1996. A Weak Theory of Strong Readings. *SALT* 6. 1996.
- Chvany, Catherine V. 1975. *On the Syntax of BE-Sentences in Russian*. Cambridge, MA: Slavica, 1975.
- de Hoop, Helen. 1995. On the characterization of the weak-strong distinction. In *Quantification in Natural Languages*, eds. Emmon Bach, Eloise Jelinek, Angelika Kratzer and Barbara H. Partee. P. 421—450. Dordrecht: Kluwer, 1995.
- Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67. 1991. P. 547—619.

- Filip, Hana. 1999. *Aspect, Eventuality Types and Nominal Reference*. New York: Garland, 1999.
- Filip, Hana. In press. Measures and indefinites. In *Reference and Quantification: The Partee Effect*, eds. Gregory N. Carlson and Francis Jeffry Pelletier. Stanford: Center for the Study of Language and Information.
- Freeze, Ray. 1992. Existentials and other locatives. *Language* 68. 1992. P. 553—595.
- Keenan, Edward. 1987. A Semantic Definition of “Indefinite NP”. In *The Representation of (In)definiteness*, eds. Eric Reuland and Alice ter Meulen. P. 286—317. Cambridge, MA: MIT Press, 1987.
- Krifka, Manfred. 1992. Thematic relations as links between nominal reference and temporal constitution. In *Lexical Matters*, eds. Ivan Sag and Anna Szabolcsi. P. 29—53. Stanford: CSLI, 1992.
- Krifka, Manfred. 1999. Manner in dative alternation. In *WCCFL 18: Proceedings of the Eighteenth West Coast Conference on Formal Linguistics*, eds. Sonya Bird, Andrew Carnie, Jason D. Haugen and Peter Norquest. P. 1—14. Somerville/Medford, MA: Cascadilla Press, 1999.
- Krifka, Manfred. 2004. Semantic and pragmatic conditions for the Dative Alternation. *Korean Journal of English Language and Linguistics* 4. 2004. P. 1—32.
- Levin, Beth, and Rappaport Hovav, Malka. 2002. Handout Spain: What Alternates in the Dative Alternation? The 2002 Conference on Role and Reference Grammar: New Topics in Functional Linguistics: The Cognitive and Discursive Dimension of Morphology, Syntax and Semantics, Universidad de La Rioja, Logrono, Spain, July 27—28. 2002.
- Levin, Beth, and Rappaport Hovav, Malka. In press. *Argument Realization: Research Surveys in Linguistics*. Cambridge: Cambridge University Press.
- McNally, Louise. 1998. Existential sentences without existential quantification. *Linguistics and Philosophy* 21. 1998. P. 353—392.
- Milsark, Gary. 1974. Existential sentences in English, MIT: Ph.D. dissertation, 1974.
- Milsark, Gary. 1977. Toward an explanation of certain peculiarities of the existential construction in English. *Linguistic Analysis* 3. 1977. P. 1—29.
- Montague, Richard. 1973. The proper treatment of quantification in ordinary English. In *Approaches to Natural Language*, eds. K. J. J. Hintikka, J. M. E. Moravcsik and P. Suppes. P. 221—242. Dordrecht: Reidel. Reprinted in Montague 1974. P. 247—270; Reprinted in Portner and Partee, eds., 2002. P. 17—34.
- Montague, Richard. 1974. *Formal Philosophy. Selected Papers of Richard Montague*. Edited and with an introduction by Richmond H. Thomason. New Haven/London: Yale University Press, 1974.
- Neidle, Carol. 1988. *The Role of Case in Russian Syntax: Studies in Natural Language and Linguistic Theory* 10. Dordrecht: Kluwer, 1988.
- Padučeva, E. V. 2004. *Dinamičeskie modeli v semantike leksiki*. Moscow: Jazyki slavjanskoj kultury, 2004.
- Partee, Barbara H. 1986. Noun phrase interpretation and type-shifting principles. In *Studies in Discourse Representation Theory and the Theory of Generalized*

- Quantifiers*, eds. J. Groenendijk, D. de Jongh and M. Stokhof. P. 115—143. Dordrecht: Foris. Reprinted in Portner and Partee, eds., 2002. P. 357—381. Reprinted in Partee, Barbara H. 2004. *Compositionality in Formal Semantics: Selected Papers by Barbara H. Partee*. Oxford: Blackwell Publishing, 2004. P. 203—230.
- Partee, Barbara H. 1991. Domains of quantification and semantic typology. In *Proceedings of the 1990 Mid-America Linguistics Conference*, ed. Frances Ingemann. P. 3—39: University of Kansas, 1991.
- Partee, Barbara H. 1995. Quantificational structures and compositionality. In *Quantification in natural language*, eds. Emmon Bach, Eloise Jelinek, Angelika Kratzer and Barbara H. Partee. P. 541—602. Dordrecht: Kluwer, 1995.
- Partee, Barbara H. 1997. Vid i interpretacija imennyx grupp (“Aspect and the interpretation of Noun Phrases”). In *Trudy Aspektologicheskogo Seminara Filologicheskogo Fakul'teta MGU im. Lomonosova, Vol. 3*, ed. Marina Chertkova. P. 121—140. Moscow: Moscow University Press, 1997.
- Partee, Barbara H. 1999. Nominal and temporal semantic structure: aspect and quantification. In *Prague Linguistics Circle Papers*, v.3, eds. E. Hajicová, T. Hoskovec, O. Leška, P. Sgall and Z. Skoumalová. P. 91—108. Amsterdam: John Benjamins Pub. Co., 1999.
- Partee, Barbara H. in press. Weak noun phrases: semantics and syntax. In *Proceedings of Dialog 2005*. Moscow.
- Partee, Barbara H., and Borschev, Vladimir. 2003. Genitives, relational nouns, and argument-modifier ambiguity. In *Modifying Adjuncts*, eds. E. Lang, C. Maienborn and C. Fabricius-Hansen. P. 67—112. Berlin: Mouton de Gruyter, 2003.
- Partee, Barbara H., and Borschev, Vladimir. 2004. The semantics of Russian Genitive of Negation: The nature and role of Perspectival Structure. In *Proceedings of Semantics and Linguistic Theory (SALT) 14*, eds. Kazuha Watanabe and Robert B. Young. Ithaca, NY: CLC Publications, 2004.
- Portner, Paul, and Partee, Barbara H. eds. 2002. *Formal Semantics: The Essential Readings*. Oxford: Blackwell Publishers, 2002.
- Quine, Willard van Orman. 1960. *Word and Object*. Cambridge: MIT Press, 1960.
- Rullmann, Hotze. 1989. Indefinite Subjects in Dutch. In *Papers on Quantification*, eds. Emmon Bach, Angelika Kratzer and Barbara Partee: Department of Linguistics, University of Massachusetts at Amherst, 1989.
- Yanko, T. E. 2001. *Kommunikativnye strategii russkoj reči*. Moscow: Jazyki slavjanskoj kultury, 2001.
- Zimmermann, Ede. 1993. On the proper treatment of opacity in certain verbs. *Natural Language Semantics* 1. 1993. P. 149—179.

**Приземлиться и промахнуться:
семантические механизмы
синтаксических ограничений ***

В. А. Плунгян

Е. В. Рахилина

Институт языкознания РАН
plungian@iling-ran.ru

ВИНИТИ РАН

В статье еще раз обращается внимание на такие случаи, когда обязательный участник ситуации не может быть синтаксически выражен при соответствующем предикате. Классическим примером такого рода в МСТ всегда служил глагол *промахнуться*. Показано, что данные ограничения свойственны достаточно большому классу единиц и могут быть объяснены с помощью механизмов взаимодействия партиципантов друг с другом, которое фиксирует значения переменных, превращая их в своего рода «ситуативные константы».

Russian *prizemlit'sja* 'to land' and *promaxnut'sja* 'to miss': semantic reasons for syntactic blocking of the expression of participants. The paper reopens a discussion of such cases when an obligatory participant cannot have a syntactic expression within the corresponding verbal phrase. MTT always used the Russian verb *promaxnut'sja* 'to miss' as a classic example of this kind. We show that the similar restrictions operate for a considerable number of Russian predicate lexemes of different nature. A possible explanation for this type of syntactic behaviour is related to the (poorly elaborated) rules of participant interaction, which make the meaning of semantic variables "fixed" and yield so-called "situational constants".

0. Введение: постановка задачи

Настоящее исследование опирается на недавнюю статью И. А. Мельчука [Mel'čuk 2004], касающуюся теории семантических актантов и ограничений на их поверхностное выражение. В этой статье, в частности, формулируется принцип наследования обязательных участников (партиципантов): если в толкование данной лексемы входят предикатные значения, то все их обязательные партиципанты наследуются

* Настоящая работа частично поддержана грантом NSF № BCS-0418311.

толкуемой лексемой и формируют множество ее обязательных участников (с. 12). Правда, не всякий такой участник может быть поверхностно выражен при лексеме. Одна из причин, как показано там же, на с. 13, связана с «фиксированностью», или полной определенностью партиципанта — когда он в большей степени константа, чем переменная.

В качестве примера И. А. Мельчук предлагает глаголы движения или каузации движения, в которых таким образом фиксирован начальный пункт движения. Нам кажется, хорошей иллюстрацией этого случая является глагол *приземлиться*, где при всех условиях исходной точкой движения является воздушное пространство — константа, которая не может быть выражена поверхностно. Пример, который использует здесь сам И. А. Мельчук — исходный пункт движения для глагола *класть*, который тоже никогда не выражается при этом глаголе, — строго говоря, не вполне константа. Надо сказать, что неточность интерпретации этого примера была допущена в работе [Рахилина 1990: 98—101], на которую дана ссылка в этом месте статьи. Для задач, которые решаются в статье И. А. Мельчука, эта неточность является незначимым нюансом — для нас, однако, она представляет некоторый интерес, поэтому мы остановимся на этом примере подробнее.

Дело в том, что предмет, который кладут, изначально находится в руке субъекта глагола *класть*. Следовательно, если бы мы подробно толковали этот глагол, фрагмент толкования, соответствующий исходной точке движения, содержал бы переменную (*рука X-a*), а значит, не был бы константой, как в *приземлиться*. Эта переменная, однако, никогда как переменная не выступает: у нее нет множества выбора значений, т. к. в каждой отдельной ситуации она связывается субъектом. Таким образом, чтобы быть совершенно точным, надо сказать, что, в отличие от *приземлиться*, содержащего в своем толковании абсолютную константу, независимую от конкретной ситуации, в *класть* имеется партиципant, фиксированный конкретной ситуацией, который может быть признан константой только внутри нее.

Нашей задачей будет показать, что поверхностная невыраженность обязательного партиципанта, даже и в тех случаях, когда он не является константой — связана с его фиксированностью относительно данной ситуации.

1. Глаголы движения: прагматика

В лингвистической литературе хорошо описан эффект дейктической (т. е. прагматической) фиксации партиципанта в глаголах дви-

жения, когда в начальной или конечной точке помещается говорящий или наблюдатель данной ситуации — впервые он был описан в работах Ч. Филлмора об английских глаголах *come* и *go* ([Fillmore 1983; 1997]; подробный анализ этого класса глаголов см. в [Падучева 2004: 373—383]). Партиципant, связанный дейктически, оказывается не выразим поверхностно, и, как показала Я. Паневова ([Paněvová 1994], см. также [Падучева 1998: 94] и [Шмелёв 1998]), к нему не может быть задан вопрос без нарушения условий успешности речевого акта. Надо сказать, что поскольку этот эффект не семантический, а прагматический, а прагматические условия очень подвижны и легко могут меняться в зависимости от контекста, строгого запрета на поверхностное выражение такого участника всё же нет: есть только тенденция, и ввиду сказанного выше, она представляется естественной.

Между тем, очень похоже на дейктические глаголы ведут себя глаголы перемещения или каузации перемещения, у которых исходная точка задана в презумпции и прагматически фиксирована — но уже не относительно говорящего или наблюдателя, а относительно субъекта движения.

Рассмотрим глагол *выпустить* (ср.: *ввиду амнистии, заключенных этой тюрьмы выпустили*); условную схему его толкования (в которой мы сознательно опустим несущественные для нас сейчас детали) можно представить так:

Презумпция: {В течение некоторого времени Q каузировал X находиться в пространстве L};

Ассерция: Q разрешил X перестать находиться в L.

В случае единичного X может возникнуть довольно жесткая прагматическая связь между ним и местом его заключения — тогда локативный участник не нуждается в поверхностном выражении и вопрос о нем свидетельствует о нарушении условий речевого акта:

— *Ходорковского выпустили.* — *?? Откуда?*

Еще раз подчеркнем, что связанность данной локативной переменной ситуативна и не определяется семантикой предиката: например, если субъект движения множествен, то имеется возможность распределения между отдельными представителями множества субъектов по разным L (тогда вопрос *откуда выпустили?* становится вполне правомерным). Связанность L может быть нарушена и при единичном объекте — если имеется прагматическая презумпция по-

следовательности постоянных L для данного субъекта — другими словами, выбора L, — и в этом случае L тоже выразим: невыразимость обязательного участника обусловлена отсутствием выбора, т. е. фиксированностью в данной ситуации.

2. Ситуативная семантика: *выиграть* и *проиграть*

Глаголы *выиграть* и *проиграть* представляют случай более жестких запретов, которые обусловлены семантическими характеристиками ситуации.

В презумпции этих глаголов имеется ситуация игры — понятно, что она содержит игроков, а также саму игру, в которую они играют. Ассерция может добавлять к этому набору участников выигрыш (или проигрыш) — в очках, деньгах, спичках и под. Ср.:

Хлестаков выиграл у городничего в карты 100 рублей.

Как видим, в этом предложении выражены все названные нами обязательные участники, и к каждому из занимаемых ими «мест» в предложении может быть задан вопрос, потому что у Хлестакова имеется выбор: партнера, игры и в каком-то смысле — выигрыша.

Совершенно другая ситуация — в случае:

Челси выиграл <у Спартака> <2 очка>.

В этом предложении не случайно остается невыраженной игра, в которую играли соперники. Вопрос к этому участнику (**Во что / в какую игру выиграл Челси?*) тоже звучит нелепо: это как раз тот случай, когда нет выбора значений участника, потому что значение его связано семантикой субъекта: Челси играет только в футбол (хотя, как маргинальную, можно представить себе ситуацию типа: *Челси выиграл у Спартака 2 очка в домино*). Говоря языком модели «Смысл $\Leftrightarrow$ Текст», в семантику имени *Челси* встроено, что это футбольный клуб, т. е. команда, которая играет в футбол — эта семантическая информация связывает переменную игры в толковании *выиграть / проиграть* — но, повторим, не вообще, то есть у предиката в целом, а в данной ситуации. Фиксированная ситуацией информация не выражается как «лишняя» — видимо, по тем же причинам, по которым мы избегаем дискурсивных повторов типа: *он пришел и (*он) тут же ушел*.

3. Антираезультативы

Теперь рассмотрим случай, когда сама глагольная семантика требует ситуативной связанности партиципентов. Это свойственно антирезультативным глаголам (подробнее о семантике и грамматике антирезультативности см. [Плунгян 2001]), и прежде всего наиболее представительному их классу — глаголам ликвидации результата (ликвидативам, по Апресян 1974: 46 и 290—292).

В качестве примера рассмотрим глагол *разоружать(ся)*. В презумптивную часть его толкования входит предикат *быть вооруженным*, который, среди прочих, имеет обязательный партиципент ‘вид оружия’:

Иракская армия вооружена автоматами, минометами и ракетами.

Следуя Принципу наследования обязательных участников, о котором мы говорили во введении, этот участник должен воспроизводиться в ликвидативном *разоружить(ся)* — однако оказывается, что он не может быть поверхностно выражен при данном глаголе, ср.:

*Буш разоружил иракскую армию (*на / *от ракет).*

Для того, чтобы выразить этот смысл, нужно выбрать другой глагол, ср.:

Буш снял с вооружения иракской армии ракеты

или:

Буш сократил вооружение иракской армии на 3 миномета.

Нам кажется, что данный запрет имеет в целом тот же механизм, что и запреты, рассмотренные выше: ситуация фиксирует род вооружений, от которого требуется освободиться, разоружаясь (это те самые вооружения, которые были «обозначены» в презумпции глагола), причем фиксирует его в полном объеме — и выбор каких-то других вооружений невозможен. Другое дело, что природа запрета здесь, в отличие от ранее рассмотренных случаев — не прагматическая, а сугубо семантическая, так что во всех употреблениях глаголов определенной (ликвидативной) семантической структуры этот эффект будет сохраняться, ср. *разгрузить, разуть, раздеть* и др. под.

З а м е ч а н и е. Следует отметить, что в семантическом отношении рассмотренные глаголы ликвидации результата связаны с так называемой

группой «полного охвата» [Апресян 1974: 279—280; Падучева, Розина 1993; Падучева 2004: 64—65], ср. *замазать стену краской, засеять поле клевером, залить каток водой* и под.; такие глаголы называют еще глаголами с ингерентной темой — под темой в этом случае имеется в виду средство обработки объекта (*краска, клевер, вода* и др.). Правда, не всякий глагол полного охвата имеет соответствующий ликвидатив, а только очень немногие, но эти ликвидативы, как кажется, полностью оправдывают и термин «полный охват» (если уж *разоружил* — то полностью, если *разгрузил* — то целиком), и термин «ингерентная тема», потому что тема (= средство) настолько встроена в семантику ситуации, что не нуждается в поверхностном выражении.

К рассмотренному (продуктивному со словообразовательной точки зрения) классу глаголов с приставкой *раз-* примыкают глаголы ликвидации результата другой поверхностной структуры, ср. *выходить*.

Глагол *выходить* имеет презумпцией острую болезнь X-а (Z), от которой он, скорее всего, должен был бы умереть — если бы не Y, который, ухаживая за X-ом, избавил его от этой болезни Z. Заметим, что И. А. Мельчук, подробно рассматривая типологию конструкций с глаголами умирания (см. с. 31 цитированной выше статьи), приходит к выводу, что для русского *умирать* (в отличие, может быть, от глаголов других языков) причина является обязательным участником (см. сноску 17 на с. 62). Таким образом, болезнь Z как партиципant входит, по правилам наследования, в число обязательных участников глагола *выходить* — но поверхностно при нем не выражается, ср.: *Няня выходила его <от тифа>*, т. к. в ситуации, заданной данным глаголом, нет альтернативы для этого партиципанта: болезнь фиксирована, если задано страдающее от нее лицо¹.

В группу антирезультативных глаголов входят не только глаголы ликвидации результата, но и глаголы неудавшегося действия: как показано в [Плунгян 2001], часто те и другие демонстрируют сходное лингвистическое поведение. Примером глаголов этого класса

¹ Любопытен в этом отношении глагол (*вы*)лечить, который по сравнению с *выходить* имеет дополнительного партиципанта — ‘средство лечения’ (ср.: *вылечил травами*). Это обстоятельство делает семантическую структуру *вылечить* принципиально другой, чем у *выходить*: каждое средство оказывается действенным по отношению к своему кругу заболеваний — поэтому вопрос: *от чего тебя вылечили травами?* прагматически звучит много лучше, чем **от чего тебя выходили?* Несколько огрубляя, можно сказать, что *выходить* — это обеспечить такой уход, при котором человек преодолевает то, от чего должен был бы умереть, тогда как *вылечить* — успешно применить лекарственное средство к некоторой (одной из) болезни. Легко видеть, что последнее сохраняет выбор значений участника Z.

можно считать близкий к *выходить* глагол *выжить*. Его анализ, в сущности, повторяет *выходить* — с точностью до каузативности (так же, как и *выходить*, этот глагол имеет в презумпции ситуацию ‘умирать от Z’), поэтому мы не будем на нем останавливаться. Зато другой глагол неудавшегося действия слишком известен, чтобы его пропустить: это глагол *промахнуться*, загадка которого (см. Мельчук 1974: 135, Апресян 1974: 147—149, Mel’čuk 2004: 41—42), как кажется, до сих пор не решена.

Согласно Mel’čuk 2004, глагол *промахнуться* имеет большой набор обязательных партиципантов, унаследованный от глагола *стрелять*, составляющего его презумпцию: ‘стрелок’, ‘мишень’, ‘оружие’, ‘средство’, ср.: *он выстрелил в меня из детского пистолета вишневыми косточками*, — однако ни один из этих партиципантов, кроме субъекта, не может быть выражен при самом глаголе *промахнуться*, ср. *он промахнулся* <*в меня / *из детского пистолета / *вишневыми косточками>. Легко видеть, что, как и в предыдущих случаях, для каждого из поверхностно невыразимых участников отсутствует ситуация выбора в случае, если задан субъект: каждый данный субъект промахивается, стреляя именно из того оружия, тем средством и по той цели, которая задана для него в презумпции и не может их изменить: действительно, нельзя стрелять в зайца, а промахнуться по отношению, скажем, к оленю.

Несколько другая ситуация с предикатом *не попасть*, действительно чрезвычайно близким по семантике к *промахнуться* (Мельчук 1974): фактически, *не попасть* повторяет ассертивную структуру *промахнуться* полностью — с точностью до связанности партиципанта ‘цель’. Дело в том, что в исходном *попасть* допускается альтернатива цели, если говорящий не отслеживает весь процесс стрельбы, а делает акцент только на ее результате: *попал не в зайца, а в оленя*; ср. в связи с этим возможное: *он не попал в зайца, зато попал в оленя*. Другими словами, в ситуации с *попасть* (но не с *промахнуться*), стрелявший мог вовсе не целиться, а выстрелить наугад, и тогда цель стрельбы *выбирается* — так сказать, волей случая. Что же касается *промахнуться*, то, по бесхитроственному толкованию В. И. Даля, этот глагол означает ‘не попасть во что метил’, т. е. не достичь именно заранее выбранной цели (ср. более «строгое», но, как оказывается, менее адекватное толкование МАС: ‘не попасть в цель’).

Глагол *промахнуться* не единичен: точно так же, например, ведет себя близкий ему морфологически антирезультативный глагол *просчитаться* (*он хотел обмануть нас, но просчитался*), который, правда, менее интересен синтаксически, чем *промахнуться*, по-

сколько содержит в презумпции всего одну переменную, не считая субъектной. Эта переменная описывает отрицательно оцениваемые говорящим планы субъекта (ср. *обмануть* в приведенном выше примере), которые в результате неправильного расчета оказываются нереализованными. Поскольку субъект просчитался в реализации всех тех планов, которые задумал (полный охват возможностей), у него нет выбора в отношении этой переменной (переменной «планов») в рамках данной ситуации — следовательно, она не может быть выражена поверхностно, ср. **просчитался в обмане*.

Для сравнения здесь можно взять глагол *ошибиться*, при котором всегда можно выразить объект ошибки (*ошибиться адресом, дверью, в расчетах* и т. п. — см. подробнее Апресян 2004, Кустова 2004: 236 и сл.). Это возможно потому, что ошибка не охватывает деятельность субъекта целиком и тем самым не фиксируется субъектом, а выбирает какой-то частный аспект его деятельности, ср. *решил задачу, но ошибся в подсчетах; принес подарок, но ошибся дверью (адресом)* и под.²

4. Заключение

Мы обсуждали механизм, который обуславливает причины поверхностной невыраженности обязательных партиципантов лексемы. С точки зрения теории «Смысл $\Leftrightarrow$ Текст», описание этого механизма важно потому, что иначе неясно, отчего принцип наследования распространяется на обязательных партиципантов, но не на слоты семантических актантов (ср. Mel'čuk 2004: 41): «партиципанты действительно наследуются, но это, однако, не обязательно означает, что наследуются Sem-A-слоты». Ответ, который мы предлагаем, связан с презумптивной связанностью одних переменных в толковании другими — т. е. с более или менее жестким взаимодействием партиципантов друг с другом, которое фиксирует значения переменных, превращая их, если так можно сказать, в «ситуативные константы».

На сегодняшний день структура толкований и моделей управления, принятая в МСТ, не позволяет такое взаимодействие отразить, потому что представляет переменные в толковании как независимые друг от друга объекты. Внимание к взаимодействию переменных

² «В случаях типа *ошибся дверью* ошибка состоит в неправильном выборе объекта (из ряда однородных), т. е. действие правильное, а объект неправильный» [Кустова 2004: 236].

сильно в школе формальной семантики, но, как нам кажется, в рамках этого направления тоже пока не обсуждались проблемы такого рода — по-видимому, из-за недооценки лексической семантики в целом как предмета исследования. Насколько и в какой форме описание лексической структуры предиката должно отражать такие взаимодействия — вопрос пока открытый.

Литература

- Апресян Ю. Д. 1974. Лексическая семантика: Синонимические средства языка // Ю. Д. Апресян. Избранные труды. Т. 1. М.: Языки русской культуры, 1995 (2-е изд.).
- Апресян Ю. Д. 2004. Синонимический ряд *ошибаться* 1 // Новый объяснительный словарь синонимов русского языка / Под общ. рук. Ю. Д. Апресяна. М.; Вена: Языки славянской культуры, Венский славистический альманах, 2004. 2-е изд. С. 752—757.
- Кустова Г. И. 2004. Типы производных значений и механизмы языкового расширения. М.: Языки славянской культуры, 2004.
- Мельчук И. А. 1974. Опыт теории лингвистических моделей «Смысл ⇔ Текст». М.: Наука, 1974.
- Падучева Е. В., Розина Р. И. 1993. Семантический класс глаголов полного охвата: толкование и лексико-синтаксические свойства // Вопр. языкознания. 1993. № 6. С. 5—16.
- Падучева Е. В. 1998. Коммуникативное выделение на уровне синтаксиса и семантики // Семиотика и информатика. Вып. 36. С. 82—107.
- Падучева Е. В. 2004. Динамические модели в семантике лексики. М.: Языки славянской культуры, 2004.
- Плунгян В. А. 2001. Антирезультатив: до и после результата // В. А. Плунгян (ред.). Исследования по теории грамматики. Вып. 1: Глагольные категории. М.: Русские словари, 2001. С. 50—88.
- Рахилина Е. В. 1990. Семантика или синтаксис? (К анализу частных вопросов в русском языке). München: Sagner, 1990.
- Шмелев А. Д. 1998. Типы «невывраженных валентностей» // Семиотика и информатика. Вып. 36. 1998. С. 167—176.
- Fillmore Ch. J. 1983. How to know wheather you're coming or going // G. Rauh (ed.). Essays on deixis. Tübingen: Narr, 1983. P. 219—227.
- Fillmore Ch. J. 1997. Lectures on Deixis. Stanford: CSLI, 1997.
- Mel'čuk I. A. 2004. Actants in semantics and syntax, I: actants in semantics // Linguistics. 42.1. 2004. P. 1—66.
- Panevová J. 1994. Valency frames and the meaning of the sentence // Ph. A. Luelsdorff (ed.). The Prague school of structural and functional linguistics. Amsterdam: Benjamins, 1994.

Process-and-Paradigm Morphology: A Separationist Sign-Based Framework

Amanda Pounder

University of Calgary
apounder@ucalgary.ca

Process-and-Paradigm Morphology is a derivative framework based on Mel'čuk's Meaning-Text Morphology as presented in e.g. Mel'čuk 1982. It maintains the structure of the linguistic sign of Mel'čuk's work, but proposes a more complex representation of the morphological operation. A process-and separationist approach to affixation allows for coherent treatment of morphological means. The morphological paradigm is elaborated as a crucial notion in morphological systems.

1. Introduction

“Process-and-Paradigm Morphology” is a framework based directly on I. A. Mel'čuk's morphological framework as it appears in Mel'čuk 1982, 2000 and elsewhere. It shares with that framework the foundational element of the linguistic sign and the morphological operation as an extension of the sign. It departs from it, however, in proposing a more complex structure for the morphological operation, and more importantly, by elaborating the notion of the morphological paradigm, a construct which is assumed to be valid for both inflectional and word-formational morphology. Process-and-Paradigm Morphology has furthermore been applied successfully in diachronic research, which confirms the general validity of frameworks of this type. The paper outlines some basic elements of Process-and-Paradigm Morphology, focusing first on the structure of the sign and expression of morphological rules and operations; differences to Mel'čuk's original proposals will be noted where they occur. It will then introduce the morphological paradigm in both its inflectional and derivational manifestations, and finally show how the paradigm can be extended to encompass complex predicates such as auxiliary constructions.

2. Fundamental Principles and Assumptions of Process and Paradigm Morphology

2.1. *The Linguistic Sign*

2.1.1. Sign Structure

Process-and-Paradigm Morphology assumes the initial premises of Meaning-Text Morphology, namely that form-meaning mappings are to be understood and represented as semiotic relationships, and that the linguistic sign is a tripartite one, consisting of *signifiant*, *signifié*, and syntactics. As a point of detail, Process-and-Paradigm Morphology has maintained Mel'čuk's original representation of the sign (cf. e.g. Mel'čuk 1982: 41) as the ordered triple (1)

$$(1) \langle X; 'X'; \Sigma_X \rangle$$

namely, with the form side first. We have not been able to identify the point at which Mel'čuk reversed the order of *signifié* and *signifiant*, but assume his motivation for the change was achieving consistency with the primacy of meaning in the Meaning-Text model. As the *signifié* and the *signifiant* imply each other, it would seem that the ordering is not particularly material. It must be said, however, that from a semiotic perspective, there is some justification for departing from the form in considering a linguistic sign: we do not have signs — or language — if we do not have forms. That is, if there is no means, including “zero”, of expressing a certain meaning in a given language (at word or subword level), then there will be no sign, not even a defective one. However, where we have morphological form but no identifiable content, we can still identify form units and defective signs. This applies both at the lexical level (we all have items in our vocabulary which are familiar but for which we have not learned the meaning) and at the subword level, where we have both defectivity and forms which do not correspond to identifiable morphological or lexical meanings. Where we can postulate formal “zeros” based on the paradigm in morphology, there is no corresponding semantic “zero” in a paradigm. We therefore attribute a certain primacy to form in this framework.

2.1.2. Morphs and Morphemes

Process-and-Paradigm Morphology is a separationist framework in the sense of Beard (e.g. 1995), Aronoff (e.g. 1994), and Stump (e.g. 2001). That is, there is considered to be a distinction between the form-meaning

relationships found in lexical units, on the one hand, and in morphological units, on the other hand. Lexemes and roots/radicals “have” meaning in the Saussurean sense and the form-meaning relationship is direct. However, affixes do not “have” meaning in this way; there is, rather, an indirect mapping relationship between morphological form on the one hand and morphological meanings on the other (for justification, see the above-mentioned works). Although Mel’čuk’s Meaning-Text framework is separationist in that it treats form and meaning separately, this difference between lexical and morphological units is not part of that framework.

As a consequence of the separationist approach taken here, (non-lexical) morphs and morphemes are primarily units of form. We therefore cannot speak of “*the* plural morpheme” of a language L (unless we are speaking loosely and there is only one abstract form in that language that is associated with the meaning ‘pl’), as one might in many approaches to morphology, including Mel’čuk’s. This means that, say, the *-s* and the *-en* of English must belong to two distinct morphemes, as there is no phonological regularity in English that will allow one to be formally derived from the other; it likewise means that if an affixal shape is associated with different meanings, but has the same phonological behaviour in all phonological contexts, then we have to do with just one affix. Thus there is just one affix *-en* in German, which may be combined with an impressive number of morphological meanings. The advantages of the separationist approach include the fact that we can consider e.g. the [z], [əs], and [s] of English as allomorphs of one morpheme and thus as more closely related to each other than /i/, /ən/ etc., which must under the traditional and Mel’čukian conception be considered as equally related to each other. The “underlying morph” of Mel’čuk (1982, 2000) is no longer an embarrassment, but corresponds to this affix morpheme¹.

2.1.3. Sets and Signs

Although many things are attributed sign status, such as morphological operations, and we are told (e.g. Mel’čuk 1975: 298) that the sign concept may be extended abstractly, Mel’čuk states (e.g. 1982: 41) that lexemes and morphemes (and any other -eme units, of course) cannot be signs, because they are sets. It is not obvious to us why this should be the case; at some abstract level, a set is a thing, and a thing may be a sign. It

¹ The underlying morph can only be postulated if all “allomorphs” are related via a phonological regularity; thus one can be postulated for *wife/wive*, but not for [s]/[z]/[əz]. See also 2.1.3. for a more important objection.

is abstraction, indeed, that makes the exclusion of the *-eme* unit artificial and embarrassing. Whatever level we pick, our supposed object will be a set: we may not congratulate ourselves on isolating a morph affix *a*, for it will still represent a set of realizations in speech, and we may even find ourselves abstracting over a number of phonetic variations in isolating it. The notion of “underlying morph” is likewise a problem; it, too, can be considered to represent a set of forms. It seems odd not to be able to say that a word-form, for example, may be a sign and thus have *signifié*, *signifiant*, and syntactics, but the abstract item, the lexeme, that it represents does not also have *signifié*, *signifiant*, and syntactics, which would in each case be an abstraction over the *signifiés*, *signifiants*, and syntactics of the set of word-forms it contains. It is surely desirable to be able to say that derivation operates on form, meaning, and syntactics of a lexeme; therefore, the lexeme needs to have these things. We also wish to avoid the problem of conceptualizing word-formation, i.e. how lexemes may be formed, if operands exist uniquely at the concrete level. Process-and-Paradigm Morphology, therefore, considers the difference between *-eme* and *allo-* to be one of abstraction primarily, while not denying the set conception of *-eme*. Morphemes and lexemes, in this framework, are abstract signs and have the same basic structure as their concrete manifestations.

2.2. Morphological Rules and Operations

2.2.1. Forms and Change of Form

In a process framework, morphology is something that happens: a morphological operation that a base undergoes consists in a change to its form, to its meaning, and/or to its syntactics. A form may undergo a change by having something added to it (affixation), by having some part of it modified (apophony), by a process in which the product is identical to the original (identity process), or by adding it to something else (composition). In all these cases, the signalling function consists in the process, or, put another way, in the relation established between the original form and the result. Therefore, the form-meaning relationships holding in the case of affixation on the one hand and in the case of apophonies and conversions on the other hand are essentially the same: we will not say, as in traditional approaches, that either a given affix, i.e. an object, “means” plural or a given event, or relation between two forms, as effected e.g. by a substitution of the base vowel, “means” plural. This approach to affixation sets Process-and-Paradigm Morphology apart from Mel’čuk’s framework.

2.2.2. Morphological Rules

The changes to *signifié*, *signifiant*, and syntactics of a base to as outlined in 2.2.1. are effected by means of rules, which are essentially independent from each other; the operation maps the rules onto each other. Given the English forms *dogs*, *mice*, and *deer*, we will then have three different form rules, one effecting an affixation, one a vowel apophony, and one an identity process respectively, to be applied to the base representing the lexeme form, each mapping onto the same semantic rule. The semantic rule, parallel to the form-rule, expresses a relationship between lexeme or base meaning and meaning of the result, and this is represented in terms of a function, e.g. $PL('X')$, where PL stands for 'plural', 'X' is the meaning of the base (this is an abbreviation of $'X' \rightarrow PL('X')$). An example in the word-formation domain would be $('X' \rightarrow) EX('X')$, which expresses an origin relation between the base meaning and the product. Syntactic rules can do such things as change the syntactico-lexical class of a lexeme in word-formation, as in e.g. $\Sigma_N \rightarrow \Sigma_V$ (meaning, here, the combinatory properties of a noun become those of a verb), or effect an agreement relation between two sentence elements, as in $AGR_{SUBJ}('1p', 'sg')$ (an abbreviation of $\Sigma_V \rightarrow AGR_{SUBJ}('1p', 'sg')$, meaning, here, an agreement relation holds with the clause subject, corresponding values '1st person' and 'singular'). In the spirit of Meaning-Text Morphology, the rule is considered to be a sign, so what we have seen so far actually corresponds to the *signifiants* of these abstract signs. Their *signifiés* are abstract, and include status of the rule in the system. Their syntactics include ("stem conditions") information as to the class of base they apply to and ("rule conditions") information as to the mechanics of applying the rule (which in the case of a form rule might involve requiring phonological adjustments in the base, for example).

2.2.3. Morphological Operations

The morphological operation, following the established pattern, is a sign, whose *signifiant* consists of the mapping between a form rule and a meaning rule and/or a syntactic rule. The *signifié* of the operation is again abstract, and includes information concerning the status of the operation in the system, such as its relative productivity. The syntactics of the operation contains information necessary for the carrying-out of the operation: the stem conditions indicate what class of base the operation applies to (and will thus represent a subset of the intersection of the stem conditions in the syntactics of the participating rules) and the operation condi-

tions indicate how the operation is to be carried out, e.g. in what slot position relative to the base, whether it entails the application of any other operation, etc. An example of a representation of a word-formation operation in German is shown in (2).

(2) Word-Formation Operation #5

$$\left\langle \left[\begin{array}{l} \langle X \oplus en; 'FR_{11}'; s.c. N, V, Adj... \rangle \\ \langle EX('X'); 'SR_9'; s.c. N... \rangle \\ \langle \Sigma_N \rightarrow \Sigma_{Adj}; 'SR_4'; s.c. N... \rangle \end{array} \right] ; \begin{array}{l} 'WOp_5'; s.c. N; sem.class Substances, \\ Metals/Minerals; o.c.: Slot I \end{array} \right\rangle$$

The operation is given the arbitrary index #5; the point is that it must be accessible to the system. The *signifiant* of the operation is the portion in square brackets and contains the three rules mapped onto each other. The first of these is the form rule (here arbitrarily named Form Rule #11, FR_{11}); its *signifiant* expresses that the affix *en* is added to qualifying bases, while its syntactics include as stem conditions (s.c.) a statement that a qualifying base must be a noun, verb, or adjective. The *signifiant* of the semantic rule (arbitrarily named SR_9) below it contains the origin function $EX('X')$ described in 2.2.2., and its syntactics states that it applies to noun bases — it thus has a narrower domain than the form rule. SR_4 , the syntactic rule, substitutes adjective properties for noun ones, and applies to noun bases. Among other things, the semantics of the operation will state that it is of low relative productivity. The syntactics of the operation states (s.c.) that the base must be a noun of the lexico-semantic classes Substance Designations or Metal/Mineral Designations. The syntactics of the operation also contains an operation condition (o.c.) stating that it applies in Slot I; this is what identifies it as a suffixation operation. Given a lexeme such as PERLE ‘pearl’, base *perl-*, we can apply this operation to it and get the adjective PERLEN ‘(made) of pearl’, or given the adjective PERLEN, we can interpret it by means of our knowledge of the operation.

3. The Morphological Paradigm

Whereas the morphological paradigm plays no role to speak of in Meaning-Text Morphology, it is of central importance in the framework described here. As is clear from historical and psycholinguistic studies, the paradigm is the principal organizational construct in the morphological component, that is, for both inflectional and word-formation morphology. It is dynamically defined here as the set of paths which a base may

(or must) follow in entering the system to create a grammatical word-form or lexeme, nodes in the path being operations. Thus, for example, a verb base for which no path has been lexically prespecified will “proceed” from slot to slot, and will select one of a set of candidate operations available at that slot for which it qualifies, and then “proceed” to the next slot, until it reaches the final slot and a correct form is put out (Figure 1). In this schematic diagram for the verb inflection paradigm of a hypothetical language L, we see that there is one prefix position, where there is a choice between two operations, one non-affixational operation that applies to the stem (in L, it is a vowel apophony), and there are two suffix positions. In L, there are three nodes in Slot I; if one selects “Operation #5” or “Operation #6”, one is obliged to select Operation 8 in Slot II.

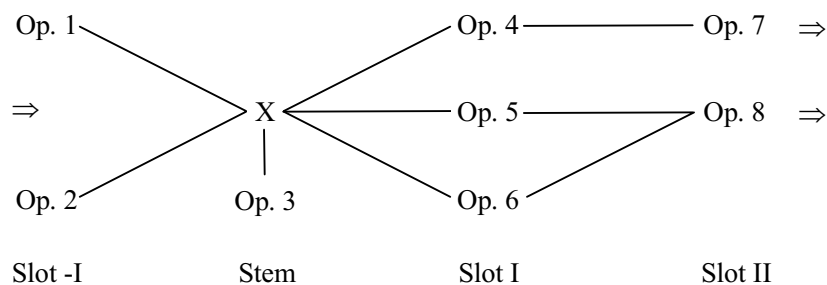


Fig. 1. Schematic Structure of the Verb Inflection Paradigm in Language L

The paradigm is essentially derivative of the set of operations in the system, in that the syntactics of the operations determine the sequence of operations which a given base entering the paradigm must undergo, as well as the properties a base must have in order to qualify to undergo them. As well as ensuring the creation of complete word-forms and lexemes, the paradigm defines the set of morphological synonymies (= the set of operations with a common semantic rule available for bases of a given class) and the set of morphological polysemies (= the set of operations with a common form rule available for bases of a given class) in any one slot position, as well as, derivable from this, morphological synonymy and polysemy relations for complete forms. It is these polysemy and synonymy relations within the paradigm on the one hand, and contrasts on the other, which are considered essential for mental structuring of morphological systems and which serve as the basis for diachronic change. The set of paths taken for an individual lexeme constitute its “lexical

paradigm”, and complete or partial replicas of this may be recorded in individual lexical entries, in addition to any lexical idiosyncrasies not produced by morphological operations.

The derivational paradigm is essentially similar to that of the inflectional paradigm, but depending on the language, the latter may be more complex, with e.g. multiple prefix and suffix slots. A base entering the derivational paradigm will likely only have one or two nodes to pass through before a complete form is created. This complex form may itself, of course, enter the paradigm to undergo further derivation. Thus a lexeme such as German *Unsterblichkeit* ‘immortality’ represents three passes through the word-formation paradigm: in each case, selection of an operation was made in either (prefix) Slot -I or (suffix) Slot I. The first pass adds the affix *lich* in Slot I to the verb base *sterb-* ‘die’, producing the adjective lexeme STERBLICH ‘mortal’; the second pass adds an affix *un* in Slot —I, creating the adjective lexeme UNSTERBLICH ‘immortal’; the final pass adds an affix *keit* in Slot I to the lexeme base *unsterblich-*, producing the noun UNSTERBLICHKEIT ‘immortality’.

4. Complex Predicates

Recently there have been numerous pleas for formally incorporating complex predicates, such as auxiliary verb constructions, in paradigms (cf. e.g. Ackerman and Stump 2002, Webelhuth and Ackerman 1998) and thus emphasizing the morphological, as opposed to the syntactic, nature of these constructions. Meaning-Text morphology goes in this direction in the sense that the lex, as a concrete manifestation of a lexeme, includes complexes such as *have eaten*, *will bite* as well as word-forms such as *ate*, *bit*, which would then all be members of the paradigms of the respective lexemes. One problem that these approaches manifest is that they neglect the forms of the participating auxiliary, which when they are finite are actually the ones being modified in the morphology; in the typical case, the word-forms of the verb are identical in its auxiliary and non-auxiliary functions, so it is undesirable to consider them to be distinct verbs. Process-and-Paradigm Morphology attempts to reconcile the desire to include complexes in the paradigm and the need to recognize that two (or more) lexemes are participating by introducing the notion of “quasi-operation”, which involves a mapping of the desired semantic rule, which will be contained in the list of semantic rules available to the inflectional morphology, onto a linking between the (path leading to the formation of the) required non-finite form of the lexeme and the lexical entry of the

auxiliary (Pounder 2004a). This lexical entry will contain information as to what subset of the lexical paradigm will be available in conjunction with the quasi-operation and what syntactic configurations are associated with the resulting forms (so, for example, while in its non-auxiliary functions the verb may subcategorize a NP complement, in association with a quasi-operation it may subcategorize a VAdjP (verb phrase headed by a participle)). In other words, the ‘perfect’ function of a verb in German will correspond to or be expressed by the linking of a non-finite (participle) form with an auxiliary. This solution has the advantage of dismissing the question of whether such meanings as ‘perfect’ are to be located in the non-finite form of the lexeme or in the auxiliary (cf. e.g. Bierwisch 1990, Klein 2000). An example of a quasi-operation for German is given in (3) and an excerpt from the lexical entry of the auxiliary in (4).

$$(3) \text{ QO}_1 = \left\langle \left[\begin{array}{l} \langle \text{PARLINK}(\text{VAdj}, \text{HABEN}); 'LR_1'; \text{V...} \rangle \\ \langle \text{PERF}(\text{X}); 'SR_{10}'; \text{V...} \rangle \end{array} \right]; 'QO_1'; \text{s.c. V...} \right\rangle$$

$$(4) \text{ HABEN... } \Sigma: \quad \begin{array}{l} \text{V}_{\text{wk}} \\ [\text{NP} \quad \_\_\_\_\_\_] \\ \text{QO}_1: [\text{VAdjP} \quad \_\_\_\_\_\_] \end{array}$$

In (3), we see a representation of “Quasi-Operation #1”. Its *signifiant* maps the Linking Rule #1 onto Semantic Rule #10 featuring the ‘perfect’ function. The *signifiant* of LR_1 contains the the paradigm linking function, which causes a participle to be linked with the lexical entry for the verb HABEN ‘have’. The syntactics of the linking rule states that it applies to verbs. The lexical entry of HABEN (4) will contain, among other things omitted here, its syntactics (Σ), e.g. that it belongs to the class of weak verbs (which itself can be interpreted as shorthand for selecting the default paths in the inflectional paradigm), that it subcategorizes NPs, and finally, that if QO_1 has been invoked during the attempt to create a form of another verb, in other words, if HABEN has been invoked as an auxiliary, it subcategorizes a VAdj Phrase.

6. Conclusion

Process-and-Paradigm Morphology has a great deal in common with its parent Meaning-Text Morphology; the commonalities outnumber the differences, which of course were highlighted here. It has been used to

model subparts of inflectional and word-formation systems (e.g. Pounder 1996a,b, 2000, 2004a) as well as to describe diachronic developments of those systems (e.g. Pounder 2000, 2001, 2004b). In word-formation studies, a focus has been on interaction between lexical information of various kinds and morphology, especially at the semantic level. Current research projects examine interfaces between syntax, prosody, and paradigmatic selection of operations (e.g. Pounder 2005).

References

- Ackerman, F. and G. Stump 2002. "Paradigms and Syntactic Expression. A Study in Realization-Based Lexicalism", in A. Spencer and L. Sadler (eds), *Constraint-Based Morphosyntax*. Stanford: CSLI, 2002.
- Aronoff, M. 1994. *Morphology By Itself*. Cambridge: MIT Press, 1994.
- Beard, R. 1995. *Lexeme-Morpheme Based Morphology*. Albany: SUNY Press, 1995.
- Bierwisch, M. 1990. "Verb Cluster Formation as a Morphological Process". *Yearbook of Morphology* 3. 1990. P. 173—199.
- Klein, W. 2000. "An Analysis of the German perfekt". *Language* 76. 2000. P. 358—382.
- Mel'čuk, I. A. 1975. "Die Konversion als morphologisches Mittel", in I. A. Mel'čuk, *Das Wort*. Munich: Fink, 1975. S. 288—318.
- Mel'čuk, I. A. 1982. *Towards a Language of Linguistics*. Munich: Fink, 1982.
- Mel'čuk, I. A. 2000. *Cours de morphologie générale*. Vol. V. Montréal: PUM, 2000.
- Pounder, A. 1996a. "Inflection and the Paradigm in German Nouns", *American Journal of Germanic Linguistics* 8. 1996. P. 219—263.
- Pounder, A. 1996b. "The Semantics of *be*-Verbs in German", *Proceedings of the 1996 Annual Conference of the Canadian Linguistic Association*. Calgary: Calgary Working Papers in Linguistics, 1996.
- Pounder, A. 2000. *Processes and Paradigms in Word-Formation Morphology: The Development of the Denominal Adjective System in German*. Berlin: Mouton, 2000.
- Pounder, A. 2001. "Adverb-Marking in German and English: System and Standardization". *Diachronica* 18. 2001. P. 301—358.
- Pounder, A. 2004a. "The Role of the Paradigm in German Auxiliary Constructions", *Proceedings of the 2003 Annual Conference of the Canadian Linguistic Association*. 2004. P. 190—201.
- Pounder, A. 2004b. "Haplology in English Adverb-Formation", in C. Kay, C. Hough, I. Wotherspoon (eds.), *New Perspectives on English Historical Linguistics. Vol. 1: Syntax and Morphology*. Amsterdam: Benjamins, 2004. P. 193—211.

-
- Pounder, A. 2005. "Syntactic Haplology, Morphological Ellipsis, and Phonological Aesthetics", paper presented at Workshop on Interfaces With Morphology, Schloss Freudental, Feb 9—13. 2005.
- Stump, G. 2001. *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge: CUP, 2001.
- Webelhuth, G. and F. Ackerman. 1998. *A Theory of Predicates*. Stanford: CSLI, 1998.

Метафора в модели «Смысл ⇔ Текст»

Tilman Reuther

Universität Klagenfurt, Austria
tilmann.reuther@uni-klu.ac.at

Metaphor in the Meaning ⇔ Text Model. First, the paper gives an overview of ways how metaphor is dealt with in MTT. Second, it presents some related approaches to metaphor in Russian linguistics. Third, it states lexicographical problems and presents solutions to them (semantic derivation, metaphor and lexical functions, writer's metaphor). Fourth, perspectives for further work on metaphor in MTT are listed.

1. Метафора в модели «Смысл ⇔ Текст» (МСТ) — общее положение дел

1.1. Общей теории метафоры в МСТ нет. Метафора трактуется как явление лексического уровня и учитывается при разработке словарных статей ТКС/ДЕС. Рассмотрим два известных случая: 1) Толкование лексем с эксплицитным компонентом 'как если бы' / 'comme si', отсылающим к зоне коннотаций другой лексемы внутри словарной статьи, и 2) Лексемы-значения лексических функций, представляющие собой явные случаи метафорических значений слов.

1.1.1 Толкование лексем с эксплицитным компонентом 'как если бы' / 'comme si' (ср. Mel'čuk/Clas/Polguère 1995, с. 195—199):

(1) VENDRE^{I.1} *Pierre vend son automobile à Jean pour trois mille dollars.*

Connotations: 1) primauté du profit sur les considérations morales
[cf. VENDRE^{II}]

2) capacité de convaincre [VENDRE^{III}]

VENDRE^{II} *Cet homme a vendu ses camarades à l'ennemi.*

'X trahit Y auprès de Z 'en échange' d'un bénéfice W [comme si X vendait^{I.1} Y à Z pour W]'

VENDRE^{III} *Cet homme a su vendre son idée au ministre.*

‘X convainc Z d’accepter^{I.1b} la proposition² Y¹ de Y² de X
[comme si X vendait^{I.1} Y à Z]’

Толкования лексем VENDRE^{II} и VENDRE^{III} и зона коннотаций при лексеме VENDRE^{I.1} демонстрируют формат полисемии типа «метафоры через коннотацию» (ср. Iordanskaja/Mel’čuk 1980). Ср. также определение коннотации в Apresjan 2004, XXXII:

(2) **Коннотации, или ассоциации** — те образные представления, которые связываются в сознании носителей языка с объектом, обозначенным данной лексемой, но не входят непосредственно в ее значение. Ср. коннотации таких названий животных, как *лиса* (хитрость), *свинья* (нечистоплотность, наглость и примитивное поведение), *овца* (покорность) и т. п. Коннотации лексикализованы, т. е. ассоциируются с предметом действительности не прямо, а через его языковое обозначение. Поэтому даже у близких синонимов они могут быть различными. Ср. коннотацию глупости и упрямства у лексемы *осел* *I* и коннотацию трудолюбия и готовности безропотно работать у лексемы *ишак* *I*; коннотацию преданности у лексемы *слуга* *I* и коннотацию сервилизма у лексемы *лакей* *I*. Коннотации, присущие слову в его прямом значении, могут становиться частью его смысла в переносных значениях, производных словах, фраземах и т. п. Ср., например, переносные значения слов *осел* ‘упрямый дурак’, *ишак* ‘трудолюб’, *слуга* ‘преданный человек’, ср. *слуга царю*, *лакей* ‘рабоподчиненный, выслуживающийся человек’ и т. д.

1.1.2 Многие лексемы-значения стандартных лексических функций (ЛФ) представляют собой явно метафорические значения слов, ср., напр., слова КОЛЬЦО, ВЕСТИ, ПИТАТЬ, РАЗГОРЕТЬСЯ в следующих примерах (о толковании глаголов см. ниже, п. 3.2.):

- (3) (а) эксплицитно: *Figur* (блокада): *кольцо блокады*
(б) имплицитно: *Oper*₁ (переговоры): *вести*; *Oper*₁ (доверие): *питать*
*IncepFunc*₀ + *IncepPredPlus* (спор): *разгореться*

1.2. Метафора эксплицитно представлена (всего лишь) в одном правиле системы перифразирования МСТ (ср. DEC III/1992, 40 с ЛФ *Figur*):

- (4) 32. $C_0 \Leftrightarrow \text{Figur}(C_0) \xrightarrow{\text{ATTR}} C_0$
Le brouillard (C_0) *nous cache le sommet* ⇔ *Le rideau* *Figur* (C_0) *de brouillard nous cache le sommet.*

1.3. Имеется определение метафорического значения и метафорического употребления слова в Apresjan 2004, XXXIII—XXXIV:

(5) **Метафора** — образное обозначение какого-то явления (предмета, действия, процесса и т. п.) А именем явления В, которое в каких-то отношениях похоже на А или воспринимается говорящим как похожее на А; ср. *черный вторник, закат жизни*. Если слово постоянно используется в качестве метафоры, то у него складывается метафорическое значение; ср. *громкий процесс, трескучие морозы, Царит тишина, Камни говорят*, где прилагательные и глаголы имеют метафорические значения; ср. также метафорические значения 2—4 существительного *молния* (см. под термином **лексема**). В противном случае говорят об авторской метафоре, или о метафорическом употреблении слова; ср. *пряжа тумана, лысины булыжников; И пели басом путников шаги* (Б. Пастернак).

(6) **Лексема** — слово в одном из его значений. В слове *молния*, например, выделяются лексемы *молния 1* = ‘разряд атмосферного электричества’ (с коннотацией быстроты), *молния 2* = ‘особо срочная телеграмма’, *молния 3* = ‘срочно выпускаемая стенная газета, посвященная особо важному вопросу’ и *молния 4* = ‘быстро задерживающаяся скользящая застежка’. (...)

Метафоре противопоставлена метонимия, ср. определение в Apresjan 2004, XXXIV:

(7) **Метонимия** — образное обозначение какого-то объекта А именем другого объекта В, который располагается на или внутри А, является частью А или как-то иначе соприкасается с А, ср. существительное *маска* как обозначение человека с маской на лице, имена столиц (*Москва, Париж*) как названия правительств соответствующих стран, названия помещений для обозначения совокупности находящихся в них людей (*Зал бурно аплодировал любимому артисту*) и т. п.

1.4. В Apresjan 1974a, 193—215 (ср. также Apresjan 1974b) перечислены типы регулярной многозначности — среди них подавляющее большинство составляют, очевидно, метонимические связи между значениями (кроме указанных выше, ср. также ‘Растение’ — ‘плод этого растения’; ‘Животное’ — ‘мех животного’, ‘Часть тела’ — ‘часть одежды, которая на ней находится’ и др.). Наряду с этим, указаны также пять типов «метафорического переноса»:

(8) ‘Животное, насекомое’ — ‘человек, похожий на него’ (*баран, комар*); ‘Больше нормы X-а’ — ‘высокая степень’ (*глубокая река — глубокие знания*); ‘Больше <меньше> нормы X-а’ — ‘положительная <отрицательная> оценка’ (*богатый крестьянин — богатый язык, уз-*

кая улица — узкие взгляды); 'Относящийся к X-у' — 'похожий на X' (деревянная ложка — деревянное лицо); 'Относящийся к X-у' — 'похожий на X цветом' (огненный столб — огненные волосы).

1.5. МСТ моделирует языковую компетенцию «среднего носителя языка», т. е. не писателя или поэта. ТКС базируется в основном на текстах обиходной, повседневной тематики и нейтрального стиля. В лексикографической работе используется методика «языкового эксперимента», примеры из художественной литературы отбираются относительно их «лексикографической ценности». Таким образом, в ТКС выбрана средняя из трех возможных альтернатив: как метафорические обозначаются такие значения слов, которые ощущает средний носитель языка. (Две другие альтернативы — либо вообще не регистрировать метафоричность значений, либо включать без разбора авторские метафоры.)

2. Экскурс: Другие подходы к метафоре в российской лингвистике

2.1. Большие словари русского языка базируются на выписках из художественной литературы, в связи с чем содержат большое количество авторских метафор. Ср., напр., следующую выписку из МАС — авторские метафоры зачеркиваем и тем самым указываем на то, что, по нашему мнению, выходит за рамки лексикализованной метафоры.

(9) РАЗГОРЕТЬСЯ (МАС 3/1984, 599)

1. Начать сильно или хорошо, ровно гореть. *Уже разгорелась зажженная лампа, и они сели за стол.* Федин. *Через несколько минут костер разгорелся как следует.* Арсеньев. ~~// Начать светиться всё ярче и ярче. Звезды разгорелись. // Стать ясным, солнечным. День, наполненный промозлой сыростью, так и не разгорелся.~~ Тендряков. *Яркое разгорелось утро, солнце пригрело плечи.* Ливеровский.

2. Приобрести огненно-красный цвет. *Смотри, как запад разгорелся. Вечерним заревом лучей.* Тютчев.

3. Стать красным, горячим (от возбуждения, волнения и т. п.). *Щеки мои разгорелись, и сердце билось от какого-то томительного и мне неведомого доселе ощущения.* Достоевский. *Антониды Иванов на разгорелась на морозе румянцем во всю щеку.* Мамин-Сибиряк.

4. перен. Оказаться охваченным все усиливающимся чувством, стремлением. *Рудин сперва как будто колебался, не решаясь высказаться, не находил слов, но, наконец, разгорелся и заговорил. ...* Тургенев. *Разгорелось у Пахома сердце. Думает: — Что же тут в тес-*

~~ноте бедствовать, коли можно хорошо жить. Л. Толстой. И толпа разгорелась. Выступали охотно сами, ... Фурманов.~~

5. перен. Развиваясь, стать очень сильным, напряженным. Узнал он, что я бывал за границей, и любопытство его разгорелось. Тургенев. Снова между нами разгорелся живой, шумный спор. Горький. Ночной бой сразу разгорелся и быстро затих, так как японцы исчезли в ночной темноте. Степанов.

2.2. О метафоре в последнее время писали и московские лингвисты, непосредственно не связанные с МСТ, но образующие более широкий круг московской семантической школы, ср., напр., Padučeva 2004, Rachilina 2000, а также сборник трудов Arutjunova 1999. Рассмотрим вкратце подход в книге Padučeva 2004, 157—176. Примеры приводятся в удобной для обзора форме, т. е. по сути дела, а не строго по нормам цитат из источника.

(10) Таблица 1

Padučeva 2004, 159:	Семантическая деривация		Комментарий:
Метафора = семантический/категориальный сдвиг			
Сеть	тонула ('погружаться в воду')	в озере.	участник Среда = ЖИДКОСТЬ
Лыжи	тонули ('вязнуть')	в рыхлом снегу.	
Петров	тонул ('погибнуть', 'прекращение существования')	в озере.	участник Субъект = ЖИВОЕ СУЩЕСТВО
Окрестности	тонули ('быть невидимым')	в туманном сумраке.	
Хлопки выстрелов	тонули ('быть неслышимым')	в свисте ветра.	
Ср. также наш пример (Т. Р.):			
Петров	тонул	в долгах.	

Padučeva 2004, 158:	Семантическая деривация		
Метонимия = сдвиг фокуса внимания (т. е. актант переходит в центральную позицию)			
Ветер	шелестит	листьями.	
Листья	шелестят.		
Мороз	затянул	лужи	тонким льдом.
Тонкий лёд	затянул	лужи.	
Ср. также наши примеры (Т. Р.):			
Судьба	затянула	её жизнь	грустью и печалью.
Грусть и печаль	затянули	её жизнь.	
Это известие	затянуло	город	покровом молчания.

Различая метафору и метонимию, Padučeva 2004, 176, отмечает, однако, следующее обстоятельство:

(11) Метафора часто возникает как следствие метонимии: метонимия чревата метафорой, поскольку при переносе фокуса внимания на смежный объект на месте согласованной категории оказывается несогласованная.

(12) Таблица 2, примеры:

Padučeva 2004, 175:		метонимия либо метафора	
<i>Левой рукой</i>	<i>он</i>	<i>сжимал</i>	<i>хрустящую бумажку.</i>
	<i>Левая рука его</i>	<i>сжимала</i>	<i>хрустящую бумажку.</i>

2.3. Неоднозначную оценку среди московских лингвистов получила теория метафоры Лакоффа-Джонсона (Lakoff/Johnson 1980). Arutjunova 1999, 371, пишет: «В метафоре стали видеть ключ к пониманию основ мышления и процессов создания не только национально-специфического видения мира, но и его универсального образца». Утверждения об универсальности ключевых, базисных метафор и о первичных, врожденных концептах в связи с классическими примерами ARGUMENT IS WAR, LOVE IS A JOURNEY кажутся нам сомнительными. О критике теории Лакоффа-Джонсона см. также Padučeva 2004, 172 сл. (обзор примеров типа TIME IS MONEY и др. в английском и русском языках).

3. Проблемы и решения

3.1. Существует обширная литература о метафоре из областей риторики, литературоведения, философии языка и др., и существует собственно лингвистическая теория метафоры. Общеизвестно противопоставление «живая»/«авторская» метафора vs. «мертвая»/«лексикализованная» метафора — ср., в частности, Ricœur 1975 и др. В рамках МСТ наблюдаем следующую картину:

(13) Таблица 3

лексическая историческая стёртая метафора	лексическая актуальная ощутимая метафора	авторская окказиональная метафора	семантическая деривация
<i>журавль I</i>	<i>осел I</i>	<i>голубая тишина</i>	<i>открыть 1</i>
<i>журавль II</i>	<i>осел 2</i>	<i>= небо</i>	<i>открыть 2</i>
в МСТ — это омонимия	в МСТ — это полисе- мия через коннотацию	в МСТ пока не учитывается	в МСТ — это регу- лярная полисемия

<i>питать 1</i>	<i>Разгореться 1</i>		
<i>питать 7</i>	<i>Разгореться 5</i>		
это тривиальная полисемия (наше решение, см. ниже, п. 3.2)	это полисемия с не-тривиальным остатком (наше решение, см. ниже, п. 3.2)	(наш анализ примера, см. ниже, п. 3.3)	(см. ниже, (14)—(16))

Относительно трактовки семантической деривации и упорядочения лексем в рамках системной лексикографии и МСТ см. Apresjan 2003:

(14) Таблица 4

Примеры наши — <i>T. R.</i> :	Семантическая деривация		
<i>Петров</i>	<i>открыл</i>	<i>дверь</i>	<i>ключом.</i>
<i>Дверь</i>	<i>открывалась</i>		<i>ключом.</i>
<i>Ключ</i>	<i>открывал</i>	<i>дверь.</i>	

(15) Самые многозначные глаголы — те, которые принадлежат к классу действий. Выше уже шла речь о типичной цепочке семантической эволюции: действия $\Rightarrow$ воздействия $\Rightarrow$ процессы $\Rightarrow$ состояния $\Rightarrow$ свойства. Действия потому наиболее многозначны, что в потенциале для них возможна самая длинная цепочка семантической деривации: из них могут развиваться деятельности, занятия, процессы, происшествия, положения в пространстве, состояния, свойства и т. п. (Apresjan 2003, 21).

(16) В словарях значения сильно многозначных глаголов уже давно упорядочиваются в следующем направлении: прямые («свободные») $\Rightarrow$ производные (переносные) $\Rightarrow$ лексически или синтаксически связанные $\Rightarrow$ грамматикализованные (в том числе связочные и чисто служебные). Этот принцип отражает универсальный семантический закон постепенного семантического выветривания глаголов: (...) (Apresjan 2003, 9).

3.2. Развивая линию лексикографического описания связей между основным значением слова и его переносным, метафорическим употреблением, предлагаем вниманию следующие формальные и неформальные толкования значений приведенных выше глаголов ВЕСТИ, ПИТАТЬ, РАЗГОРЕТЬСЯ, выступающих в качестве значений ЛФ. Значения за номером 1 — это основные значения. Значения ВЕСТИ 1 и ВЕСТИ 14 истолкованы формально, т. е. в формате ТКС с помощью сем, отсылающих в эксплицитной форме от метафорического значения к основному значению; значения ПИТАТЬ 1 и ПИТАТЬ 7, а также РАЗГОРЕТЬСЯ 1 и РАЗГОРЕТЬСЯ 5 приведены из БАС и МАС и истолкованы неформально:

- (17) (а) **ВЕСТИ** (ср. Reuther 1996)
1. *X ведет Y-а из A в B по трассе C* = X контролирует движение Y-а из A в B, X и Y идут из A в B по трассе C.
 14. X Oper₁ C₀ = X контролирует C₀ попеременно с Y-ом, X делает C₀ (в определенный отрезок времени), C₀ — языковая деятельность X-а и Y-а.
- (б) **ПИТАТЬ** (БАС 9/1959, 1253—1254)
1. Давать еду, пищу; кормить.
 7. *Перен.* Питать что к кому, чему. Испытывать к кому-, чему-либо какое-либо чувство (в сочетании с существительными, обозначающими соответствующее чувство).
- (в) **РАЗГОРЕТЬСЯ** (МАС 3/1984, 599)
1. Начать сильно или хорошо, ровно гореть.
 5. *перен.* Развиваясь, стать очень сильным, напряженным.

Из этих примеров следует, с одной стороны, что толкования традиционных словарей недостаточно эксплицитны, а, с другой стороны, что имеется возможность эксплицитного толкования лексем в рамках МСТ. Формальные толкования глаголов ПИТАТЬ 7 и РАЗГОРЕТЬСЯ 5 в метафорическом значении представляем себе так:

- (18) (а) **ПИТАТЬ 7.** X Oper₁ C₀ к Y = ‘X находится в ментальном или эмоциональном отношении C₀ с Y-ом’
- (б) **РАЗГОРЕТЬСЯ 5.** IncepFunc₀ + IncepPredPlus C₀ = ‘Внезапно началось и стало интенсивным межличностное отношение C₀’

Здесь сема ‘находиться’ определяет тривиальную полисемию слова ПИТАТЬ, а семемы ‘внезапно начаться и стать интенсивным’ показывают нетривиальную полисемию слова РАЗГОРЕТЬСЯ.

3.3. Обратимся теперь к примеру авторской метафоры и возможности ее лингвистической трактовки.

- (19) (а) ...*шары впитывались в голубую тишину*...
- (б) шар $\leftarrow$ впитываться $\rightarrow$ тишина $\rightarrow$ голубая

Авторская метафора реализуется часто в рамках нескольких предложений предыдущего контекста (между тем, МСТ на данном этапе функционирует в рамках одного предложения!). Рассмотрим предыдущий контекст нашего примера (Bruno Šul’c (Schulz), Pan, в: *Traktat o manekenach. Sbranie prozy*. Per. s pol’skogo Leonida Civ’jana. SPb.: Inapress, 2000, с. 57):

(20) Буйное некошенное разнотравье мохнатой шубой укрывало волнистую землю. Тут были (...) Все это, перепутанное и пушистое, было напоено ласковым воздухом, приправлено лазоревым ветром и пропитано небом. Стоило лечь в траву, и тебя накрывала голубоватая география облаков и плавающих континентов, ты начинал вдыхать карту небосвода. (...) А одно растение, желтое, с бледным надутым стеблем, наполненным млечным соком, само выбрасывало из своих побегов воздух и пух в виде перистых молочно-белых *шаров*, которые рассыпались от малейшего дуновения и беззвучно *впитывались* в голубую тишину (курсив наш — Т. Р.).

Путь к пониманию авторской метафоры открывают её богатые текстовые связи, в том числе очевидно лексические. Пройдемся обратно по тексту:

(21) Таблица 5

<i>шар</i>		<i>впитываться</i>	<i>тишина</i>	<i>голубая</i>
			Еpit (небо)	голубое
<i>воздух и пух в виде шаров</i>		<i>рассыпаться</i>		
<i>воздух из побегов</i>				
	<i>Ты</i>	<i>вдыхать</i>	<i>карта небосвода</i>	
	<i>Тебя</i>	<i>накрывать</i>	<i>география облаков</i>	<i>голубоватый</i>
<i>в траву</i>		<i>лечь</i>		
		<i>припитано</i>	<i>небом</i>	
<i>воздухом</i>		<i>напоено</i>		
<i>разнотравье</i>		<i>укрывать</i>		

На наш взгляд общий ключ к решению проблем авторской метафоры в МСТ — это постройка семантических сетей хотя бы целых абзацев. Помимо того, очевидную роль должны играть экстралингвистические факторы, модели мира и фреймы (ср. позицию лежащего в траве человека, смотрящего в небо и видящего над собой растения).

4. Перспективы трактовки метафоры в МСТ мы видим в следующих направлениях:

4.1. Разработка теории метафоры как части метатеории словаря — это, прежде всего, контроль над полнотой/системностью описания разных классов лексем и метафор в ТКС.

4.2. Разработка словарных статей для значений ЛФ, содержащих метафорические значения слов. Примеры см. выше (ВЕСТИ, ПИТАТЬ, РАЗГОРЕТЬСЯ).

4.3. Формализация результатов собственно лингвистических исследований метафоры в рамках МСТ по следующим линиям (ср. идеи из сборника Arutjunova 1999, 313—402):

(а) По линии видов дискурса, типологии текстов и функциональных стилей (ср., с одной стороны, законодательство, судопроизводство и руководства к эксплуатации приборов, где метафора встречается редко, а, с другой стороны, язык рекламы или поэзии, где она обязательно наличествует);

(б) По линии синтаксиса: подлежащее (метонимия) vs. сказуемое (метафора);

(в) По линии коммуникативной структуры: тема vs. рема;

(г) По типам номинации: конкретная референция vs. идентифицирующая функция;

(д) По линии значения: конкретно-физическое значение vs. абстрактное значение (*вода течет, жизнь течет, время течет, мысли текут* и т. д. (ср. Arutjunova 1999, 374: «Без метафоры не существовало бы лексики “невидимых миров” (внутренней жизни человека), зоны вторичных предикатов, т. е. предикатов, характеризующих абстрактные понятия»);

(е) По линии «пяти чувств»: установление сходства между областями, воспринимаемыми разными органами чувств, ср. явление синестезии: *твердый металл и твердый звук, теплый воздух и теплый тон* (Arutjunova 1999, 374).

4.4. Исследование и формализация механизмов синтеза метафор в направлении Смысл ⇒ Текст: Как сделать так, чтобы заданный смысл выражался «нестандартно, метафорично»? («моделирование писателя»).

4.5. Исследование и формализация механизмов анализа метафор в направлении Текст ⇒ Смысл: Как сделать так, чтобы «нестандартный, метафорический» текст был хоть как-то интерпретирован, т. е. чтобы тексту была поставлена в соответствие одна или несколько семантических сетей? («моделирование лингвиста-литературоведа»).

4.6. Исследование и формализация механизмов обработки метафор в процессе перевода, т. е. в направлении Текст ⇒ Смысл ⇒ Текст. Тут возможны два противоположных случая: дословный перевод или замена по смыслу (*ты мне нужна как воздух — ты мне нужна как вода* (к жизни), ср. Iordanskaja/Mel'čuk 1980); ср. также проблемы перевода неологизмов (терминологических метафор) в философском трактате или в техническом тексте (*hard disk — винчестер — твердый диск — хард-диск*).

4.7. Исследование и формализация механизмов динамического развития лексики. Как сделать так, чтобы были охвачены процессы постоянной «метафоризации/полисемизации» в языке?

Литература

- Apresjan, Ju. D. 1974a. *Leksičeskja semantika. Sinonimičeskie sredstva jazyka*, Moskva, 1974.
- Apresjan, Ju. D. 1974b. Regular polysemy, *Linguistics* 142, 1974. P. 5—32 (= *Izvestija AN SSSR, SLJa XXX*, 1971. P. 509—523).
- Apresjan, Ju. D. 2003. Fundamental'naja klassifikacija predikatov i sistemnaja leksikografija, In: Chrakovskij, V. S. (otv. red.), *Grammatičeskie kategorii: ierarchii, svjazi, vzaimodejstvie. Materialy meždunarodnoj konferencii*, SPb., 2003. P. 7—21.
- Apresjan, Ju. D. 2004. Lingvističeskaja terminologija slovarja, In: *Novyj ob''jasnitel'nyj slovar' sinonimov russkogo jazyka*. Pod obščim rukovodstvom akad. Ju. D. Apresjana, Moskva-Vena (= Wiener Slawistischer Almanach, Sonderband 60), XXII—LIV.
- Arutjunova, N. D. 1999. *Jazyk i mir čeloveka*, Moskva.
- Iordanskaja, L. N., Mel'čuk, I. A. 1980. Konnotacija v lingvističeskoj semantike, *Wiener Slawistischer Almanach* 6, 1980. S. 191—210.
- Lakoff, G., Johnson, M. 1980. *Metaphors We Live By*, Chicago-London, 1980.
- Mel'čuk, I. A., Clas, A., Polguère, A. 1995. *Introduction à la lexicologie explicative et combinatoire*, Louvain-la-Neuve: Duclot, 1995.
- Padučeva, A. V. 2004. *Dinamičeskie modeli v semantike leksiki*, Moskva, 2004.
- Rachilina, E. V. 2000. *Kognitivnyj analiz predmetnyh imen: semantika i sočetaemost'*, Moskva, 2000.
- Reuther, T. 1996. On Dictionary Entries for Support Verbs: The Cases of Russian VESTI, PROVODIT' and PROIZVODIT', In: Wanner, L. (ed.), *Lexical Functions in Lexicography and Natural Language Processing*, Amsterdam-Philadelphia, 1996. P. 181—208.
- Ricœur, P. 1975. *La métaphore vive*, Paris, 1975.
- Словари:
- BAS 1959. *Slovar' sovremennogo russkogo literaturnogo jazyka v 17 tomach*, T. 9, Moskva, 1959.
- DEC 1992. Mel'čuk, I. A. (ed.), *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques, III*, Montréal: Les Presses de l'Université de Montréal, 1992.
- MAS 1984. *Slovar' russkogo jazyka v četyrech tomach*, 2-e izd. Tom III, Moskva, 1984.
- TKC 1984. Mel'čuk, I. A., Žolkovskij, A. K. *Tolkovo-kombinatornyj slovar' sovremennogo russkogo jazyka. Opyty semantiko-sintaksičeskogo opisanija ruskoj leksiki*, Vena (= Wiener Slawistischer Almanach, Sonderband 14), 1984.

An Attempt of Implementing the Generative Grammar Island Constraints in the «Meaning — Text» Model *

Elena Rudnitskaya

IV RAN
erudnits@rambler.ru

The paper considers a number of Island Constraints of the Generative Grammar and options of integrating these constraints into the «Meaning — Text» Model as Island Filters on the stage of synthesis/ linearization. The proposal is that a Filter includes (1) a fragment of the Surface Syntactic tree corresponding to a particular Island, and (2) constraints on placement of the wh-word contained in the Island in the corresponding string after linearization.

Попытка реализации островных ограничений генеративной грамматики в модели «Смысл — Текст». Статья посвящена возможности интеграции некоторых Островных Ограничений Генеративной Грамматики в модель «Смысл — Текст» в качестве «Островных Фильтров» на стадии синтеза/линеаризации. Мы предлагаем, что Фильтр состоит из (1) фрагмента Поверхностно-Синтаксического дерева, соответствующего отдельно взятому Острову, и (2) ограничений на позицию вопросительного слова, содержащегося внутри Острова, в цепочке слов после преобразования линеаризации.

1. Introduction

The Island Constrains (Ross 1967/1986) are one of the major empirical grounds of the Movement Theory of the Generative Grammar. In English and some other (mostly European) languages, the wh-word (interrogative word) in a wh-question is placed clause-initially instead of staying *in situ* (in its «base» (argument) position, in which its non-interrogative counterpart would be). This displacement of the wh-word is called **wh-movement**: the «moved» wh-NP is sentence-initial (1a),

* The author would like to thank Igor Boguslavsky and anonymous Reviewers for useful comments.

whereas the empty «trace» NP stays in the base-position (1b). In a considerable number of languages, the wh-word remains *in situ* (2a-b)¹.

- (1) a. *John bought* [_{NP} *a pencil*] b. [_{WH-NP} *What*] *did John buy* *t*_{WH-NP}?
 (2) a. *Chelswu-nun* [_{NP} *chayk-ul*] *sa-ss-ta* [*Korean*]
 Chelswu-TOP [_{NP} *book-ACC*] *buy-PAST-DECL*
 «Chelswu bought books»
 b. *Chelswu-nun* [_{WH-NP} *mwues-ul*] *sa-ss-ni?*
 Chelswu-TOP [_{WH-NP} *something-ACC*] *buy-PAST-QUEST*
 «What did Chelswu buy?»

Wh-movement (fronting of the wh-word) is usually allowed, but it is banned in case of certain syntactic constructions inside the sentence, and then the wh-question is ungrammatical. These constructions are called «Islands», so if the «base» position of a wh-word is inside an «Island», the wh-word cannot be fronted, or moved out of the «Island». These cases are called Island Constraints. The most typical Islands are a Complex Noun Phrase (CNP), (a clause which is complement to a noun or a relative clause modifying a noun) (3a-b), and a Coordinate Structure (5). In (4a-c), it is shown that if a wh-word is moved out of an embedded clause which is not Island, the question is grammatical.

- (3) a. *_[WH-NP Who] *did Piro make* [_{CNP} *the claim* [_{SENT-COMPL} *that he saw* *t*_{WH-NP} *last week*]]? [Haegeman 1994: 401]
 b. *_[WH-NP Who] *do you know* [_{CNP} *the town*_{NP-REL} [_{REL-CL} (*where*) *Bill met* *t*_{WH-NP} *e*_{NP-REL}]]?
 («You know the town (where) Bill met somebody, and who was it?»)
 (4) a. _[WH-NP Who] *did Piro claim* [_{SENT-COMPL} *that he saw* *t*_{WH-NP} *last week*]]? [Haegeman 1994: 401]
 b. _[WH-NP Who] *do you think* [_{SENT-COMPL} *that John wants to invite* *t*_{WH-NP}]]?

¹ The claim that in some languages (such as Chinese, Japanese, Korean, etc.), wh-words remain *in situ* and are not sentence-initial (do not undergo syntactic movement) does not imply that the corresponding language has no syntactic wh-movement. Such a claim is only one alternative (Haegeman 1994: 494–500). Another alternative is to say that all languages have syntactic wh-movement, but if the wh-word stays *in situ*, a null/abstract wh-operator undergoes syntactic wh-movement instead of the wh-word (Watanabe 1992). Also, note that a wh-word in a wh-question such as (2b) can front optionally. Optional fronting is called scrambling, which is arguably movement, and, in any case, is not wh-movement (wh-movement is obligatory).

- (5) a. *[_{WH-NP} *What*] *did John buy* [_{COORD-P} *t_{WH-NP} and sugar*]?
 b. *[_{WH-NP} *What*] *did John buy* [_{COORD-P} *sugar and t_{WH-NP}*]?²

Other Islands which are present in English and a number of other languages are Adjunct (adverbial clause), Sentential Subject and Left Branch Islands², cf. (6)–(8).

- (6) *[_{WH-NP} *What*] *will you be mad* [_{CADJ-CLAUSE} *if John buys t_{WH-NP}*]?
 (7) a. *[_{WH-NP} *Which candidate*] *did they say* [_{SENT-SUBJ} *to get t_{WH-NP} to agree to a debate*] *was hard*? [Merchant 1999: 222]
 b. [_{WH-NP} *Which candidate*] *did they say* [_{SUBJ} *it*] *was hard* [_{SENT-COMPL} *to get t_{WH-NP} to agree to a debate*]? [Merchant 1999: 222]
 (8) *[_{WH-AP} *Whose*] *did you see* [_{OBJ-NP} *t_{WH-AP} car*]? [Merchant 1999: 192]

There are a lot of formal explanations trying to account for all or most of Islands Constraints, e.g. the Subjacency/ Bounding Nodes account by Chomsky 1973, the Barriers account by Chomsky 1986, the Economy Principles of Movement account by Chomsky 1995. On the other hand, at least some of the Island Constraints are consistent with an Information Structure («functional») account, which does not contradict any of their syntactic account (e.g. Erteschik-Shir & Lappin 1979). In section 2, we will briefly outline Erteschik-Shir & Lappin's explanation, and in section 3, we will try to couch some of the Island Constraints mentioned above in the «Meaning — Text» framework (Mel'čuk 1974/1999, Apresian et al. 1989, Mel'čuk 2001).

2. The Information Structure account of Island Constraints

Erteschik-Shir and Lappin 1979 undertake an attempt to reformulate Island Constraints in terms of Information Structure units («topic»/ «theme» vs. «focus»/ «rheme», «foreground» vs. «background») rather than syntactic constituents (Subject vs. Object). Their main claim is that an NP can only be moved out of a focal constituent, but movement is often disallowed from a topical constituent, or from a constituent which is background. For instance, Subject is most often topic, a relative clause is typically background, whereas Object is usually focus.

² Movement of the Left Branch of an NP is allowed in Russian. We do not illustrate other Island-like effects, such as subject-object and argument-adjunct asymmetries in details, as well as WH-Islands effects, for shortness of presentation. The issue of scrambling (as opposed to wh-words movement) will not be discussed either.

Therefore, movement out of a Subject constituent is banned (Sentential Subject Constraint, (7a)); movement out of a relative clause is banned (CNP Constraint, (3b)); but an NP can be moved out of a sentential Object/Complement, cf. (4a-b) and (7b). This correlation between grammatical and Information Structure categories works best for languages without scrambling, such as English, but is also partially valid for other languages (e.g. Russian)³.

Since Information Structure is to a considerable degree incorporated into syntactic structure in English (cf. the use of the Passive Voice transformation), Island Constraints mentioned above in English are to be formulated in terms of syntactic rather than information structure. However, Erteschik-Shir & Lappin 1979 give exceptional examples even for English. For instance, movement out of a relative clause is possible if the matrix (main) clause is a «there» (existential) clause, cf. (9a-b) from Erteschik-Shir & Lappin 1989: 58⁴.

- (9) a. *This is* [_{MOVED-NP} *the kind of weather*] [_{MATR-CLAUSE} *that there are many people* [_{REL-CLAUSE} *who like* *t*_{MOVED-NP}]]
 b. **This is* [_{MOVED-NP} *the kind of weather*] [_{MATR-CLAUSE} *that I am familiar with many people* [_{REL-CLAUSE} *who like* *t*_{MOVED-NP}]]

As mentioned above, relative clause in English is usually background, that is why it is an Island for movement. However, since «there»-clauses are also usually background, the relative clause in (9a) is exceptionally foreground; thus, movement of *the kind of weather* out of this clause is possible.

Regarding CNP Islands, in which the «that» clause is a complement (3a), Erteschik-Shir and Lappin present judgements varying depending on

³ For instance, in Russian, movement out of postverbal subjects is allowed, cf. (7a-b) and (i a-b).

- (i) a. ?* [_{WH-NP} *Chto*] *ty xochesh' chtoby* [_{PREV-SUBJ} *kupit' t*_{WH-NP}] *bylo legko?*
 b. ?? [_{WH-NP} *Chto*] *ty xochesh' chtoby bylo legko* [_{POSTV-SUBJ} *kupit' t*_{WH-NP}]?
 «What do you want to be easy to buy?»

⁴ Erteschik-Shir & Lappin assume that the NP *this kind of weather* in (9a-b) is moved from the object position in the relative clause. However, this is a less obvious case of «movement» than wh-movement, in particular, wh-movement is disallowed in both (ii a-b) parallel to (9a-b).

- (ii) a. * [_{MOVED-NP} *Which weather*] [_{MATR-CLAUSE} *are there many people* [_{REL-CLAUSE} *who like* *t*_{MOVED-NP}]]?
 b. * [_{MOVED-NP} *Which weather*] [_{MATR-CLAUSE} *are you familiar with many people* [_{REL-CLAUSE} *who like* *t*_{MOVED-NP}]]

the lexical content of the matrix (main) verb. In particular, they claim that (3a) is degraded, but acceptable (cf. (10a)). The contrast is illustrated in (10a-b) (Erteschik-Shir & Lappin 1979: 61).

- (10) a. [?]_[WH-NP What] *did she make* [_{CNP the claim} [<sub>SENT-COMPL that he had done t_{WH-NP}]]]?
 b. ^{*}_[WH-NP What] *did she discuss* [_{CNP the claim} [_{SENT-COMPL that he had done t_{WH-NP}]]]?}</sub>

As Erteschik-Shir and Lappin put it, *make* in (10a), unlike *discuss* in (10b), belongs to basic set of verbs «used to introduced the NP» (*claim*) and adds no semantic content to this NP. Therefore, *make* is background, whereas *discuss* is foreground. The relative clause in (10a) is foreground, so the wh-word *what* can move out of it⁵.

To conclude, Erteschik-Shir & Lappin try to present most constraints on (wh)-movement as related to «background», in particular those that have systematic correlates in the syntactic structure («sample» constraints from section 1)⁶. The «background» explanation would probably work better in non-configurational languages than in configurational ones, but this issue requires more research.

3. Island Constraints and the «Meaning — Text» model

The main levels of the traditional Generative Grammar version are Deep Structure and Surface Structure (Chomsky 1981). (Wh-)-movement (in English) proceeds between Deep Structure and Surface Structure, whereas Island Constraints are constraints on well-formed Surface Structure.

First, we will consider the Surface Syntax Representation fragments of the «Meaning — Text» model (Mel'čuk 1974/1999, Apresian et al. 1989)

⁵ A similar contrastive pair is probably available in Russian, cf. (10a-b) vs. (i a-b). In any case, (i c) with a relative clause CNP is more ungrammatical than (i a-b) with complement CNP-s.

- (i) a. [?]_[WH-NP Čto] *on vnes* [_{CNP predloženie} [<sub>SENT-COMPL organizovat' t_{WH-NP}]]]?
 b. ^{???}_[WH-NP Čto] *oni obsuždali* [_{CNP predloženie} [<sub>SENT-COMPL organizovat' t_{WH-NP}]]]?
 c. ^{*}_[WH-NP Čto] *on nanjal* [_{CNP special'nogo sotrudnika} [_{SENT-COMPL organizovat' t_{WH-NP}]]]?}</sub></sub>

⁶ We omit Erteschik-Shir & Lappin's discussion of less familiar Islands, such as the Factive Island.

which can account for some Island Constraints (section 3.1), and second, certain Lexical Functions from the model (Mel'čuk 1974/1999: 78—109) related to certain Island Constraints-like phenomena presented in section 2 (3.2). Finally, some concepts of the Communicative Structure in the «Meaning — Text» model (Mel'čuk 2001) will be shortly considered in their relation to Erteschik-Shir & Lappin's account (3.3).

3.1. The Surface Structure in the Generative Grammar is not equivalent to the Surface Structure Representation in the «Meaning — Text» model. Though the former is a constituent tree, it is linear, that is, the (essential) word order in the resulting sentence is the same as in the Surface Structure. The latter is a dependency tree, in which the linear order of its nodes is not specified.

As Mel'čuk 1974/1999: 268—289 notes, linearization of the nodes of the Surface Structure tree proceeds at the stage of converting the Surface Structure Representation into the Deep Morphological Representation. Linearization rules are mainly based on types of syntactic relations in the Surface Syntax tree (see Mel'čuk 1974/1999: 221—235, Apresian et al. 1989: 71—121, 204—215, Boguslavsky et al. 2004: 9—22). For instance, the subject connected with the verb by the «predicative» relation in the tree must precede the verb in the linear string (e.g. *John left*); a preposition must precede its complement (e.g. *to school*), etc.

The main rules of linearization listed in Mel'čuk 1974/1999: 268—289 do not include special rules for questions. For instance, *bought* and *books* in (i) *Mary bought books* are connected by the same «1-st complement» (1-compl) relation as *buy* and *what* in (ii) *What did Mary buy?*

Apresian et al. 1989: 242—260 (based on Apresian 1985) include «syntactic features» (*syntaksičeskie priznaki*) into the model. Syntactic features are assigned to words (lexemes) in the dictionary and determine or restrict the set of real grammatical constructions (syntagmas) in which these words are used. For instance, the feature ADRES is assigned to bi-transitive verbs (Apresian et al. 1989: 245, cf. *He gave Mary a book*). Interrogative pronouns and adverbs (wh-words) are assigned the feature INTERROG that, among other restrictions, determines their displacement to the beginning of the clause/sentence. For instance, *what* in (ii) above, unlike *books* in (i), has the feature INTERROG, and its position is therefore clause-initial rather than postverbal⁷.

⁷ The clause-initial position of an INTERROG word causes the syntactic tree corresponding to an interrogative clause (question) to be non-projective, so that wh-questions are treated as special syntagmas and must be handled by a separate group of linearization rules.

The INTERROG syntactic feature does not solve the Island Constraints problem, so a wh-question with an initial wh-word «originating» from an island (see sections 1-2 above) must be ruled out by additional filters or conditions. In the computer implementations of the «Meaning — Text» model (Apresian et al. 1989: 19—22), filters are only used on the stage of syntactic analysis (for the Surface Structure tree well-formedness), but not on the stage of synthesis and linearization. Therefore, it would be good to introduce a set of special conditions (or filters) on the stage of linearization. Let us call these conditions Island Filters.

If Island Constraints are regarded as syntactic (see section 1), Island Filters have to be formulated in terms of Surface Structure relations (as some of the syntactic features are). For instance, in case of the CNP Island (3a-b) are repeated below), the Filter would have to specify the set of relations in the embedded clause inconsistent with the sentence-initial position of the wh-word.

- (3) a. *[_{WH-NP}Who] *did Pairo make* [_{CNP} *the claim* [_{SENT-COMPL} *that he saw* _{t_{WH-NP}} *last week*]]? [Haegeman 1994: 401]
 b. *[_{WH-NP}Who] *do you know* [_{CNP} *the town*_{NP-REL} [_{REL-CL} (*where*) *Bill met* _{t_{WH-NP}} _{c_{NP-REL}}]]?
 («You know the town (where) Bill met somebody, and who was it?»)

Let us consider the most prototypical instance of the CNP Island (3b) in detail. In the Surface Structure tree of (3b), the relevant relations⁸ are:

- (11) V (know) — (1-compl) → N (town) — (relat) → V (meet) — (1-compl) → PRON [INTERROG] (who)

The corresponding fragments of the Surface Structure tree of (3a) and of the grammatical (4a) would be (13a-b):

- (4) a. [_{WH-NP}Who] *did Pairo claim* [_{SENT-COMPL} *that he saw* _{t_{WH-NP}} *last week*]]? [Haegeman 1994: 401]
 (12) a. V (make) — (1-compl) → N (claim) — (1-compl) → CONJ (that) — (subord-conj) → V (see) — (1-compl) → PRON [INTERROG] (who)
 b. V (claim) — (1-compl) → CONJ (that) — (subord-conj) → V (see) — (1-compl) → PRON [INTERROG] (who)

The relevant difference between (12a) and (12b) is the N node (*claim*) which is daughter of an argument (e.g. 1-compl) relation of the matrix

⁸ The names of relations are taken from the list of syntactic relations for English in Apresian et al. 1989: 71—121.

clause verb and mother of a clause argument (e.g. 1-compl). This node is present in (12a) and absent in (12b) and rules out the wh-question with a fronted wh- (or [INTERROG]) word (3a).

The condition ruling out (3a-b) (in a more generalized form) would be formulated as follows:

- (13) a. $V \text{ — (predic/ n-compl) } \rightarrow N \rightarrow \text{(relat/n-compl)} \rightarrow V \text{ — (predic/n-compl) } \rightarrow \text{PRON/ ADV [INTERROG]}$
 b. If a pronoun/ adverb (**Wh**) which has an [INTERROG] feature is daughter of an «n-compl»/«adverb» relation, so that its mother is a verb (**V**) and daughter of a «relat»/ «n-compl» relation, and **V**'s mother is a **noun N** being daughter of a «predic»/«n-compl» relation, **Wh** cannot be sentence-initial.
 ***Wh** $N_{\text{SUBJ}} V N \dots (N_{\text{EMB-SUBJ}}) V (N_{\text{EMB-ARG}}^*)$.

There are two more issues to address in relation to the [INTERROG] feature of wh-words and the condition in (13b). First, as (14) shows, a wh-word need not be sentence-initial, but it can be clause-initial in an embedded clause which is a wh-question (14a). In this case, the CNP constraint is also valid, cf. (3a) and (14b):

- (14) a. *I know* [_{EMB-QUEST} *who*_{WH-NP} [*Poirot claimed* [_{SENT-COMPL} *that he saw* *t*_{WH-NP} *last week*]]] [The Reviewer's example]
 b. **I know* [_{EMB-QUEST} *who*_{WH-NP} [*Poirot denied* [_{CNP} *the claim* [_{SENT-COMPL} *that Mary saw* *t*_{WH-NP} *last week*]]]]

In (14b), it is crucial that *who* moves out of the CNP, but it is not sentence-initial. Thus, the condition in (13b) must be modified so that the banned position of the wh-word be defined not only in terms of linear order but also of syntactic relations:

- (13) b'. If a pronoun/adverb (**Wh**) which has an [INTERROG] feature is daughter of an «n-compl»/ «adverb» relation, so that its mother is a verb (**V**) and daughter of a «relat»/ «n-compl» relation, and **V**'s mother is a **noun N** being daughter of a «predic»/«n-compl» relation, **Wh** cannot immediately precede $N_{\text{(SUBJ)}}$ or the first sister of N in the linear string corresponding to the tree fragment described above)⁹.
 ...***Wh** $(N_{\text{SUBJ}}) V (N) \dots (N_{\text{EMB-SUBJ}}) V (N_{\text{EMB-ARG}}^*)$.

⁹ In (13b-b') and (16c), we disregard the auxiliary (*do*), which appears only in matrix but not in embedded wh-questions, for the sake of a more generalized version of the rule.

For the Coordinate Structure Island (5a-b), the relevant fragments of the Surface Syntax tree would be (15a-b), and the condition (16a-c) would rule (5a-b) out.

- (5) a. $*[_{\text{WH-NP}} \textit{What}] \textit{did John buy} [_{\text{COORD-P } t_{\text{WH-NP}}} \textit{and sugar}]?$
 b. $*[_{\text{WH-NP}} \textit{What}] \textit{did John buy} [_{\text{COORD-P } \textit{sugar and } t_{\text{WH-NP}}}]?$
- (15) a. $V \text{ (buy)} \text{ --- } (1\text{-compl}) \rightarrow \text{PRON [INTERROG]} \text{ (what) --- } (\text{coord}) \rightarrow \text{CONJ (and) --- } (\text{coord-conj}) \rightarrow N \text{ (sugar)}$
 b. $V \text{ (buy)} \text{ --- } (1\text{-compl}) \rightarrow N \text{ (sugar)} v \text{ --- } (\text{coord}) \rightarrow \text{CONJ (and) --- } (\text{coord-conj}) \rightarrow \text{PRON [INTERROG]} \text{ (what)}$
- (16) a. $V/N/P/... \text{ --- } (\text{predic}/n\text{-compl}/\text{adverb}/\text{prepos}/...) \rightarrow \text{PRON/ADV[INTERROG]} \rightarrow (\text{coord}) \rightarrow \text{CONJ [COORD]} \text{ --- } (\text{coord-conj}) \rightarrow N,$ or
 b. $V/N/P/... \text{ --- } (\text{predic}/n\text{-compl}/\text{adverb}/\text{prepos}/...) \rightarrow N \rightarrow (\text{coord}) \rightarrow \text{CONJ [COORD]} \text{ --- } (\text{coord-conj}) \rightarrow \text{PRON/ADV [INTERROG]}$
 c. If a pronoun/adverb (**Wh**) which has an [INTERROG] feature is daughter of a «predic»/«n-compl»/«adverb»/«prepos» relation and mother of a **coordinate conjunction (Conj)**; or if **Wh** is a daughter of a «coord-conj» relation (a daughter of a **Conj**), **Wh** cannot be the first among its sisters in the linear string corresponding to the tree fragment described above.
 $... *Wh N_{\text{SUBJ}} V N \text{ Conj}, \text{ or } ... *Wh N_{\text{SUBJ}} V \text{ Conj } N.$

Another interesting problem is treatment of clauses with two or more wh-words. In English, only one of the wh-words can be fronted, cf. (17c)¹⁰. Also, the fronted wh-word must be the highest in the valence/ argument hierarchy: cf. (17b), in which the 2-compl wh-phrase *for whom* is fronted, whereas the 1-compl wh-phrase *which book* remains in its regular position.

- (17) a. *John wondered* [EMB-QUEST [WH-NP-1-COMPL *which book*] *Bill bought* $t_{\text{WH-NP-1-COMPL}}$ [WH-NP-2-COMPL *for whom*]]
 b. $*\textit{John wondered}$ [EMB-QUEST [WH-NP-2-COMPL *for whom*] *Bill bought* [WH-NP-1-COMPL *which book*] $t_{\text{WH-NP-1-COMPL}}$]
 c. $*\textit{John wondered}$ [EMB-QUEST [WH-NP-1-COMPL *which book*] [WH-NP-2-COMPL *for whom*] *Bill bought* $t_{\text{WH-NP-1-COMPL}}$ $t_{\text{WH-NP-2-COMPL}}$] [Haegeman 1994: 384]

The problem of questions with several wh-words (related to the so-called Superiority Effects) is relevant for Island Filters, at least in lan-

¹⁰ We do not consider languages, in which more than one wh-words/ phrases can be fronted, such as Russian:

(i) $[[\check{C}to]_1 [dlja \ kogo]_2] \textit{Ivan kupil } t_1 \ t_2?$

guages such as English. However, Superiority Effects are usually treated separately from other Island Constraints. For instance, in case of a CNP Island with several wh-words/phrases inside the CNP, none of them can be moved out; cf. (3b), (17a), and (18):

- (3) b. *[_{WH-NP} *Who*] *do you know* [_{CNP} *the town*_{NP-REL} [_{REL-CL} (*where*) *Bill met* *t*_{WH-NP} *e*_{NP-REL}]]?
 («You know the town (where) Bill met somebody, and who was it?»)
- (18) *[_{WH-NP-1-COMPL} *What*] *do you know* [_{CNP} *the party*_{NP-REL} [_{REL-CL} (*during which*) *Bill gave* *t*_{WH-NP-1-COMPL} [_{WH-NP-2-COMPL} *to whom*] *e*_{NP-REL}]]?
 («You know the party, during which Bill gave something to someone, and who and what were these?»)

To conclude, additional Island Filters needed to rule out ungrammatical wh-questions due to Island Constraints have to be formulated as follows. Sentences with certain Surface Structure tree fragments (corresponding to Island configurations and containing wh-words ([INTERROG] pronouns/ adverbs)), at the stage of linearization, do not allow placement of the wh-word before the linear string corresponding to the Island tree fragment. These Filters have to be applied at the stage of converting the Surface Structure tree Representation into the linear Deep Morphological Representation¹¹.

3.2. The Factive Island and the Oper₁ Lexical Function (Mel'čuk 1974/1999: 94). As we have shown in section 2, Erteschik-Shir & Lappin 1979 propose the «background — foreground» account of certain Island-like phenomena, such as (10a-b). In (10a), the main verb *make* (*the claim*) has fewer lexical semantics features than *discuss* (*the claim*) in (10b), and therefore is background. That is why the CNP is foreground in (10a), and movement out of this CNP is less degraded in (10a) than in (10b).

¹¹ The Reviewer raises the question of whether, instead of a separate Filter for each Island Constraint, there can be proposed «some generalized version of the rule that can capture a large number of different Island Constraints». So far, our answer to this question is negative. First, even in the Generative Grammar tradition, most of the attempts to treat all the Island Constraints using one principle or condition (mentioned in the end of section 1) actually COULD NOT account for all of the major Island Constraints. Second, in each of the Island Constraints, a different syntactic configuration with its set of syntactic relations is involved. That makes it very difficult to formulate a unitary condition which would include, for instance, both (13b-b') (for the CNP Island) and (16c) (for the Coordinate Structure Constraint) in a more generalized form.

According to the Lexical Function Theory, in (10a), the main verb *make* in the phrase *make the claim* is an instance the Oper₁ lexical function (a lexically deficient word) — «the verb connecting the noun — the 1-st actant/ argument (the subject [*she*]) with the noun denoting the situation (the 1-st complement [*claim*])» (p. 94), e.g. Oper₁ (*claim*) = *make*. In the fully ungrammatical (10b), on the other hand, *discuss* is an instance of a regular, non-deficient verb.

This analysis accounts for the background status of *make* in (10a) and, correspondingly, the foreground status of the sentential complement clause (which makes wh-movement more grammatical) in terms of Lexical Functions. Lexical Functions are part the Semantic Representation and are also concerned with certain fragments of the Deep Structure tree (pp. 191—208); they are substituted by words in the Surface Structure tree (pp. 237—238). Thus, filters which distinguish wh-questions in (10a) and (10b) have to be based on fragments of the Deep Structure Representation. For instance, the Deep Structure Representations of the relevant fragments of (10a-b) will be (19a-b).

- (19) a. Oper₁ — 2 → claim b. discuss — 2 → claim

3.3. The Information Structure option of interpreting the Island Constraints (Mel'čuk 2001: 181—210). In section 2, an attempt has been presented to formulate Island Constraints in terms of Information Structure, in particular, «foreground — background». It was claimed that a wh-word cannot be moved out of a background fragment of a sentence. For instance, since relative clauses are usually background, they are Islands for wh-movement. Mel'čuk 2001: 199—210 proposes a different understanding of «foreground — background», these concepts are tied in Mel'čuk's understanding primarily to the Semantic and Deep Syntactic Representations, in particular, to Syntactization (e.g. regular vs. raised possessor). Mel'čuk (pp. 200—201) proposes to analyze the «backgrounding» effect in relative clauses as a case of «secondary Thematization» in the Semantic Representation. For instance, the main Theme (T₁), the Rheme (R₁) and the secondary Theme (T₂) of (17a) (p. 200) are shown in (20b). The secondary Theme T₂ would be the node *a man* belonging to T₁.

- (20) a. A man [_{REL-CLAUSE} who was wearing a frilly dress] came into the room
 b. [_{T₁} [a frilly dress]] ← 2 — [was wearing] — 1 → [_{T₂} [a man]]]
 ← 1 — [_{R₁} [came]] — 2 → [the room]]

Thus, there can not be found a straightforward parallel between Mel'čuk's Semantic Representation-based concept of «foreground —

background» and the widespread use of these terms. Mel'čuk 2001: 197 does not consider fronting (as Focalization) in detail (probably including wh-fronting) because (1) it is a «strictly word order operation» (p. 97), and (2) because both the Theme and the Rheme from the Semantic Representation can be fronted.

To conclude, it is difficult to find in the «Meaning — Text» model a fragment corresponding to Erteschik Shir & Lappin's «foreground — background» because Erteschik-Shir & Lappin's «foreground — background» opposition is formulated in terms of Information Structure, whereas Mel'čuk's 2001 Communicative Structure concepts are strictly tied to the Semantic (and Deep Syntactic) Representation.

References

- Apresian, Yu. D. (1985). «Sintaksičeskie priznaki leksem» [Syntactic features of lexemes]. In: *Russian Linguistics* 9.2/3. 1985. P. 289—318.
- Apresian, Yu. D. et al. (1989). Yu. D. Apresian, I. M. Boguslavsky, L. L. Iomdin, A. V. Lazursky, N. V. Pertsov, V. Z. Sannikov, L. L. Cinman. *Lingvističeskoe obespečenie sistemy ETAP-2* [Linguistic Support of the ETAP-2 System]. Moskva, «Nauka», 1989.
- Boguslavsky, I. M. et al. (2004). *Morfo-sintaksičeskaja i leksiko-semantičeskaja razmetka podmassiva sbalansirovannogo korpusa* [Morpho-Syntactic and Lexico-Semantic Marking of a Sub-Massive of Balanced Corpora]. Yearly report, IPPIRAN. 2004.
- Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht, Foris, 1981.
- Erteschik-Shir, N. and S. Lappin (1979). «Dominance and the Functional Explanation of Island Phenomena». In: *Theoretical Linguistics* 6.1. 1979. P. 41—86.
- Haegeman, L. (1994). *Introduction to Government and Binding Theory*, 2nd edition. Blackwell, Cambridge (Mass.), 1994.
- Mel'čuk, I. A. (1974/1999). *Opyt teorii lingvističeskix modelej «Smysl — Tekst»* [The Linguistic «Meaning — Text» Models Experience]. Moskva, «Yazyki russkoj kul'tury», 1999.
- Mel'čuk, I. A. (2001). *Communicative Organization of Natural Language. The Semantic-Communicative Structure of Sentences*. John Benjamins, Amsterdam/Philadelphia, 2001.
- Merchant, J. (1999). *The Syntax of Silence: Sluicing, Islands, and Identity in Ellipsis*. PhD Dissertation, University of California, Santa Cruz, 1999.
- Ross, J. R. (1967/1986). *Constraints on Variables in Syntax /Infinite Syntax!.* PhD Dissertation, MIT, 1986.
- Watanabe, A. (1992). *S-structure Movement and Wh in situ*. Ms., MIT, 1992.

Individual-Level and Stage-Level Predicates: the Spanish Emotion Nouns

Begoña Sanromán Vilas

University of Helsinki
begona.sanroman@helsinki.fi

The goal of this paper is to present a set of characteristics of Spanish emotion nouns, aspectually classified as states, that shows the possibility to delimitate these stative predicates as individual-level or stage-level predicates. As a starting point we assume that there are two groups of Spanish emotion nouns according to the semantic role played by the cause/origin of the emotion: the internal cause emotion nouns (IC) and the external cause emotion nouns (EC). In IC nouns, such as *respeto* ‘respect’, the emotion is born in the experiencer as a result of a judgement made about an object. EC nouns, such as *asombro* ‘amazement’, require the existence of an external fact that triggers the emotion. This semantic difference between both groups is not isolated. We consider that it goes through all the grammatical levels in such a way that it is possible to outline morphological, syntactic and lexical co-occurrence patterns of behaviour. We shall use some of these patterns of behaviour — for instance, their paradigmatic and syntagmatic co-occurrence — as evidence for individual-level/stage-level contrast between the nouns.

1. Introduction

It is well known that within the aspectual class of states there is a distinction between individual-level and stage-level predicates. The goal of this paper is to present a set of characteristics of Spanish emotion nouns, aspectually classified as states, that shows the possibility to delimitate these stative predicates as individual-level or stage-level predicates.

There is no need to insist that the individual-level/stage-level contrast is not original. Starting from the initial work of Carlson in 1977 (apud Carlson 1995), there have been many linguists who have outlined a division within the aspectual class of states (see, e.g., Luján 1981, Bennett et al. 1990, Demonte 1999, Maienborn 2005). However, what gives some originality to this study is the application of this distinction to a group of

Spanish nominal predicates¹, because, as far as we know, there exist only studies dealing with verbal and adjectival predicates.

As a starting point we consider that there are two groups of Spanish emotion nouns according to the semantic role played by the cause/origin of the emotion: the internal cause emotion nouns (IC) and the external cause emotion nouns (EC) (Sanromán 2003)². In IC nouns, such as *respeto* ‘respect’, the emotion is born in the experiencer as a result of a judgement made about an object. On the other hand, EC nouns, such as *asombro* ‘amazement’, require the existence of an external fact that triggers the emotion. This semantic difference between both groups is not isolated. We consider that it goes through all the grammatical levels in such a way that it is possible to outline morphological, syntactic and lexical co-occurrence patterns of behaviour. We shall use these patterns of behaviour as evidence for individual-level/stage-level contrast between the nouns.

The research is carried out within the theoretical and methodological framework of the Explanatory and Combinatorial Lexicology (Mel’čuk et al. 1995), developed in the Meaning-Text Theory (see, e.g., Mel’čuk 1997). More specifically, the description of the lexical units (LUs) given in the *Dictionnaire explicative et combinatoire* (DEC) (see Mel’čuk et al. 1984–1999), in general, and those presented in the *Diccionario de colocaciones del español* (DICE) (see, e.g., Alonso Ramos 2002 and <http://www.dicesp.com>), in particular, have served as a helpful model (and corpus, in the case of DICE) throughout this study.

The paper is organized as follows: Section 2 deals with the nature of the aspectual class of states and the criteria given to distinguish individual-level from stage-level predicates. Section 3 presents two groups of Spanish emotion nouns and the most relevant differences between them. In Section 4, after a brief analysis of the problems entailed in the aspectual delimitation of emotion nouns, we compare these nouns with other

¹ A similar distinction to the one proposed here was made by Anscombre (1995) within French emotion nouns.

² The idea of two groups of emotions nouns has been specially developed within the French language. Anscombre (1995) and Mathieu (1999) talk about “*prédicats de sentiment endogènes*” when the emotion is born in the experiencer, and “*prédicats de sentiment exogènes*” when the emotion is created by an external fact. Van de Velde (1995: 79–82) makes a similar classification and uses the terms “*sentiments*” and “*états*”, respectively. Later, in Flaux and Van de Velde (2000: 87–92), the authors refer to the former ones as “*sentiments*” and to the latter ones as “*emotions*”, and they adopt the generic term “*affects*” to include both predicates.

Spanish nominal predicates as a way to enhance the aspectual differences among them. On the basis of these observations, a set of characteristics of the emotion nouns used as evidence for the distinction between individual-level/stage-level predicates is provided in Section 5. Specifically, we present the variation of emotion nouns in grammatical number (5.1) and some lexical units in paradigmatic (5.2) and syntagmatic (5.3) lexical co-occurrence with them. Section 6 summarizes the paper and draws some conclusions.

2. States as an aspectual class: individual-level and stage-level predicates

Aspectual classifications aim to delimitate different semantic classes of predicates taking into account the internal time of the events, that is to say, how these events are developed and distributed with time. In this paper we use the term “event” in its wider sense (Vendler 1967, Pustejovsky 1995), so we consider events as a general category consisting of four aspectual classes: states (*to know, to be, to love*), activities (*to read, to smoke*), accomplishments (*to eat an apple, to smoke a cigarette*) and achievements (*to arrive, to find, to win*).

Within the previous classification, stative events (states) are usually opposed to dynamic events (activities, accomplishments and achievements). According to this, emotion nouns are states because they are defined as durative and unbounded events; in other words, emotion nouns keep themselves without changes or progress along the period they last (durative states) and as they cannot progress, also they cannot direct themselves to a limit neither reach it (unbounded states) (de Miguel 1999: 3012).

States, at the same time, can be separated into two groups: individual-level and stage-level predicates. The difference between them depends on whether the possibility exists to establish spatial-temporal limits in the state. In that way, we see that while individual-level predicates characterize individuals independently of the spatial-temporal situation, stage-level predicates describe the state in which individuals are in a particular spatial-temporal situation (op. cit. 3012—3013).

Individual-level/stage-level distinction has been the object of a great number of studies, particularly applied to Spanish adjectival predicates. Demonte (1999: 142—144), for instance, distinguishes two classes of adjectives depending on the internal structure of the quality assigned by the adjective. In that sense, she talks about individual-level adjectives

such as *apto* ‘apt’, *idóneo* ‘suitable’ or *psicópata* ‘psychopathic’ and stage-level adjectives, where she includes qualifying adjectives derived from participles such as *harto* ‘full’, *contento* ‘happy’ or *descalzo* ‘bare-foot’ and all the adjectival participles (*abierto* ‘open’, *destrozado* ‘smashed’, *conmovido* ‘moved’). As some authors note (Demonte 1999: 142–143, Fernández Leborans 1999: 2366–2367, 2426–2428), a categorical proof in Spanish to differentiate among adjectives which refer to permanent properties from those which refer to temporal properties is their syntagmatic co-occurrence with the Spanish copula forms *ser* ‘to be’ and *estar* ‘to be’. Adjectives denoting individual-level states can combine only with *ser*; on the contrary, those denoting stage-level states are used with *estar*.

In the following sections we try to show the possibility to apply this distinction to stative nominal predicates denoting emotion in Spanish in a similar way to that made for adjectival predicates. The next I will begin presenting emotion nouns in Spanish.

3. Spanish emotion nouns: IC and EC emotion nouns

Spanish emotion nouns can be divided into two major groups. These groups represent two ways of conceptualizing emotional reality. The first group is integrated by the so called IC emotion nouns such as *respeto* ‘respect’, *amor* ‘love’ or *odio* ‘hate’. In this case, the emotion is conceived as having its origin in the experiencer him/herself as a result of an internal thought through which the experiencer judges an object from the outside world. The second group includes EC emotions nouns such as *asombro* ‘amazement’, *ira* ‘rage’ or *desesperación* ‘desperation’. With these nouns, for the emotion to start it is required an external fact that triggers the emotional reaction in the experiencer.

Evidence for these two groups is provided through many of their semantic, syntactic and lexical co-occurrence properties. From a semantic viewpoint, the second actant of IC nouns is an “object” towards which the emotion directs (*el respeto de Pedro hacia su profesor* ‘the respect of Pedro towards his teacher’); with EC nouns, the second actant is the “cause” which gives rise to the emotion (*el asombro de Pedro por la proeza* ‘Pedro’s amazement because of his feat’).

Syntactically, with IC nouns, the second actant can be expressed by means of the prepositions *a* ‘to’ or *hacia* ‘towards’, but never with *ante* ‘in the presence of’ (*el respeto de Pedro a <hacia, *ante> su profesor*); on the contrary, the second actant of EC nouns is expressed by *ante* and rejects *a* and *hacia* (*el asombro de Pedro ante <*a, *hacia> la proeza*).

Just to illustrate the behaviour of these nouns with regard to lexical co-occurrence, we offer here only one example related to the lexical function (LF) Adv₂. Only EC nouns present values for this LF: *ante el asombro del público, para su asombro* ‘to his/her amazement’.

4. A comparative approach between emotion nouns and other Spanish nominal predicates

As stated above, from the aspectual viewpoint, emotion nouns are states, that is to say, they are defined by their duration in time and by their unbounded nature. This means that when a state is taking place there is no way to fix its internal limits. However, sometimes, it is possible to set external limits. When this happens, we can establish the stage-level (with external limits)/individual-level (without external limits) contrast.

The main problem of this approach concerning emotion nouns, and in general every nominal predicate, is that only adjectives have the irrefutable evidence to be combined with *ser* and/or *estar* verbs. Added to this difficulty is the fact that the individual-level/stage-level nature of a state cannot be derived from the definition of the LU *sensu stricto*. When a definition contains the semantic component ‘state’, it is understood, invariably, that the LU is characterized by two properties: its duration in time and its unbounded nature — we have to take into account that these are the unmarked variants of the punctual/durative and bounded/unbounded oppositions. Both features bear implicitly the assumption that the state by itself does not have spatial-temporal limits. So, *el amor de Pedro por Ana* ‘Pedro’s love for Ana’ or *el asombro de Pedro ante la noticia* ‘Pedro’s amazement at the news’ denote emotional states that remain homogeneously within their duration.

Despite obstacles, it is undeniable that there are aspectual differences between emotion nouns such as *amor*, *respeto* or *odio* and nouns such as *desesperación*, *ira* or *asombro*. We explain these differences taking as a starting point three classes of nouns: events nouns, moral quality nouns and emotion nouns. Event nouns (Bosque 1999: 51–53, Picallo 1999: 368–371) such as *viaje* ‘journey’, *reunión* ‘meeting’ or *accidente* ‘accident’ are predicates with well defined boundaries. Bosque (ibid.) shows examples where these nouns appear in combination with verbs such as *tener lugar* ‘to take place’, *ocurrir* ‘to happen’ or *presenciar* ‘to witness’ — which imply a concrete spatial and temporal location —, and verbs such as *comenzar* ‘to start’ or *terminar* ‘to finish’ — which point out the beginning and the end of the event. Otherwise, moral quality

nouns such as *honestidad* ‘honesty’, *bondad* ‘goodness’ or *orgullo* ‘pride’ lack spatial-temporal boundaries but have intensity and degrees of intensity (Van de Velde 1999: 63). In that way, it is impossible to say **La bondad de Pedro tuvo lugar el jueves a las siete en el parque* ‘The goodness of Pedro took place on Thursday at seven in the park’, but the expression *Pedro tiene una bondad infinita* ‘Pedro has an infinite goodness’ is possible.

If we analyse emotion nouns following the same patterns, we find, as a first observation, that these nouns do not behave as a homogeneous class. On the one hand, all the emotion nouns, IC (1) as well as EC (2), can be intensified:

- (1) *Siento un respeto inmenso hacia su obra*
‘I feel a great respect for his work’
- (2) *con gran asombro del público*
‘with great amazement from the public’

On the other hand, only EC nouns combine with LUs which have external spatial-temporal boundaries. In this way, while expressions such as those in (3) are possible with EC nouns; they are ungrammatical with IC nouns (4):

- (3) a. *Mariana, en un momento de cólera, gritó desde el barandal.*
‘Mariana, in a moment of fury, shouted from the banister’
- b. *En su desesperación, no pudo ajustar bien el cinto.*
‘In his desperation, he could not tighten his belt properly.’
- (4) a. **En un momento de aversión gritó.*
‘In a moment of aversion, he shouted’
- b. **En su aprecio, le regaló uno de sus cuadros.*
‘As a sign of appreciation, he gave one of his pictures’

In (3a) and (3b), it is shown what a person is doing or not at the time in which he/she is feeling the emotion, but precisely the duration of the external time is bounded with *en* or the expression *en un momento de*.

In brief, IC nouns behave in a similar way as moral quality nouns. When a person is feeling love, or respect, the emotion is seen from the interior. In this sense, the emotion can be intensified but not located in a concrete space and time. EC nouns behave partly as quality nouns, partly as event nouns. When a person is feeling anguish or despair, these emotions are within the person, so they can be intensified; however, as far as it is possible to talk about the moment of the anguish or the desperation, it is also possible to fix limits, so the emotion can be seen from the exterior, as a whole.

5. Evidence for the aspectual delimitation of the Spanish emotions nouns

In this section we present some characteristics of emotion nouns used as proof for the individual-level/stage-level distinction.

5.1. Variation of emotion nouns in grammatical number

Between IC and EC emotion nouns, there are some differences regarding the grammatical number variation. In general, emotion nouns are mass predicates and for this reason, it should be expected that, at least, IC nouns did not admit plural number — plural is a property only for the entities located in space and time. It is true that some nouns, such as *remordimiento* ‘remorse’ or *esperanza* ‘hope’, can be used in the plural, but in these cases, they do not denote quantities, but the magnitude of the emotion (*¡Tenía tantas esperanzas de volver a verte!* ‘a great hope’). Other nouns such as *amor* or *desprecio* ‘contempt’, when they are used in the plural, they do not refer to the emotion, they take another sense: *amor* as ‘the object of the emotion’ (*Tuvo varios amores a lo largo de su vida* ‘several persons to whom someone felt the emotion’) and *desprecio* as ‘action showing the emotion’ (*Me hizo tantos desprecios que dejé de hablarle* ‘several signs of the emotion’). We are inclined to affirm that, in the rest of the cases, when IC nouns are used in the plural, they refer to the persons who experience the emotion (*Suscita muchos odios, Levanta muchas pasiones* ‘somebody makes several persons feel hate/passion towards him/her’).

However, when the EC nouns are in the plural, it seems that they refer to ‘periods of time during which the emotion is felt’ (5):

- (5) a. *Comprendía que todas las desesperaciones eran momentos de tensión transitoria.*
 ‘He understood that all the despair were moments of temporal tension’
 b. *Sus enfados son muy largos.*
 ‘His moments of anger are very long’
 c. *Entonces montó en cólera, una de aquellas cóleras espectaculares de Joaquín.*
 ‘Then he got angry, one of those Joaquín’s spectacular moments of anger’

There are, otherwise, some cases showing that EC nouns are discontinuous or countable. Let’s observe the contrast between sentences in (6) and (7):

- (6) a. *Ayer me dio dos sorpresas, pero hoy no volverá a darme otra.*
 ‘Yesterday he surprised me twice, but today he will not surprise me anymore’
 b. *Ayer me dio dos sobresaltos, pero hoy no volverá a darme otro.*
 ‘Yesterday he scared me twice, but today he will not scare me again’
 c. *Ayer me dio dos sustos, pero hoy no volverá a darme otro.*
 ‘Yesterday he frightened me twice, but today he will not frighten me again’
- (7) a. **Ayer me causó dos asombros, pero hoy no volverá a causarme otro.*
 ‘Yesterday he amazed me twice, but today he will not amaze me again’
 b. **Ayer me causó dos desesperaciones, pero hoy no volverá a causarme otra.*
 ‘Yesterday he caused me despair twice, but today he will not cause me to despair again’

In (6) and (7) we have used different EC emotion nouns in a context which allows us to test if a noun is countable or uncountable. The result shows that nouns such as *sorpresa* ‘surprise’, *sobresalto* ‘start’ and *susto* ‘fright’ are exceptions within the group of emotion nouns, exactly because they are countable. In our opinion, the discontinuous character of these nouns can be related to the intense brevity in which they appear and disappear. It seems that there has been established some kind of identity between the emotion itself — difficultly delimited — and the objective cause which triggers it — perfectly delimited; for instance, a telephone call late in the night, some shouts, in the case of *sobresalto* o *susto*; the unexpected arrival of somebody, for *sorpresa*.

5.2. LUs in paradigmatic co-occurrence with emotion nouns

In this section we comment briefly on the aspectual differences showed by the verbs and adjectives in paradigmatic co-occurrence with IC emotion nouns, on the one side, and with EC emotion nouns, on the other.

The sense of emotion nouns can be expressed by means of verbs. From IC nouns are derived transitive verbs; for instance, the values for the LF V_0 of *respeto* ‘respect’, *amor* ‘love’ and *odio* ‘hate’ are *respetar* ‘to respect’, *amar* ‘to love’ and *odiar* ‘to hate’, respectively. All these verbs are stative. From EC nouns, such as *asombro* ‘amazement’, *enfado* ‘anger’ and *desesperación* ‘desperation’, are derived two types of

verbs: one of them is constructed with *se* such as *asombrarse* ‘to be amazed’, *enfadarse* ‘to get angry’ and *desesperarse* ‘to despair’, and the other one is a conversive of the previous one: *asombrar* ‘to amaze’, *enfadar* ‘to anger’ and *desesperar* ‘to despair’. The former is an inchoative verb in which it is possible to distinguish the initial (*está a punto de enfadarse* ‘he is going to get angry’), medium (*se está enfadando* ‘he is getting angry’) and final (*está completamente enfadado* ‘he is totally angry’) phases. The latter refers aspectually to a change of state, that is to say, the transformation from not to be angry, for instance, till to be angry, but not the process that is experienced when one is already in this state. The aspectual differences between these verbs reflect, in a certain way, that EC nouns have some kind of limits.

The adjectives derived from emotion nouns present also aspectual differences. Adjectives in *-ado*, for example, are values for the LF A₂ with IC nouns. In this way, *respetado* ‘respected’, *amado* ‘loved’, *odiado* ‘hated’ refer to the object of the emotion. However, adjectives in *-ado* derived from EC nouns are values of A₁: *asombrado* ‘amazed’, *enfadado* ‘angry’, *desesperado* ‘desperate’ refer to the experiencer of the emotion. With regard to aspect, we can see that adjectives derived from IC nouns are combined with the copula form *ser* (*es respetado/amado* ‘he is respected/loved’); otherwise, adjectives derived from EC nouns occur with the copula form *estar* (*está asombrado/enfadado* ‘he is amazed/angry’).

5.3. Verbs in syntagmatic co-occurrence with emotion nouns

This section deals with the differences presented by IC and EC emotion nouns regarding the verbal values to LF Oper₁ and to the combination of Incep, Cont and Fin with Oper₁.

IC nouns select as more frequent values for Oper₁ the following verbs: *tener* ‘to have’, *sentir* ‘to feel’, *rendir* ‘to render’, *tributar* ‘to pay’, *profesar* ‘to profess’, *dispensar* ‘to confer’, etc. EC nouns combine with *tener* ‘to have’, *sentir* ‘to feel’, *sufrir* ‘to suffer’, *experimentar* ‘to experience’, *llevarse* ‘to take’, *pasar* ‘to pass’, *padecer* ‘to suffer’, etc.

It is worthwhile to make some observations regarding *tener*, because the co-occurrence of this verb with the IC and EC nouns present significant differences. When *tener* appears with IC nouns, the emotion noun, which is a DO, can be used without determinants or modifiers. The semantic actant Y (object of the emotion) has to be expressed as a syntactic dependent of the verb (IO), normally in the form of a pronoun (*le/les* ‘him, her/them’) or redoubled: with a pronoun and a prepositional phrase (8).

- (8) a. *Le tenía cariño a su vecino.*
 ‘He felt affection for his neighbour’

When *tener* is used with EC nouns, the noun, with the role of DO, is accompanied necessarily by a determinant and a modifier (intensifier). It is not compulsory the syntactic expression of the semantic actant Y (cause of the emotion); if it is expressed, it depends always on the noun, never on the verb (9).

- (9) *Tuvo una desesperación terrible por la muerte de su marido.*
 ‘She felt a terrible desperation because of her husband’s death’

We present now a list of the phasal verbs in co-occurrence with IC and EC emotion nouns³:

- (10) IncepOper1:
 a. IC nouns: *tomar, coger, cobrar* [~ a N];
 b. EC nouns: *caer* [en ART ~], *montar* [en ~];
- (11) ContOper1:
 a. IC nouns: *conservar, guardar, mantener* [ART ~];
 b. EC nouns: *no salir* [de ART ~]; *seguir* [con ART ~]
- (12) FinOper1:
 a. IC nouns: *perder* [ART ~]
 b. EC nouns: *salir* [de ART ~]

Considering what has been exposed, it is possible to make some observations: verbs selected by IC nouns, when they are used as full verbs, express ownership, possession or transfer. We include among them *tener* ‘to have’, *deber* ‘to owe’, *tomar* ‘to take’, *coger* ‘to catch’, *cobrar* ‘to earn’, *conservar* ‘to preserve’, *guardar* ‘to keep’ o *perder* ‘to lose’ (13):

- (13) a. *Le tiene odio.*
 ‘He hates him’
 b. *Le cogió (cobró, tomó) cariño.*
 ‘He became fond of him’
 c. *Conserva su rencor.*
 ‘He bears a grudge against him’

Otherwise, verbs combined with EC nouns express, in their full sense, spatial or/and temporal location, for example, *caer* ‘to fall down’, *montar* ‘to get on’, *pasar* ‘to pass’, *salir* ‘to come out’, *entrar* ‘to come in’ o *hundir* ‘to sink’ (14):

³ In the appendix, we offer a more complete list of LFs and their values for IC/EC emotion nouns.

- (14) a. *Cayó en una desesperación.*
 ‘He despaired’
 b. *Montó en cólera.*
 ‘He got angry’
 c. *No sale de su asombro.*
 ‘He continues in amazement’

If we compare these observations with the use of Spanish verbs *ser* and *estar*, we verify that precisely *ser*, the copula form which occurs with individual-level adjectives, is the one used to express ‘possession’ (*Ese libro es mío* ‘That book is mine’). On the contrary, one of the distinctive values of *estar*, the verb combined with stage-level adjectives, is ‘to occupy a place in the space, to stay or live in a place’; in other words, *estar* means spatial location (*Está en casa* ‘She is at home’). This is another piece of evidence for the claim that IC nouns behave as individual-level predicates and EC nouns as stage-level predicates.

6. Conclusions

The aim of this paper was to apply the known aspectual difference between individual-level/stage-level predicates to a group of Spanish emotion nouns. We have presented two groups of Spanish emotions nouns, IC and EC nouns, and we have provided evidence for the claim that IC nouns behave as individual-level predicates and EC nouns, as stage-level predicates.

As a further development, this typology could be integrated into an explanatory and combinatorial dictionary in order not to repeat in the particular entries of the nouns all the information shared by each one of the groups. In this sense, the shared information for IC and EC nouns could be encoded in a generic entry for the emotion’s semantic field.

References

- Alonso Ramos, Margarita (2002): “Hacia un Diccionario de colocaciones del español y su codificación”. In: M. A. Martí (ed.), *Estudios sobre el léxico*, Barcelona: Edicions de la Universitat de Barcelona, 2002.
- Anscombe, Jean-Claude (1995): “Morphologie et représentation événementielle: le cas des noms de sentiment et d’attitude”. *Langue française* 105. 1995. P. 40—53.

- Bennett et al. (1990): "Toward a Computational Model of Aspect and Verb Semantics", *Machine Translation* 4, 1990. P. 217—250.
- Bosque, Ignacio. 1999. "El nombre común". In: I. Bosque/V. Demonte (drs.). P. 3—75.
- Bosque, Ignacio / Violeta Demonte (drs.) (1999): *Gramática descriptiva de la lengua española. Vols. 1—3*. Madrid: Espasa Calpe, 1999.
- Carlson, Gregory N. / Francis J. Pelletier (eds.) (1995): *The Generic Book*. Chicago: University of Chicago Press, 1995.
- Demonte, Violeta (1999): "El adjetivo: Clases y usos. La posición del adjetivo en el sintagma nominal". In: I. Bosque / V. Demonte (drs.), 1999. P. 129—215.
- Fernández Leborans, M^a Jesús (1999): "La predicación: Las oraciones copulativas". In: I. Bosque/V. Demonte (drs.). P. 2357—2460.
- Flaux, Nelly/D. Van de Velde (2000): *Les noms en français: Esquisse de classement*. Paris: Ophrys, 2000.
- Luján, M. (1981): "The Spanish copulas as aspectual indicators". *Lingua* 54, 1981. P. 165—209.
- Maienborn, C. (2005): "A Discourse-Based Account of Spanish *ser/estar*". *Linguistics* 43: 1, 2005. P. 155—180.
- Mathieu, Yvette Y. (1999): "Les prédicats de sentiment". *Langages* 136, 1999. P. 41—52.
- Miguel, Elena de (1999): "El aspecto léxico". In: I. Bosque / V. Demonte (drs.), 1999. P. 2977—3060.
- Mel'čuk, Igor (1997): *Vers une linguistique Sens-Texte. Leçon Inaugurale*. Paris: Collège de France, 1997.
- Mel'čuk, Igor et al. (1984—1997): *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques I—IV*. Montréal: Les Presses de l'Université de Montréal, 1984—1997.
- Mel'čuk, Igor / André Clas / Alain Polguère (1995): *Introduction a la lexicologie explicative et combinatoire*. Louvain-la-Neuve: Duculot, 1995.
- Picallo, M. Carme (1999): "La estructura del sintagma nominal: las nominalizaciones y otros sustantivos con complementos argumentales". In: I. Bosque / V. Demonte (drs.), 1999. P. 363—393.
- Pustejovsky, James (1995): *The Generative Lexicon*. Cambridge: MIT Press, 1995.
- Sanromán Vilas, Begoña (2003): *Semántica, sintaxis y combinatoria léxica de los nombres de emoción en español*. Helsinki: Yliopistopaino [doctoral dissertation], 2003.
- Van de Velde, Danièle (1995): "Les verbs dits "psychologiques" revus à la lumière des noms correspondants". *Revue de Linguistique Romane* 59, 1995. P. 67—97.
- Van de Velde, Danièle (1999): "La multiplication des sentiments". *Travaux de linguistique* 38, 1999. P. 63—87.
- Vendler, Zeno (1967): *Linguistics in Philosophy*. Ithaca: Cornell University Press, 1967.

APPENDIX: Some of the most regular LFs
with Spanish emotion nouns

IC emotion nouns (individual-level predicates)		
AFECTO, AGRADECIMIENTO, AMOR, ANTIPATÍA, APEGO, APRECIO, AVERSIÓN, CARÍÑO, DESDÉN, DESPRECIO, DEVOCIÓN, ESTIMA, HOSTILIDAD, ODIO, OJERIZA, RENCOR, RESPETO, TIRRIA, SIMPATÍA, VENERACIÓN, ...		
LFs	'GLOSS' OR "METALINGUISTIC EXPLANATION"	values
Gener	"generic term"	sentimiento, actitud [de ~]
V₀	"derived verb"	amar, apreciar, estimar, despreciar, desdenar, odiar, etc.
S₁	'the one who feels the E(motion)'	amante2, adorador2, admirador2, etc.
S₂	'E's object'	amor1.1b, afición1b.
S0Caus1Manif	'E's manifestation'	cariño1b, desdén2, desprecio2, afecto2b, etc.
A₁	'feeling the E'	admirador1, amante1
A₂	'being the E's object'	amado, odiado, aficionado, etc.
Able₂	'prone to cause the E'	odioso, apreciable, cariñoso, deseable.
Adv₁	'X feeling the E'	con [~], afectuosamente, etc
Propt	'because of'	por [~], de [~] (menos frecuente).
Oper₁	'X feels the E'	sentir, tener, rendir, tributar, profesar, deber, dispensar [~ a/por N]
IncepOper₁	'X begins to feel the E'	tomar, coger, cobrar [~ a N]
ContOper₁	'X continues feeling the E'	conservar, guardar, mantener [(ART) ~]
FinOper₁	'X stops feeling the E'	perder [ART ~]
Oper₂	'Y is the object of X's E'	gozar [de ART ~], tener [ART ~]
Labor₁₂	'in his behaviour with Y, X feels the E'	tratar [(a) N con ~]
CausOper₁	'Y or W makes X feel the E'	llenar [a N de ~]
CausFunc	'Y o W makes the E be in X'	causar, provocar, producir, suscitar [~(a N)]

EC emotion nouns (stage-level predicates)		
ABATIMIENTO, ANGUSTIA, ASOMBRO, CHASCO, CÓLERA, DECEPCIÓN, DESESPERACIÓN, DESPECHO, DISGUSTO, ENFADO, EXASPERACIÓN, FASTIDIO, FURIA, ILUSIÓN, INDIGNACIÓN, IRA, IRRITACIÓN, PREOCUPACIÓN, RABIA, REGOCIJO, SORPRESA,...		
LFs	'GLOSS' OR "METALINGUISTIC EXPLANATION"	values
Gener	"generic term"	estado [de ~]
V₀	"verb with <i>se</i> "	enfadarse, asustarse, desesperarse, irritarse, indignarse.
Conv₂₁(V₀)	"derived verb with inversed actants"	enfadar, asustar, irritar, desesperar, indignar.
S₂	'E's cause'	admiración 2b , chasco 1b , sorpresa 1b , decepción 1b , etc.
S0Caus1Manif	'E's manifestation'	gesto, grito, señal, muestra [de ~], etc.
A₁	'feeling the E'	abatido, desesperado.
A₂	'causing the E'	desesperante, irritante.
Able₁	'inclined to feel the E'	asombradizo, irascible, irritable.
Adv₂	'Y, being the E's cause for X'	para [~ de N/A _{pos} ~], ante [ART~ de N/A _{pos} ~]
Propt	'because of'	de [~], por [~]
Oper₁	'X feels the E'	tener, sentir, sufrir, experimentar, llevarse, pasar, padecer [~]
IncepOper₁	'X begins to feel the E'	caer, montar [en (ART/Sing) ~]
ContOper₁	'X continues feeling the E'	seguir [con ~], no salir [de ART ~]
FinOper₁	'X stops feeling the E'	salir, recuperarse [de ART ~]
Oper₂	'Y causes the E in X'	dar, producir, causar, provocar, suscitar, etc.
Labor₂₁	'Y makes X feel the E'	llenar [a N de ~], empujar, llevar [a N a ART ~], hundir, sumir [a N en ART ~]
Perm₁Fact₀	'X lets the E direct his way of acting'	abandonarse [a ART ~], dejarse llevar (arrastrar) por [ART~]
IncepFunc₁	'the E starts to be in X'	apoderarse [de N], entrar [a N]

ContFunc₁	'the E continues being in X'	durar [a N]
FinFunc₁	'the E stops being in X'	pasarse(1e) [a N], abandonar [a N].
Adv₁Able₂	'in such a way that causes the E'	sorprendentemente, asombrosamente.

Object or Situation: the Factors of Expressing the Arguments of Mental Verbs in Russian

A. B. Letuchiy,

Russian State University for Humanities
alexander_letuchiy@hotmail.com

N. V. Serdobolskaya

Moscow State University
natasha_serd@yahoo.com

The paper examines mental, emotive and perception verbs that can take both noun and clausal complements (e.g. *Vas'a uvidel Pet'u* 'Vas'a saw Peter' and *Vas'a uvidel, čto Pet'a risujet* 'Vas'a saw Peter drawing' lit. 'that Peter draws'). We examine the pragmatic factors influencing the speaker's choice between the noun and clausal complement. It is shown that the same factors are relevant for both perception and emotive verbs, despite the difference in argument structure.

Между объектом и ситуацией: способы выражения актантов ментальных глаголов в русском языке. В фокусе внимания настоящего исследования находятся ментальные, эмотивные глаголы и глаголы восприятия, вторая валентность которых может заполняться именем или ситуацией (напр., *Вася увидел Петю* и *Вася увидел, что Петя рисует*). Мы рассматриваем факторы прагматического порядка, регулирующие выбор говорящим именного или сентенциального актанта. В работе показано, что несмотря на различие в составе семантических ролей, одни и те же факторы оказываются релевантными для глаголов восприятия и эмотивных глаголов.

0. The goal of the paper

This paper examines Russian mental, emotive and perception verbs that can take both noun and clausal complements ¹, e.g. the verb *uvidet'* 'see' in *Vas'a uvidel Pet'u* 'Vas'a saw Peter' and *Vas'a uvidel, čto Pet'a risujet* 'Vas'a saw Peter drawing something' lit. 'that Peter draws'. In

¹ Perception verbs (*u*)*videt'* (see), (*u*)*slyšat'* (hear), (*po*)*čuvstvovat'* (feel), emotive verbs (*ras*)*serdit'sja* (be/become angry), *rasstroit'sja* (be defeated, troubled), *udivit'sja* (be surprised), (*razo*)*zlit'sja* (be/become mad), (*ob*)*radovat'sja* (be glad), (*po*)*nравit'sja* (like), *zavidovat'* (envy), and mental verbs (*vs*)*pomnit'* (remember) and (*u*)*znat'* (know, get to know).

these sentences the noun *Pet'a* and the dependent clause *čto Pet'a risujet* occupy the same valency slot of the verb *uvidet'* and encode the semantic role of Stimulus. We focus on the properties of the Stimulus that are relevant for the speaker when choosing, whether the object or the situation is to be expressed. To achieve this goal, we first find out morphological and syntactic devices, which are used to encode the Stimulus of the verbs in question; then we identify semantic and pragmatic factors determining the choice of the construction encoding the Stimulus.

It is well-known that in Russian one and the same emotive predicate can have different number of syntactic valency slots in the same lexical meaning (A. Zaliznjak 1992: 416). For instance, the verb *serditsja* 'to be angry' can have the argument structure $\langle X \text{ serditsja na } Y \rangle$ 'X is angry at Y' and $\langle X \text{ serditsja na } Y \text{ za } Z \rangle$ 'X is angry at Y for Z', where Y denotes the Stimulus, and Z denotes the Cause:

- (1) *Papa serditsja na Pet'u. Papa serditsja na Pet'u za to, čto on razbil časy.*
'Dad is mad at Peter. Dad is mad at Peter, because he has broken his watch.'

It is assumed that *serditsja* is a three-place predicate, taking the Experiencer, Stimulus, and Cause participants (see the article on *serdit'sja* in Apresjan (1999)), where Stimulus and Cause may be unexpressed (e.g. *Papa serditsja* 'Dad is angry'). Taking this argument structure for *serdit'sja* we could assume that all the three participants exist independently, i.e. denote different entities in the real world. However, this presumption is evidently wrong for the Stimulus and Cause participants. The verb *serditsja* denotes the emotional reaction of the Experiencer at the situation (Cause), where the Stimulus is a participant of the Cause situation (most often, the subject of the dependent clause encoding the Cause). This intuitive reading is justified by the possibility of the construction, where Cause is encoded as a dependent clause, where the Stimulus is the subject:

- (2) *Papa serditsja, čto Pet'a razbil časy.*
Dad is angry, because Peter has broken the watch. (lit. that Peter...)

The verb *serdit'sja* has a much wider range of possibilities, concerning Stimulus and Cause encoding, cf. (we give below the variants elicited from the informants):

- (3) *Papa serditsja, potomu čto kot xuliganil vs'u noč.*
Dad is angry, because the cat has been playing tricks all night long.

- (4) *Papa serditsja iz-za tvoix kaprizov.*
Dad is angry because of your whims.
- (5) *Papa serditsja iz-za razbitoj vazy.*
Dad is angry because of the broken vase.
- (6) #² *P'otr Ivanovič serditsja na radikulit.*
P'otr Ivanovich is angry because of the sciatica (lit. at the sciatica).
- (7) # *P'otr Ivanovič serditsja na ne sdelannoje domašneje zadanije.*
P'otr Ivanovich is angry because of the homework (lit. at the uncompleted homework).

In (3) Cause is encoded by an adverbial clause with *potomu čto* 'because'). The Stimulus (*kot* 'the cat') is not an argument of *serditsja*, as it fills the subject position in the dependent clause. (4) seems to resemble (3) in that Cause does not fill any valency slot of the matrix verb, and is expressed through a prepositional phrase; nor does the Stimulus, expressed as the underlying subject of the verbal noun. In both (3) and (4) the NP, most likely to occupy the Stimulus valency slot of the matrix verb, appears as the subject of the dependent clause. In (5) Cause (*to, čto kto-to razbil vazy* 'the fact that someone has broken the vase') does not occupy an argument position in the matrix clause, nor an adverbial position, but is expressed as an attribute of the noun *vaza* 'vase'. The sentences (6) and (7) are of interest, because the NP taking the preposition seems to refer at the same time both to Stimulus and Cause (see Volf (1989) for such cases). I.e. the referent of the noun phrase is "a mix" of the two participants, which would be impossible if they would be conceived as independent entities. Therefore, the presumption that Stimulus and Cause denote different entities in the real world, seems to be untenable. Intuitively, they are parts of one and the same entity. The question arises, why they are often realized as different syntactic units. This question is to be answered below.

What can be observed on the syntactic level, is the striking diversity of possible alternations in argument representation, that can be observed not only with the verb *serdit'sja*, but, as shown below, with most of the perception, mental and emotive verbs in Russian. Such a diversity arises problems for any formal theory of language. As far as MTT is concerned, the question arises, whether the constructions in (1)-(7) can be listed in

² This symbol (#) marks the examples that have been elicited from the informants, but do not seem to be quite acceptable as to our judgement.

one and the same lexical unit, or should correspond to the different lexical meanings of the matrix verb in question. On the one hand, *serdit'sja* in (1)-(7) is believed to have one and the same lexical meaning (see Apresjan (1999)). On the other hand, the sentences (1)-(7) differ not only in the overt marking of Stimulus and Cause, but also in their position in the argument structure of *serdit'sja*: in (2) Cause appears as a complement of *serdit'sja*, in (3) rather as an adverbial (see the discussion of the argument structure of emotive verbs in A. Zalznjak (1992)).

We explain the observed diversity by the factors lying outside the idiosyncratic semantic properties of the matrix verb. We argue that it resides in the semantic and pragmatic factors discussed below. To verify our point, we 1) find out the morphological and syntactic devices that are used to encode the non-Experiencer arguments of the listed verbs, and 2) check the relevance of the above-mentioned pragmatic factors in the series of psycholinguistic experiments.

To avoid any misunderstanding arising from difference in argument structure of perception and emotive verbs we do not use the notions Stimulus, Cause, or Theme. For the sake of simplicity, we use the notions SST (situation as Stimulus/Cause/...) and OST (object as Stimulus/Cause/...), where SST denotes the clausal complement (Stimulus or Cause) of the perception/emotion verbs (e.g. *Pet'a razbil časy* 'Peter has broken the vase' in (2), *Vaza razbita/razbilas'* 'The vase is broken / broke' in (5)), and OST denotes the object, stimulating the Experiencer's emotional reaction by its actions, properties, location in space etc. (*Pet'a* in (1)–(2), *kot* 'the cat' in (3), *ty* 'you' in (4), *vaza* 'vase' in (5)).

As we aim at analyzing the colloquial speech rather than the literary language, some examples offer constructions, "unusual" or even unacceptable in literary Russian.

1. Morphological and syntactic constructions used to express OBJECT (OST) and SITUATION (SST)

1.1. Primarily, our purpose was to list all the possible constructions appearing with the verb in question. To achieve this goal we studied the data from the National Corpus of Russian Language. To estimate the approximate frequency of these constructions in colloquial speech we undertook an experiment where the informants were to complete sentences (as *Vas'a rasserdils'a*, *Ty znaješ...?*). This way we endeavoured to elicit the constructions, most frequent for colloquial speech. The most interesting peculiarities of the constructions elicited are given below.

The list of elicited constructions is given below:

- OST alone is expressed (*rasserdils'a na Vas'u* 'got angry at Vas'a');
- SST alone is expressed:
 - a) as a clausal complement (*udivils'a, čto Vas'a ne prišol* '(he) was surprised that Vas'a had not come');
 - b) as an adverbial clause with a temporal, causal or conditional conjunction (*rasserdils'a, kogda mne na nogu uronili batareju* '(I) got angry when somebody let a radiator drop on my leg')
 - c) as a converb (*razozlils'a, uslyšav ob etom* '(he) got mad after having heard about it');
- Both OST and SST are expressed:
 - a) OST is expressed as a noun complement, SST — as an action/state/property nominal (*rasserdils'a na Vas'u za jego povedenje* 'got angry at Vas'a for his behaviour');
 - b) OST is expressed as a genitive noun or a possessive pronoun modifying an action/state/property nominal, which refers to the SST (*uslyšal voj sobaki* '(he) has heard a dog's howling', *udivilsja Vasinomu povedeniju* '(he) was surprised at Vas'a's behaviour');
 - c) SST is expressed as an adverbial clause with a temporal, causal or conditional conjunction, and OST — as an NP with a preposition (*Otec rasserdilsja na Pet'u, tak kak Pet'a polučil dvojku* 'Father got angry at Peter, because he had got a bad mark').

It is significant that (both in this and other experiment series) the most frequently elicited construction is the one when the verb has a noun complement only. The exact frequency of different constructions depends on the matrix verb; still, for all the verbs analyzed, the noun complement construction is much more frequent, than the clausal complement one. Even if a situation is conceived as a Stimulus of emotion or perception, often it is the agent alone which is expressed (see (13)–(15) below).

At this primary stage of investigation, the following results seem to be of interest. A range of constructions have been elicited, uncommon or even unacceptable in literary Russian:

- "Unusual" arguments' marking (e.g. instrumental case on SST with the verb *udivilsja* 'to be surprised'):
 - (8) # *Ja udivilsja svojim postupkom.*
I was surprised by my (own) action.
- "Unusual" syntactic constructions: a number of matrix verbs can have both a noun and clausal complement occupying one and the same syntactic position:

- (9) # *On uvidel Kuz'ku, čtó on tvorit / čto on š'ekočet čeloveku p'atki.*
He saw Kuz'ka doing something (lit. He saw Kuz'ka what he did) / He saw Kuz'ka tickling the man's heels (lit. He saw Kuz'ka that he was tickling the man's heels).
- (10) # *Nafan'a rasserdilsja na Kuz'ku, čto on xuliganit.*
Nafan'a got angry at Kuz'ka for playing tricks (lit. at Kuz'ka that he was playing tricks).

1.2. Most frequent is the construction where the SST remains overtly unexpressed:

- (11) *Otec rasserdilsja na t'ot'u. Žena obradovalas' mužu.*
The father got angry at the aunty. The woman was glad at seeing her husband (lit. at her husband).

Most frequently, the OST must be overtly expressed; with some verbs, as *rasserditsja*, its omission leads to unacceptability:

- (12) [?] *Papa rasserdilsja na/za dvojki / [?] Papa rasserdilsja iz-za dvojki.*
lit. Dad got angry at the bad marks / Dad got angry because of the bad marks.

Such a construction is possible, though not frequent, with other verbs:

- (13) *Ja obradovalsja uspexu.*
I was glad to have such a success (lit. at the success).

The SST can be expressed as a prepositional phrase or a relative clause, modifying the OST:

- (14) *Žena obradovalas' buketu cvetov ot muža.*
The woman was glad that her husband gave her a bunch of flowers (lit. glad at the bunch of flowers from her husband).
- (15) *Žena obradovalas' cvetam, kotoryje podaril jej muž.*
The woman was glad that her husband gave her flowers (lit. glad at the flowers that her husband gave her).

As a rule, the frequency of the construction where SST is expressed as a clausal complement (and OST as a subject of SST clause, see (16), (17)) is unexpectedly low (less than 40 %).

- (16) *Otec rasserdilsja, tak kak Pet'a polučil dvojku.*
Father got angry, because Peter had got a bad mark.
- (17) *Žena obradovalas' vozvraš'eniju muža. Žena obradovalas', čto jeje muž polučil povyš'enije.*
The woman was glad at her husband's coming back home. The woman was glad that her husband was given a promotion.

Thus, any of the matrix verbs in question allow a number of argument structure alternations in the same lexical meaning. We argue that the choice between the alternative argument structures is regulated by the semantic and pragmatic properties of OST and SST.

What seems to be an interesting result is that the frequency of the construction where OST and SST are encoded separately (i.e. occupying independent syntactic positions) is unexpectedly low, though it is grammatically correct. If both OST and SST are expressed separately, most often only one of them fills a valency slot of the matrix verb. If the matrix verb assigns the case marking to the SST (17), then OST is realized as subject in the complement clause. If the matrix verb assigns the case marking to the OST, SST is most often encoded as a prepositional phrase, adverbial, or relative clause. In 2.2 we show that this peculiarity is due to the pragmatic factors discussed below.

1.3. A rather frequent construction (as far as emotive verbs are concerned) is the one where SST is encoded as a causal adverbial clause headed by conjunctions *tak kak* and *potomu što*, temporal adverbial clauses with *kogda* and adjunct clauses. The problem is, that in this case, argument status of the Cause participant can be disputable. On the one hand, it is encoded by the syntactic devices usually marking the cause, time e.a. adverbials (cf. *Vas'a ne pošol v školu, potomu što zabolet* 'Vas'a didn't go to school, because he was ill'). On the other hand, the Cause of emotion (as in *Papa rasserdilsja na Vas'u, potomu što tot rugajetsa ploximi slovami; On udivilsja, uvidev strašiliš'e*) belongs to the definition of the emotive verb in question (see the definition of *serdit'sja* in (Apresjan 1999)), and therefore fills the valency slot of such a verb. That is, Cause of emotion can take an "intermediate" position between an argument and an adverbial (see Xrakovskij 1999 about the scalarity of arguments/ adverbials distinction. The question is, whether Cause, so encoded, is a complement or an adverbial in the matrix clause.

Following the point argued in (A. Zaliznjak 1992), we distinguish between the following cases: *On ispugalsja, potomu što uvidel volka* 'He was frightened, because he had seen a wolf', where introduces the direct Cause of emotion and *On usrugalsja, potomu što on voobšč'e truslivyj* 'He was frightened, because he is faint-hearted', where it introduces a more distant Cause. "Sledujet... različat' pričinu kak rol' odnogo iz učastnikov situacii, opisyvajemoj predikatom vnutrennego sostojanija, i pričinu kak osnovanije dl'a ocenki... kotoraja ne javl'ajetsa učastnikom situacii"³ (A. Zaliznjak 1992: 416). We consider the constructions of the first sort only.

³ It is important to distinguish Cause as an obligatory participant in the situation introduced by inner-state predicates, and Cause as evaluation base... that is not an obligatory participant of the situation introduced by inner-state predicates.

2. Semantic and pragmatic factors determining the choice of OST and SST encoding

2.1. The second experiment aims at identifying the most important factors, that influence the choice of constructions with the mental verbs in texts. The experiment was organized as follows: the informants were to fill gaps in the texts that described four two minute-fragments taken from the cartoons (“Prikl’učenija domovenka” (“Little bogy’s adventures”) and “Sledstviye vedut kolobki” (“Kolobki hold an investigation”)). The texts were compiled so as to elicit the constructions with verbs in question as heads.

The experiment has shown that the following factors can regulate the choice of the diatheses of the verbs under consideration:

- 1) Pragmatic properties of the situation (for instance, aforementionedness).
- 2) Properties of the OST: salience (protagonist/non-protagonist), animacy.
- 3) Semantic and syntactic properties of the main predicate; idiosyncratic lexical features.
- 4) Markedness of the SST.
- 5) Dynamicity of the SST.

The most important factor is the aforementionedness: for designating aforementioned situations and objects informants use other means, than for the new ones. For the SST the pronoun *eto* ‘this’ and abstract nouns like *povedenije* ‘behaviour’, *proisxod’aščeye* ‘the things which happen’ are used; the OST is encoded by the anaphoric pronouns. The factor under consideration is of interest, because usually the SST and the OST are both aforementioned or both new. If the situation has been mentioned in the previous text, it increases the probability that the OST alone (not the SST) would be expressed. On the contrary, new situations, which have not been mentioned, tend to be expressed as a whole. For example, according to our data, the verb *ispugat’sja* ‘fear’ chooses the sentential strategy in 80 % of cases and *dogadat’sja* ‘guess’ in 100 %, although the construction with a noun would be grammatically correct — cf. *dogadat’sja o proisšedšem* ‘guess what happened’. Presumably it is motivated by the fact that the Stimulus of these verbs is most often new. New situations tend not to have any salient participants, that could be used as OST instead of the name of the situation.

The pragmatic status of the participant and its animacy is also significant for the choice of a diathesis. If the participant, that appears in the fragment of the cartoon, is animate, especially if it is the protagonist of the situation, it tends to occupy the argument position, hence, the infor-

mants choose the OST construction (cf. the sentence *Vorona uvidela (Babu-Jagu i kota)* ‘The crow saw (Baba-Jaga and the cat)’, where the OST construction was used by 53 % of informants). If the participant is inanimate and is not a protagonist, SST constructions tend to be chosen.

Properties of participants are closely related to another parameter — properties of the matrix verb. It is a well-expected result that the semantics of the verb in many cases determines the choice of the diathesis: for example, the OST of the verb *rasserdit'sja* is expressed much more often than the SST, and with the verbs like *radovat'sja* ‘be glad’ and *udivit'sja* ‘be surprised’ the OST is rarely expressed in a separate verbal or nominal phrase (cf. *Vas'a obradovalsja, čto Pet'a prišel* ‘Vas'a was glad, that Pet'a came’, where the OST *Pet'a* is expressed inside the SST phrase *Pet'a prišel* ‘Pet'a came’ vs. *On obradovalsja za tebjā* ‘He was glad for you’, see also Apresjan 1999). However, the choice from all possible diatheses is influenced to a certain degree by pragmatic factors. The first and second experiments give some interesting data, proving that even semantically close predicates can choose different constructions: for example, probability of the sentential strategy with the verb *slyšat* ‘hear’ is greater than with *videt* ‘see’ and it is even greater with *čuvstvovat* ‘feel’.

Finally, the markedness and dynamicity of the described situation is relevant. Let us compare, for example, the frequency of sentential and nominal strategies for two sentences with the verb *videt* ‘see’: the first of them describes the situation ‘Nafan'a saw that Kuz'a tickles the man's feet’ — the SST is a dynamic situation. Only 36% of informants filled the gap with OST constructions (like *On uvidel Kuz'ku* ‘He saw Kuz'ka’), whereas frequency of the nominal strategy in the whole sample is 57 %. The second sentence describes the situation ‘Kuz'ka saw (that there is) a pot on the table’, i. e. a state. In this case 96 % of informants chose the nominal strategy (*On uvidel kastr'ul'u* ‘He saw a pot’).

2.2. The purpose of the last experiment was to define the relational degree of importance of different factors. The informants were to describe photographs, using constructions of the type “I see...”. In this experiment only the predicate *videt* ‘see’, which is especially important for our purposes, has been analyzed.

The photographs, used for this experiment, differed on the basis of those characteristic properties discovered in the second experiment. Only the factor “the number of participants” has been added:

I. Properties of the situation as a whole:

1) Dynamic/static. Even though photographs are always in a sense static, they let us define, whether the situation is dynamic (for example, the photograph representing a football match) or static (people, lying on a sofa).

2) Marked/unmarked situation. Unmarked situations are those, which are usual for everyday life, marked are unusual (for example, the situation of meditation is not usual).

II. Properties of the participants of the situation.

1) Quantity of participants (one/two/more than two).

2) “Unmarked”/“marked” participants. The participants can be “marked” as to their appearance (e.g. have “unusual” clothes, haircut etc.).

3) Known/unknown (identifiable/non-identifiable by the informant) participant. (The value of this feature “unknown” was substituted by “known” if the answer showed that the informant identified a participant, that we first considered as non-identifiable, and vice versa.)

In the course of the experiment the following ways of expressing the second argument of the verb *videt* have been found:

1) Subordinate clause (clausal argument): *Ja vizhu, čto l’udi lov’at rybu* lit. ‘I see that people are fishing’;

2) Nominal object: *Ja vižu Valerija Leont’jeva* ‘I see Valery Leontiev’;

3) Verbal noun (name of a situation): *Ja vižu igru v futbol* ‘I see a football match’;

4) Combination of a nominal object and a participle: *Ja vižu čeloveka, igrajuščego v m’ač* ‘I see a person playing with a ball’;

5) Combination of a nominal object and a relative clause: *Ja vižu l’udej, kotoryje tancujut* ‘I see people dancing’ (lit. people, who are dancing);

6) Combination of a nominal object and a prepositional phrase: *Ja vižu l’udej na pl’aže* ‘I see people on the beach’.

It has been argued for the verbs of perception that it is basically a situation (not its participants) that is perceived (cf. Kirsner, Thompson 1976). On the syntactic level the perceived situation can be realized as a dependent clause or a NP. If Stimulus is expressed as a noun, such construction is considered as deficient realization of the dependent clause (for example, *a cat* in the sentence *He saw a cat* is a reduced realization of the clause *A cat was walking/liing/eating* etc.). In terms of [Gak 1976], the speaker chooses “reductive nominalization” — a metonymical transfer, like in *He cannot go to the cinema because of his brother* (i.e. *because of his brother’s actions / his brother’s sickness*). In other words, the nuclear pattern for verbs of perception is supposed to be ⟨Nominative, S⟩, where S is an embedded clause. Adopting this approach, we must suppose that the construction with a clausal complement expressing the Stimulus of the verb *videt* ‘see’ should be the most frequent. But this is not the case. Such a pattern has a relatively low frequency:

Table 1. Frequency of different constructions
(with the verb *videt'* 'see')

Total ⁴	Embedded clause	Verbal noun	Noun	Noun + participle	Noun + rel. clause	Noun + PP
356	45 (17%)	31 (12%)	141 (54%)	14 (5%)	17 (7%)	12 (5%)

To sum up, the construction where the perceived situation does not occupy a syntactic position as an argument of the verb *videt'*, but is overtly expressed (by a participial phrase, a relative clause or a prepositional phrase) was chosen in 20 % of all examples only. The most frequent strategy of expressing the Stimulus of *videt'* is the nominal one.

The frequency of the clausal strategy increases (and the frequency of the nominal strategy decreases) if the perceived situation is highly marked (18) or dynamic (19):

(18) *Otec, mat' i doč' stranno sid'at.*

'The father, the mother and the daughter are sitting in a strange way' (the photograph represents a scene of meditation).

(19) *Čelovek igrajet na pianino, koška mešajet.*

'Somebody is playing piano, and the cat is interfering'.

Tables 2 and 3 represent how the dynamicity of the situation and its markedness trigger the choice of the constructions:

Table 2. Dynamicity of the situation (SST)

	Total	Embedded clause	Noun	Verbal noun	Noun + participle	Noun + rel. clause	Noun + PP
Dynamic	144	21 (15%) ⁵	59 (41%)	11 (8%)	7 (5%)	6 (4%)	5 (3%)
Static	210	23—25 (12%)	82 (39%)	20 (10%)	7 (3%)	11 (3%)	7 (3%)

Table 3. Markedness of the situation (SST)

	Total	Embedded clause	Noun	Verbal noun	Noun + participle	Noun + rel. clause	Noun + PP
Marked	144	19 (13%)	53 (37%)	8 (6%)	6 (4%)	7 (5%)	6 (4%)
Unmarked	209	26 (12%)	88 (42%)	23 (11%)	8 (4%)	10 (5%)	6 (3%)

⁴ Here and further the total number of answers does not suit the sum in columns, because in columns only the relevant constructions were presented (for example, we omit the "direct speech" comments, as *Oh, it's so beautiful* etc.).

⁵ Here and further the ratio of the examples for this strategy to the whole number of dynamic/static situations, accordingly, is given.

A much more relevant factor turns out to be the properties of one of the participants of the situation, to be precise, their markedness (according to the informant's judgement). The clausal strategy is likely to be chosen, if the object is unmarked as to its appearance, is not known by the informant or if the photograph includes several participants (the more participants are present in the forefront, the more the importance of a single one decreases). Tables 4, 5 and 6 show the relevance degree of these three factors:

Table 4. Number of participants

	Total	Embedded clause	Noun	Verbal noun	Noun + participle	Noun + rel. clause	Noun + PP
1	110	9 (8%)	62 (56%)	2 (2%)	7 (6%)	7 (6%)	3 (3%)
2	106	13 (12%)	41 (39%)	9 (8%)	1 (1%)	3 (3%)	3 (3%)
> 2	140	24 (17%)	38 (27%)	20 (14%)	6 (4%)	7 (5%)	6 (4%)

Table 5. Markedness of the OST

	Total	Embedded clause	Noun	Verbal noun	Noun + participle	Noun + rel. clause	Noun + PP
Marked	104	8—9 (9%)	47 (45%)	10 (10%)	6 (6%)	5 (5%)	1 (1%)
Unmarked	249	36—37 (15%)	94 (38%)	21 (8%)	8 (3%)	12 (5%)	11 (4%)

Table 6. Known/unknown participant (OST)

	Total	Embedded clause	Noun	Verbal noun	Noun + participle	Noun + rel. clause	Noun + PP
Known	109	16—17 (15%)	54 (49%)	4 (4%)	4 (4%)	3 (3%)	4 (4%)
Unknown	244	28—30 (12%)	87 (36%)	27 (11%)	10 (4%)	14 (6%)	8 (3%)

It can be seen that the properties of the situation itself (dynamicity, markedness) are less important than the properties of the participants: the latter are much more relevant for the choice of the construction.

We assume that such variation of patterns depends not only on the meaning of particular verbs, but is also conditioned by discourse factors. The point is that the situation, when the valency slot of the verb is not filled by a noun complement (*Ja seržus' na Vas'u* 'I am angry with Vas'a'), but by a clausal argument (*Ja seržus', čto Vas'a ujexal* 'I am angry that Vas'a went out') is marked in Russian. It is well-known that choosing a noun/NP the speaker makes abstraction: he ignores the individual properties of an object that do not have to correspond to the prototypical properties of the class of objects denoted by the said noun/NP

(see Kobozeva 2000: 35—36). The expressing of the situation as a whole requires a higher level of abstraction, because the situation cannot be perceived directly. Therefore, conceiving a situation from the world of reality requires a greater cognitive effort.

It is important that names of material objects are elements of the lexicon, that the speaker takes “as they are”, whereas clausal complements, have to be generated in the course of generating the text, which requires a greater effort from the speaker. Let us consider which way the speaker avoids it.

The SST is a fragment of reality, including a number of participants (*Ja uvidel, čto devuška sidit na motocikle* ‘I saw a girl sitting on a motorbike’). In most cases, one of the participants has the highest discourse status according to his semantic and pragmatic properties: animacy, definiteness, pragmatic salience etc. The speaker generates a sentence as if this participant were the only participant of the perception situation: (*Ja uvidel devušku* ‘I saw a girl’). The SST can in such cases be expressed by means of a relative clause (*Ja uvidel devušku, kotoraja sidit na motocikle* lit. ‘I saw a girl, who is sitting on a motorbike’), a prepositional phrase (*Ja uvidel devušku na motocikle* ‘I saw a girl on a motorbike’) or even be omitted.

However, if the situation is highly salient pragmatically (more salient than any of its participants), the speaker chooses to express it. In such cases the SST becomes an argument of the matrix verb (*Ja uvidel, čto... / Ja slyšal jego koncert* ‘I saw that... / I have heard his concert’) and the OST is demoted to the position of subject in the embedded clause, or genitive subject of the deverbal noun, or it can even be omitted.

Therefore, the OST and the SST divide the object valency of the matrix verb, attracting the speaker’s attention by means of semantic and pragmatic properties, including animacy (see section 2.1), markedness of the appearance (colour, clothes and so on), salience among other participants, quantity of the participants; less relevant is the dinamicity and markedness of the situation.

OST:		SST:
markedness of the object		markedness of the situation
animacy		dinamicity
		
noun	noun + prepositional phrase noun + relative clause	embedded clause verbal noun

The properties of situations that make expressing of the OST highly probable (markedness of the object, salience and animacy), are presented

on the left side of the scale and the ones that make SST probable (markedness of the situation, dinamicity) are on the right side.

We assume that the scale explains the results of the first experiment: in particular, the fact that OST and SST are rarely expressed both as arguments of the matrix verb (see 1.2). The reason is that the situation as a whole and its participants “compete” for the right to fill the valency slot of the matrix verb: if one of the participants is more salient, the situation tends to be expressed as an adverbial, and vice versa, when the situation is more salient, its arguments are demoted.

3. Conclusions

We suppose to have shown that choice of the diathesis of a particular verb is influenced not only by the idiosyncratic semantic properties of the verb, but also by pragmatic factors. In this case we face an important question: whether the diversity of constructions reflects distinct lexical meanings or distinct constructions within the same lexical meaning. In the reality, the situation, designated by the verbs under consideration, has two participants: the one referring to the Stimulus and the one referring to the Experiencer. The syntactic realization of the Stimulus referent can be “narrow”, including one participant only, or “wide” — in that case it includes the whole situation (these two cases are related by means of metonymy, as in Gak 1976).

We suppose the main result of our work to be the following: we have shown that the observed diversity of syntactic constructions is not an idiosyncratic feature of one or several verbs — it is a common property of all the matrix verbs under consideration. Above we have defined the factors that influence the choice of a construction in a particular context. On the level, where the speaker's cognitive task is formed, SST and OST compete for the right to fill the valency slot of the verb. The choice between SST and OST is made after defining the relative weight of a number of pragmatic factors. The speaker decides, first, whether the SST / OST should be “promoted” to a complement position or “demoted” to an adjunct position and, second, whether they are obligatory to express.

We argue that these constructions arise as a result of competition between the object and the situation on the cognitive level. It can be said that mental verbs have “narrow” and “wide” diatheses: in the former only a participant of the situation is realized. The latter is a result of abstraction occurring when all the participants of the situation are not salient.

What is important, there are two types of diatheses independently of the relations of these diatheses to each other: for example, the verb *videt*

‘see’ has only one valency slot that can be filled either by OST or by SST, and emotive verbs (like *serdit’sja* ‘be angry’, *obižat’sja* ‘be offended’) have two different valency slots for SST and OST (see in detail Padučeva 2004). However, both perception and emotive verbs have narrow and wide diatheses and must choose between them: it is not accidentally that SST and OST are rarely expressed in the same sentence.

We analyzed only some of the principles relevant for the expression of SST/OST. For example, we did not pay any attention to the idiosyncratic semantic properties of particular matrix verbs. To account for some differences a more detailed analysis of concrete matrix verbs (as the analysis of *videt’*) is required.

References

- Apresjan Ju. D. 1964. O sil’nom i slabom upravlenii (opyt količestvennogo analiza) // *Voprosy jazykoznanija*. 1964. № 3. P. 32—49.
- Apresjan Ju. D. 1998. Kausativy ili konversivy? // *Tipologija. Grammatika. Semantika. K 65-letiju Viktora Samuilovicha Xrakovskogo*. Saint-Petersburg, 1998. P. 273—281.
- Apresjan Ju. D. 1999. Slovarnye stat’i *znat’ I*, *videt’ I.I.*, *serdit’sja*, *radovat’sja*. // Ju. D. Apresjan et al. *Novyj ob’jasnitel’nyj slovar’ sinonimov russkogo jazyka*. Vol. 1. Moscow: Jazyki slavyanskoj kul’tury, 1999.
- Arutjunova N. D. 1976. *Predloženie i jego smysl*. Moscow: Nauka, 1976.
- Volf E. M. 1989. Emocional’nye sostojanija i ix predstavlenija v jazyke. // *Logičeskij analiz jazyka. Problemy intensional’nyx i pragmatičeskix kontekstov*. Moscow: Nauka, 1989.
- Gak V. G. 1976. Nominalizacija skazujemogo i ustraneniye subjekta. // *Sintaksis i stilistika*. Moscow: Nauka, 1976. P. 85—102.
- Kirsner R., Thompson S. A. 1976. The role of pragmatic inference in semantics: a study of sensory verb complements in English // *Glossa*, v. 10 (1976).
- Kobozeva I. M. 2000. *Lingvističeskaja semantika*. Moscow: URSS, 2000.
- Padučeva E. V. 2004. *Dinamičeskije modeli v semantike leksiki*. Moscow: Jazyki slavyanskoj kul’tury, 2004.
- Plungian V. A., Raxilina E. V. 1998. Paradoksy valentnostej. // *Semiotika i informatika*, vol. 36. Moscow: Jazyki russkoj kul’tury, 1998. P. 108—119.
- Xrakovskij V. S. 1999. *Ponjatije sirkonstanta i jego status*. // Xrakovskij V. S. *Teorija jazykoznanija. Rusistika. Arabistika*. Saint-Petersburg: Nauka, 1999. P. 102—112.
- Zaliznjak Anna A. 1992. *Issledovanija po semantike predikatov vnutrennego sostojanija*. München, 1992.

Многофункциональность аппарата лексических функций и некоторые возможности его применения и развития *

З. М. Шаляпина

(Институт востоковедения РАН)
languages@bk.ru

The Formalism of Lexical Functions: Some of its Possible Uses, Applications and Extensions. The formalism of lexical functions (LF), designed as a means of generalizing and controlling lexical aspects of periphrastic transformations, is as multifunctional as such transformations are manifold. In automatic text processing, this makes possible its employment for more than one purpose. The paper outlines some applications and extensions of this formalism and its underlying ideas within an automatic translation framework.

Аппарат лексических функций (ЛФ) был создан для обобщения и моделирования лексического аспекта перифрастических преобразований; однако поскольку эти преобразования весьма многогранны, многофункциональным оказывается и аппарат ЛФ, что открывает для него различные приложения. В данной статье описаны некоторые пути применения и развития данного формализма и заложенных в нем идей в рамках задачи автоматического перевода.

Аппарат лексических функций (ЛФ) создавался, как известно, в первую очередь как инструмент моделирования перифрастических преобразований при семантическом синтезе естественно-языковых текстов [1]. В этих целях он должен прежде всего обеспечивать в словаре систему «горизонтальных» связей — перекрестных ссылок от каждой лексической единицы (ЛЕ) ко всем другим ЛЕ, которые могли бы заменять или сопровождать ее при таких преобразованиях.

* Работа выполнена при частичной финансовой поддержке РФФИ, грант № 04-01-00623.

В принципе, в том идеальном случае, когда все единицы словаря имеют в нем полное и точное формализованное описание всей совокупности своих лингвистических свойств, в качестве требуемых перекрестных ссылок было бы достаточно давать в статье каждой ЛЕ простой перечень тех единиц, которые способны заменять или сопровождать ее в трансформах включающих ее фрагментов текста. Выбор тех из единиц данного перечня, которые требуются в конкретном случае таких преобразований, мог бы осуществляться путем сравнения описаний, сопоставленных этим единицам в их собственных словарных статьях, со свойствами преобразуемого контекста и с теми языковыми требованиями к нему, нарушение которых инициирует в данном случае процедуры его преобразования.

Понятно, однако, что такая организация модуля языковых преобразований была бы неоптимальной даже при данном (очевидно, нереальном) предположении о полноте и точности описания всех возможных ЛЕ. Более эффективно заранее установить наиболее характерные типы возможных преобразований и, выявив допустимые в их рамках соотношения между исходными и отсылочными ЛЕ, классифицировать их в соответствии с полученной типологией — с тем, чтобы в ситуациях, подпадающих под те или иные из этих типов, обращаться только к соответствующим классам отсылочных единиц. Такую типологию и позволяет задавать аппарат ЛФ, благодаря чему на его основе может строиться не просто индекс горизонтальных межлексемных связей, но развернутая система управления базой лексических знаний в процессе текстовых преобразований, обеспечивающая контроль за их типом и направлением.

В различных случаях преобразования фрагментов текста могут затрагивать, естественно, разные его аспекты, так что обобщение участвующих в таких преобразованиях межлексемных отношений при их классификации в терминах ЛФ должно опираться на множественные основания. Эта классификация отражает, в частности, элементы смысла, часто встречающиеся в толкованиях разных ЛЕ в сходных позициях; ограничения, налагаемые отдельными ЛЕ на выбор в содержащем их текстовом фрагменте лексических означающих для других участников задаваемой им ситуации; лексико-синтаксические возможности передачи морфологических категорий; валентностные соотношения между синтаксически связанными в тексте участниками его структуры; различия средств обозначения одних и тех же или близких ситуаций по коммуникативной организации, по степени идиоматичности, избыточности, стилистической маркированности и т. п.

Все эти типы лингвистической информации не специфичны только для перифрастических преобразований, но могут участвовать и в других видах обработки текстов. Поэтому система ЛФ, «вбирая» в себя элементы такой разноаспектной языковой информации, общие для межлексемных отношений различных классов, неизбежно приобретает многофункциональный характер. Рассмотрим в этой связи интерпретацию и реализацию некоторых из функций ЛФ в экспериментальном комплексе для исследований по японско-русскому автоматическому переводу ЯРАП [5—7] с его тремя основными подсистемами: японского анализа, японско-русского межъязыкового перехода и русского синтеза.

1. Собственно перифрастические функции аппарата ЛФ применяются в комплексе ЯРАП прежде всего в соответствии с основным назначением этого аппарата — для синтеза выходных текстов, выполняемого подсистемой русского синтеза RussLan (*Russian Language generation*). Структуры, формируемые при межъязыковом переходе и поступающие на синтез, в силу как ограниченности средств обработки входных текстов, так и различий между входным и выходным языками, могут быть неполными и/или неточными, что и должны, по возможности, компенсировать их перифрастические преобразования. Для этого в подсистеме RussLan разработан механизм так наз. локальных трансформаций, позволяющий, среди прочего, оперировать с ЛФ, которые вместе со своими значениями задаются в статьях участвующих в трансформациях лексем в специальной зоне ЛФ - о т с ы л о к. Данный механизм активируется, в частности, при несоответствии семантических характеристик обрабатываемых ЛЕ селекционным ограничениям, налагаемым на них их структурными спутниками — противочленами этих ЛЕ в релевантных для них отношениях зависимости.

Примером могут служить структуры, где субъектную валентность при глаголе информационного действия заполняет обозначение информационного объекта или результата. В одних языках это вполне нормативно, в других, включая русский, здесь требуется имя одушевленного деятеля, ср. *The paper studies...* vs. **Статья изучает...* В подсистеме RussLan структуры данного типа преобразуются одним из двух способов. Если для лексемы, задающей информационный объект (результат), не заполнена ее собственная субъектная валентность, эта лексема вытесняется в позиции глагольного субъекта общим именем ее собственного субъекта — значением ЛФ S₁: *Автор статьи изучает*. Если же эта валентность у исходной лексемы заполне-

на, глагол пассивизируется, а его субъект преобразуется в локатив с использованием ЛФ = Loc: *В статье <этого автора> изучается.*

2. Отображение в терминах ЛФ часто повторяющихся элементов лексических толкований позволяет в отсутствие общепринятого семантического языка использовать инвентарь ЛФ в качестве его частичной эмуляции. Многочисленные примеры формализованного описания (контекстной) лексической семантики в терминах ЛФ представлены, в частности, в [3], где, например, значение существительного *место* в ситуации *кинотеатр* определяется как *единица способности* $Fact_2-S_{loc}S_1Real_2(\text{кинотеатр})$, наречия *легко*, *хорошо*, *быстро* при глаголе *засыпать* описываются как $Adv_{1B}MagnPredAble_1(\text{засыпать})$ и т. д. В общем случае синтаксис таких суперпозиций слишком сложен для автоматической обработки — тем более, что они могут включать, как видно из первого примера, и обычные лексемы. Но ЛФ можно применять для представления лексической семантики в упрощенном варианте — как вид семантических дескрипторов.

Такая возможность реализована в системе RussLan за счет того, что в ее словарных статьях для описания семантических свойств лексем предусмотрено две разные зоны: в зоне Сем-классов даются семантические признаки обычного (не лексико-функционального) типа, в зоне ЛФ-семантики — ЛФ, значением которых может являться заглавная лексема. И те и другие могут указываться в числе селекционных ограничений, свойственных валентностям тех или иных ЛЕ, и привлекаться при проверке таких ограничений в качестве семантических характеристик проверяемых лексем.

Большинство ЛФ берется из их стандартного инвентаря (с изменением формы их записи так, чтобы ввести подстрочные индексы в строку), другие строятся по аналогии с первыми. Например, помимо локативных предложных ЛФ Loc_in, Loc_ab, Loc_ad, вводится ЛФ Inf_source — предлог, указывающий на информационный источник. При этом в словаре системы не делается попытки использовать ЛФ для описания лексической семантики сколько-нибудь точным и исчерпывающим образом: они указывают только на факт релевантности для лексемы соответствующих смыслов. Так, для глагола *принимать* в зоне ЛФ-семантики даются две ЛФ: Oper1 и Oper2 (каждая из которых актуализируется в контекстах своего типа), для предлога *из* — три: AdvCaus, Inf_source и Loc_ab и т. п.

3. Некоторым ЛФ свойственна специфическая для них функция эмуляции морфологических категорий, прежде всего категорий синтаксического словообразования [1, с. 187].

В системе RussLan эта функция получает дополнительное развитие, позволяя использовать ЛФ во всех случаях невозможности построения требуемой при синтезе словоформы. В этих рамках вводится ряд дополнительных ЛФ, например, *Neg* и *Pos*, которые задают соотношения типа *все — никто*, *любой — никакой* и т. д., требуемые при выборе (или модификации) соответствующих лексем в зависимости от наличия / отсутствия отрицания при управляющем глаголе. Аналогичным способом может моделироваться супплетивизм: при лексемах с супплетивным выражением тех или иных категорий достаточно указать эти категории в качестве ЛФ, имеющих своими значениями требуемые супплетивные лексемы.

Реализован также особый морфолого-синтаксический (МС-) тип локальных трансформаций, на котором базируется как построение аналитических форм, так и перифрастические способы «выхода из парадигматического тупика», когда требуемая форма отсутствует либо для данной конкретной лексемы (как форма множественного числа для лексемы *молоко*), либо вообще в системе морфологии данного языка (как деепричастие будущего времени или сослагательного наклонения в русском языке). По возможности, при этом используются стандартные ЛФ — скажем, при необходимости образовать форму 1-го лица единственного числа от глагола *победить* вводится *Oper1*: *одержжу победу*. Но наряду с ними вводятся и некоторые дополнительные, например, ЛФ *Spec* (от *species*) со значениями *вид (молока)*, *сорт (продукции)* и т. п. Предусмотрены также МС-трансформации, действующие «по умолчанию» — в отсутствие пригодных ЛФ. Так, в качестве глагольного коррелята от лексемы *бюрократ* строится перифраза *делать то, что характерно для бюрократа*; форма деепричастия сослагательного наклонения преобразуется в конструкцию с *если бы* и т. д.

Механизм МС-трансформаций действует и при разворачивании фразеологических единств, расширяя тем самым возможности применения аппарата ЛФ для описания фразеологии. Именно: компоненты фразеологического единства типа *собаку_съесть* даются в его статье в зоне ЛФ-отсылок как значения особых «компонентных» ЛФ и замещают его при МС-синтезе по правилу МС-трансформации, определяемому типом самого единства.

4. Одной из центральных функций ЛФ является отображение ими отношений лексической обусловленности — зависимости лексических означающих одних участников выражаемой в тексте ситуации от означающих других ее участников. Данный тип межлексемных

отношений, в свою очередь, имеет несколько аспектов, поскольку такая обусловленность может быть как полностью идиоматической, так и связанной с какими-либо содержательными факторами. Так, недопустимость выражения *кулек молока (в отличие от *пакет молока*) определяется не столько свойствами самих данных лексем, сколько видом упаковки, используемой для молока в современном производстве: она не делается в виде кульков.

Аппарат ЛФ ориентирован прежде всего на языковую идиоматику. Но обработка текстов требует учета и тех межлексемных отношений, которые отражают устройство мира. Развитие идеи ЛФ в этом направлении в свое время привело, в частности, к определению понятия энциклопедических функций [2]. На этом же пути лежит распространение данного аппарата на задачу выбора конкретизирующих переводных эквивалентов в ситуациях так наз. «нерасчлененности» («неопределенности») смысла переводимой единицы. Она особенно актуальна при переводе на русский язык с таких языков, как французский или японский, где стилистически немаркированными именованиями действий и процессов часто являются единицы с более общей семантикой, чем в русском. Ср. франц. *sortir* или яп. *дэру* с русск. *покидать*, которому в нейтральном стиле соответствуют глаголы, уточняющие способ движения: *выходить*, *выезжать*, *вылетать* и т. п.

В системе ЯРАП японские единицы подобного типа заменяются при межъязыковом переходе либо русскими ЛЕ той же смысловой общности (отвлекаясь от стиля), либо, если таковых нет, наиболее частотными из конкретизирующих эквивалентов. В обоих случаях переводы получают помету о желательности их коррекции. В русском словаре в статьях лексем, допускающих такую коррекцию в зависимости от предпочтений их контекстных спутников, также может даваться аналогичный признак, а в зоне ЛФ-семантики перечисляются те элементы их смысла, которые могут при этом уточняться. В свою очередь, в статьях лексем, имеющих предпочтения для выражения в их контексте тех или иных смыслов, в зоне ЛФ-отсылок указываются эти предпочтения. В ходе синтеза при обработке ЛЕ с пометами о коррекции происходит обращение к их структурным спутникам, проверяется, есть ли у тех предпочтения для выражения задаваемых данными ЛЕ смыслов, и при наличии таких предпочтений исходные ЛЕ заменяются на предпочитаемые.

Понятно, что применение ЛФ в этих целях требует соответствующего расширения их инвентаря: он должен быть дополнен набором всех смыслов, для которых в данном языке возможна подобная

коррекция. В частности, для русского языка в этот инвентарь включаются ЛФ *Move*, *Mv_in*, *Mv_out*, *Mv_thru*. В статьях *перемещаться*, *проникать*, *покидать*, *идти*, *входить*, *выходить* и т. п. они даются в зоне ЛФ-семантики, в статьях лексем, обозначающих объекты, для которых типичны определенные способы движения, — в зоне ЛФ-отсылок вместе с лексемами, задающими эти способы. В результате, скажем, по входной структуре, которая без лексической коррекции давала бы результат: «Птицы покидают гнезда, а змеи покидают норы», синтезируется более нормативная фраза: *Птицы вылетают из гнезд, а змеи выползают из нор*. Тот же механизм реализует собственно лексическую обусловленность, позволяя вместо «Публика начала сильные аплодисменты» получать на выходе *Публика разразилась бурными аплодисментами*.

Важную роль играет лексическая коррекция данного типа для обработки предлогов. В системе ЯРАП действует принцип максимального сохранения при анализе и межъязыковом переходе всех лексических и структурных характеристик исходного текста. В частности, сохраняются и переводятся все функциональные единицы, включая японские аналоги предлогов — «падежные» частицы и послелого, заменяемые при этом на свои наиболее стандартные русские корреляты. В случаях, когда такими коррелятами являются словоизменительные категории — скажем, аккузатив и датив для частиц, маркирующих, *gesp.*, прямое и адресатное дополнения, переводным эквивалентом исходной единицы считается «пустой» предлог, вводящий требуемую категорию. В результате в структурах, поступающих на синтез, присутствуют как слабоуправляемые, так и сильноуправляемые предлоги, и все они, в том числе пустые, подвергаются при синтезе лексической коррекции по требованиям как управляющего предиката, так и припредложного существительного (если только не указан явный запрет на такую коррекцию). При этом управляющий предикат может модифицировать семантику предлога, что, в свою очередь, может влиять на способ его уточнения со стороны припредложного существительного. Так, если при межъязыковом переходе лексема *фотография* оказалась подчинена предлогу *от* или *из*, то в случае подчинения его глаголу *смотреть*, актуализирующему для этого предлога смысл отношения к «исходной точке», данная лексема «предпочтет» предлог *с*; если же предлог, подчиняясь предикату *видно*, получает смысл «информационного источника», то она потребует предлога *по*, ср.: *Это лицо смотрело на него с фотографии / из окна; Это ясно по фотографии / из анкеты*.

5. Наконец, последняя из интересующих нас здесь функций ЛФ состоит в способности многих из них задавать валентностные соотношения своих аргументов и значений.

Такие соотношения имеют принципиальное значение для синтаксической структуры текста, поскольку за их счет создается возможность не только прямой реализации валентностей — в виде связей зависимостей, в которых непосредственно участвуют лексемно-носители реализуемых валентностей, — но и косвенной, когда валентности «делегируются» от своих исходных носителей к другим ЛЕ, связанным с первыми по цепочке зависимостей, и реализуются в качестве валентностей уже этих других ЛЕ. Среди ЛФ данную функцию выполняют прежде всего так наз. лексические оформители синтаксических отношений [1, с. 204]: Oper_i, Func_i, Labor_{ij}, а также все другие ЛФ, если они сопровождаются индексами — номерами валентностей, «заимствуемых» словами-значениями этих функций от их аргументов. Однако косвенную реализацию валентностей могут обеспечивать и ЛЕ, не укладывающиеся в классическое определение ЛФ: от вспомогательных, модальных, фазовых глаголов до сильноуправляемых предлогов, с одной стороны, и конструкций с «расщепленным посессором» (типа *он махнул рукой*, где зависимость между глаголом *махнул* и его подлежащим *он* представляет собой прямую реализацию субъектной валентности самого глагола и косвенную реализацию посессивной валентности его дополнения *рукой*), а также предикатно-актантных структур со словами типа *опоздать* (напр., с *завершением работы*), *ожидать* (напр., *похвалы*), *суждено* (напр., *встретиться*) и т. п., с другой.

Во всех таких конструкциях можно усматривать проявление единого, весьма общего языкового механизма — контекстного валентностного (КВ-) наследования [4], в рамках которого ЛЕ, делегирующие свои валентности для реализации при их структурных спутниках, — в том числе аргументы ЛФ, — выступают как КВ-наследодатели, а ЛЕ, которым они передают свои валентности, — как КВ-наследники.

В зависимости от свойств тех и других, данный механизм может иметь несколько модификаций. В наиболее стандартной из них исходная валентность наследодателя «совмещается» при ее наследовании с одной из собственных валентностей наследника таким образом, что семантическая интерпретация и селекционные ограничения «совмещенной» валентности заимствуются от представляемой ею валентности наследодателя, а способы ее оформления при реализации определяются наследником (тот, так сказать, реализует доставшееся

ему валентностно-семантическое наследство «по своему усмотрению»). В других модификациях КВ-наследования наследуемая валентность передается наследнику вместе с требованиями к ее оформлению при реализации (= наследодатель как бы «оговаривает» способы реализации передаваемого им наследства). Иногда при этом также происходит совмещение наследуемой валентности с представляющей ее валентностью наследника, и требования к ее оформлению «встраиваются» в сведения об оформлении самой соответствующей валентности наследника. Ср.: *Задача сводилась к тому, чтобы найти виновника* vs. *Вопрос сводился к тому, кому это выгодно* vs. *Ответ сводился к тому, что не хватает информации* (одинарным подчеркиванием здесь выделены оформители, вводимые наследником — глаголом *сводиться*, двойным — единицы, наличие которых обусловлено наследодателями — *resp.*, *задача*, *вопрос* и *ответ*). Но чаще при наследовании валентности вместе с ее оформлением она просто добавляется к модели управления наследника как новая его валентность, свойственная ему только в данном конкретном контексте. Такой полный перенос валентностей особенно хорошо виден на примерах, где наследодатель отделен от наследника границей клаузы, ср.: *занятия, которые X вел у первокурсников* vs. *агитация, которую X вел среди молодежи* vs. *полемика, которую X вел с Y по поводу Z* и т. п.

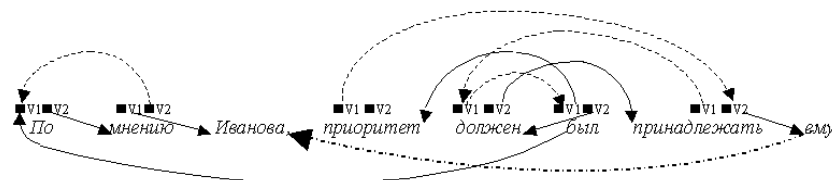
Обычно КВ-наследодатель заполняет (прямо или косвенно) одну из первых трех валентностей КВ-наследника, что в аппарате ЛФ отражается противопоставлением *Func_i* vs. *Oper_i* vs. *Labor_{ij}*. Но есть и конструкции, в которых он реализует 4-ю валентность, ср.: *X довел Y-а до нервного срыва постоянными придирками*, где субъектная валентность глагола *довел* одновременно представляет субъектную валентность его четвертого актанта *придирки*. Этот же пример демонстрирует способность КВ-наследников наследовать валентность сразу от двух разных наследодателей: объектная валентность глагола *довел* совмещена с субъектной валентностью его третьего актанта (*нервный*) *срыв*. То, какая валентность наследодателя передается КВ-наследнику, обычно определяется свойствами КВ-наследника (его способностью к субъектному, объектному и т. п. контролю). Однако это может зависеть и от наследодателя — точнее, от того, какая из его валентностей в большей степени семантически совместима с «принимающей» валентностью наследника, ср.: *X-у грозит опасность / арест / конфискация*; *X рассчитывает на успех / премию / вручение награды* и т. п.

В общем случае как та валентность КВ-наследника, по которой он связан с наследодателем, так и та, которая представляет при нем

наследуемую валентность, может, в свою очередь, реализоваться косвенным образом, так что в контексте может строиться целая цепочка последовательных «передач» одной и той же валентности, ср.: *Генерал мог решить отказаться выполнить приказ об отводе войск или Его должна была начать мучить совесть.*

Возможна и неоднозначность определения таких цепочек. Так, в японских конструкциях *X ва Y га Z*, где *X* и *Z* — *главный* и *зависимый* предикаты, в зависимости от их семантического соотношения, от *Y* к *Z* может делегироваться и реализоваться в виде его собственного актанта *X* как агенса *Y*-а, так и его пациенс (в том числе при наличии у *Y* обеих валентностей).

Отношения КВ-наследования должны учитываться, очевидно, на различных этапах обработки текстов наряду со связями зависимостей, устанавливаемыми при прямой реализации валентностей, и отношениями кореферентности. Поэтому в системе ЯРАП они в явном виде включаются в синтаксическую структуру текста, которая определяется, таким образом, как суперпозиция всех трех типов отношений. В графическом представлении их можно обозначать стрелками разных типов — например, так, как это сделано ниже:



(зависимости представлены здесь сплошными стрелками, связывающими реализуемую валентность и ее заполнитель, КВ-наследование — пунктирными стрелками, идущими от делегируемой валентности наследодателя к представляющей ее валентности КВ-наследника, кореферентность — штрих-пунктирной стрелкой, идущей от местоимения к его антецеденту).

Формальные средства, разрабатываемые в рамках системы ЯРАП для моделирования КВ-наследования как самостоятельного аспекта структуры текста, не совпадают, естественно, с аппаратом ЛФ. В частности, грамматические признаки, приписываемые словарным единицам для отображения их свойств как потенциальных КВ-наследников и/или наследодателей, отличаются от ЛФ тем, что входят в иерархию дальнейшего парадигматического обобщения. Однако в основе своей эти средства во многом восходят к «индексированным» ЛФ и реализуют заложенные в них идеи.

Резюмируя сказанное, подчеркнем, что аппарат ЛФ, «вырезая» тот участок языковой системы, для которого наиболее характерны перифрастические преобразования, и тот пучок межлексемных соотношений, которые наиболее активно в них участвуют, предстает тем самым в качестве стимула и инструмента для изучения и описания «высвечиваемых» им лексико-синтаксических явлений далеко за пределами этого первоначального участка.

Литература

- Жолковский А. К., Мельчук И. А.* О семантическом синтезе // Проблемы кибернетики. Вып. 19. М.: Наука, Гл. ред. физ.-мат. лит.-ры, 1967. С. 177—238.
- Леонтьева Н. Н., Волковская Е. В., Молчанова Т. В.* Автоматическое индексирование с помощью словаря энциклопедических функций // Автоматическая переработка текста методами прикладной лингвистики. Кишинев, 1971. С. 257—259.
- Мельчук И. А., Жолковский А. К.* Толково-комбинаторный словарь современного русского языка. Вена: Wiener Slawistischer Almanach. Sonderband 14. 1984.
- Шаляпина З. М.* Структурные валентности как универсальный инструмент описания языковой сочетаемости (при сущностном подходе к ее моделированию) // Московский лингвистический журнал. 2001. Т. 5. № 2. С. 35—84.
- Шаляпина З. М., Модина Л. С., Канович М. И. и др.* Экспериментальный комплекс ЯРАП для лингвистических исследований в области японско-русского автоматического перевода: первая очередь // Деп. в ИНИОН РАН 5.11.01. №. 56804. М.: ИВ РАН, 2001. 404 с.
- Shalyapina Z. M.* Japanese with relation to Russian in the JaRAP system of Japanese-Russian automatic translation // International Conference. Japan Phenomenon: Views from Europe. Moscow, 2001. P. 132—139.
- Shalyapina Z. M., Kanovich M. I., Shternova O. A. et al.* RUSSLAN: A System of Russian Language Generation // Investigations into Formal Slavic Linguistics. Contributions of the Fourth European Conference on Formal Description of Slavic Languages — FDSL IV. Part I // Linguistik International. Band 10. Frankfurt am Main et al.: Peter Lang Verlag, 2003. P. 385—403.

Meaning-Text Theory for Textual Input Analysis and Proofing in a Generation System

Svetlana Sheremetyeva

Department of Computational
Linguistics,
Copenhagen Business School
lanaconsult@mail.dk

Olga Babina

Department of Intercultural
Communications,
Southern Ural State University
olga_babina@mail.ru

The paper presents a new tool for textual input proofing developed for AutoPat, — a patent claim generator. AutoPat relies on human input to build a deep content representation and then transforms it into a surface text using metadata stored in the lexicon. The quality the input is thus essential for the quality of output. The tool has been implemented for the English language but can be extended to other languages.

Модель Смысл-Текст и коррекция текстовых данных в системе автоматического порождения документов. В статье описан новый компонент системы АвтоПат, генератора формул изобретений, предназначенный для автоматической коррекции поступающих на вход системы текстовых фрагментов, на основе которых с использованием лексикона системы строится глубинная структура, трансформируемая в текст. Корректность вводимых пользователем фрагментов во многом определяет качество генерации. Компонент имплементирован для английского языка.

Introduction

The patent claim is the focal point of a patent disclosure and is the actual subject of legal protection. The claim must consist of a single, albeit possibly very complex, sentence with a well specified conceptual, syntactic and stylistic/rhetorical structure. Figure 1 illustrates a fragment (claims can be over a page long) of a US claim text.

*A cassette for holding excess lengths of light waveguides in a splice area **comprising** a cover part and a **pot-shaped** bottom part **having** a bottom disk and a rim **extending** perpendicular to said bottom disk, said*

*cover and bottom parts **are superimposed** to enclose jointly an area **forming** a magazine...*

Fig. 1. A fragment of a patent claim text. Predicative words, which are heads of individual phrases describing essential features of the invention, are bold faced

Superficially our system conforms to the standard emerged in natural language generation, in that it includes the stages of content specification resulting in the production of a meaning representation of a nascent claim, text planning and surface realization (Reiter, 1994).

The specificity of our system is that AutoPat is equipped with a knowledge elicitation scenario and relies on joint human-computer content specification in a textual form. The quality of this textual input is a crucial point in getting a high quality output that makes proofing tools its essential components (see, also, e.g., (Nyberg et al., 2003) about the importance of proofing tools for NLP, dealing with a human input). AutoPat data flow is shown in Figure 2.

Using semantico-syntactic metadata stored in the lexicon and joint human-computer content specification AutoPat first builds a deep content representation and then transforms it into a surface text (Sheremetyeva and Nirenburg, 1996; Sheremetyeva, 2003). As our theoretical framework we thus adopted *MEANING-TEXT THEORY* (Mel'cuk 1981) that operates on the principle that language consists in a mapping from the content or meaning representation to its text through a number of intermediate levels of representation.

In what follows we first describe the lexicon of the system, the core component of any application based on *MEANING-TEXT THEORY* and then show how the MTT framework is used to enhance analysis and human input proofing including grammar, and content checking.

The Lexicon

The essential part of AutoPat knowledge is coded in the system lexicon. Following Mel'čuk (1992: 332), we define lexicon as "a specific list of lexical units of a language, arranged in a specific way and supplied with specific information, the whole being designed for a specific purpose". The set of parameters (or fields) for lexicon specification in the AutoPat knowledge base is strictly determined by the needs of application, and draws heavily on the sublanguage corpus analysis. It includes: *a tagging lexicon* of claim wordforms and *a deep (information-rich) lexicon*

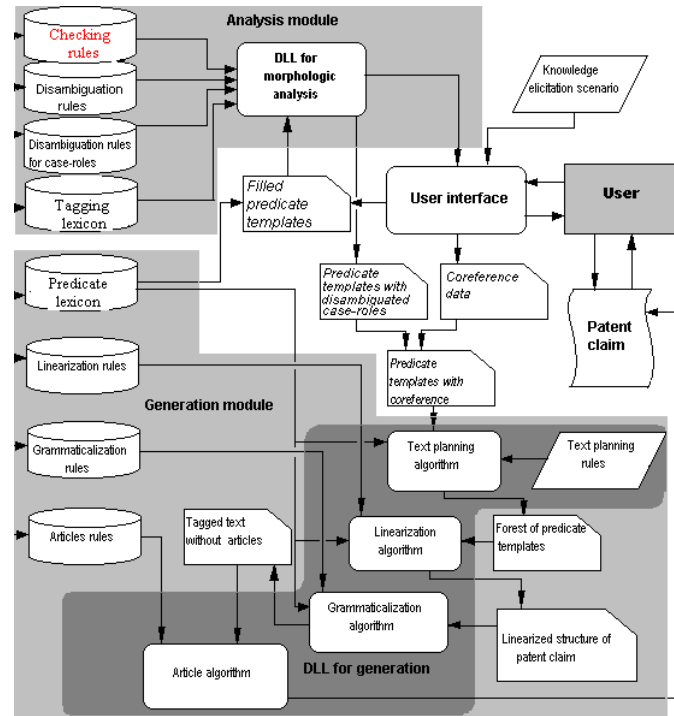


Fig. 2. Architecture and data flow in AutoPat

of predicates. Predicates are lexical units that describe properties or relations between invention elements (bold-faced in Figure 1). Proofing tools are based on the domain specific knowledge stored in these lexicons.

2.1. Tagging lexicon

A tagging lexicon is a lexicon of corpus-based wordforms listed together with their supertags. We define the term “supertag” as a label, which codes richer information than standard POS tags. The use of supertags, as noted in (Joshi and Srinivas, 1994), who coined this term, localizes some crucial linguistic dependencies, and thus show significant performance gains.

The content of supertags differs from work to work and is tailored for the needs of an application. For example, Joshi and Srinivas (1994) use elementary trees of Lexicalized Tree-Adjoining Grammar for supertagging lexical items. In (Gnasa and Woch, 2002) it is grammatical structures of the ontology that are used as supertags.

Our supertag codes morphological information (such as POS and inflection type) and semantic information, an ontological concept, defining a word membership in a certain semantic class (such as object, process, substance, etc.). At present we use 43 supertags that are combinations of 1 to 4 features. For example, the feature structure of noun tags is as follows:

Supertag [POS [Noun [object [pl, sg] process-ing [pl, sg] process-other [pl, sg] substance [pl, sg] measuring unit [pl, sg] parameter [pl, sg] other [pl, sg]]]]]

For example, the supertag “process-other, singular (coded Nf) means that a word is a noun in singular (N), means a process (f), and does not end in *-ing*. This tag will be assigned, for example, to such words as *activation* or *alignment*. This lexicon is used for user input analysis and proofing.

2.2. Predicate Lexicon

The lexicon of predicates is the essential part of the system knowledge, which covers both the lexical, and, crucially for our system, the knowledge that links syntax and semantics. We call predicates those lexical units that are used to specify properties and relations of the invention elements (see, for example, boldfaced words in Figure 1). The structure of a predicate lexicon entry conforms to what is called “a government model of a verb” (Apresyan, 1974); it stores valence properties of predicates on semantic and syntactic levels. The entries in this lexicon include

Semantic knowledge: SEM_CI — semantic class such as meronymy, property, spatial, connection, etc., and CASE_ROLES, — a set of case-roles (Fillmore, 1970) associated with a predicate, such as, *place*, *manner*, *purpose*, etc. This information is used to build internal meaning representation of the nascent claim.

Meaning-text linking knowledge: FILLERs, — sets of most probable surface fillers of case-roles in terms of types of phrases and lexical preferences, PATTERNs, — linearization patterns of predicates that code both the knowledge about co-occurrences of predicates with their case-roles and their linear order. Thus, for example, the following phrase from an actual claim: (1: *the splice holder*) *: *is arranged* (3: *on the cover part*) (4: *to form a rotatable splice holder*) (where 1, 3 and 4 are case role codes and “*” shows the position of the predicate), will match the linearization pattern (1 * 3 4), and RANK, — corpus-based predicate frequency within one semantic class (affects lexical selection).

Morphological knowledge: MORPH, — corpus based predicate word-forms (Sheremetyeva, cf.). This information is used at the generation stage. We base our proofing on the check whether the user's input conforms to the knowledge stored in the lexicons.

Analysis

The AutoPat Analyser switches on simultaneously with the knowledge elicitation procedure and starts with the user supplying natural language phrases into the system in the process of computer interview. Every analysis step relies on static knowledge of the system and on the dynamic knowledge collected by the previous analyzing procedures. The Analyser, includes

a submodule of interactive semantico-syntactic analysis; the top-level procedure here is **retrieving a predicate template**. Following the user selection in a predicate menu it presents the user with a predicate template from the predicate lexicon (See Figure 4). The next procedure is **filling a predicate template by the user**. It creates filled predicate structures, records phrase-filler borders and semantic status of case-roles; and

a submodule of automatic morphological analysis which in our system is reduced to **supertagging** (further « tagging »), a look-up of words in the Tagging lexicon followed by the **tag disambiguation** procedure that attempts to disambiguate multiple tags. The characteristic feature of our Analyser is that tagging is done after the interactive semantico-syntactic analysis rather than before which makes it possible for tag disambiguation to use meaning-text linking knowledge from the predicate lexicon (semantic class of a predicate and its lexical meaning, phrase boundaries, semantic status of case-roles, fillers allowed for a certain case-role). Constraint-based hand-crafted disambiguation rules are thus specified over a much richer feature space compared to traditional taggers. The Analyser produces an internal meaning representation for each quantum of knowledge supplied by the user in the form of a shallow meaning representation as follows

```
text::={ template){template}*
template::={predicate-class predicate ((case-role)(case-role})*1
case-role::=(rank status value)
value::={word tag}*
```

¹ **template** in the meaning representation is retrieved from a predicate lexicon following the user's predicate selection from a system menu. The interface presentation of this template is shown in Figure 3.

where *predicate-class* as specified in the predicate lexicon, *predicate* is a string corresponding to one of predicates from the predicate lexicon, *case-roles* are *ranked* according to the frequency of their co-occurrence together with each predicate in the training corpus, *status* is the semantic status of a case-role, such as agent, theme, place, instrument, etc., *value* is a string which fills a case-role. Tag is a label from a tagging lexicon.

Proofing

User input proofing is of special importance in AutoPat due to the specificity of generation algorithm and user constraints, — English language proficiency and general human limitations. The generator treats case-role fillers as segments of text, which appear in the nascent, claim exactly as supplied by the user, with the exception of determiners, which are inserted at the last stage of the generation procedure. The order of the case-role fillers and predicates is to a large extent determined by the case-role linear pattern from the predicate lexicon. This requires well-formed English phrases put in the right slots of a predicate template. For example, if the phrase “*basically perpendicular*” were typed in the slot “What” (see figure 4), a predicate phrase (sentence) with the wrong word order would be generated: “**basically perpendicular to said bottom disk extends at least one steel rim for rotation*”. Unfortunately the user does such kinds of mistakes especially if s/he is not a native English speaker². Another type of mistakes is wrong or omitted prepositions. For example, if all the fillers were put in the right template slots, but the phrase filling the slot “What for” (case-role “purpose”) were left uncorrected as it is shown in Figure 4, it would produce semantically incorrect sentence “*at least one steel rim extends basically perpendicular to said bottom disk rotation*”. The phrase “*for table*” would also be considered as an erroneous filler of the slot “What for” (case-role “purpose”), as the “purpose” valence requires in such a case an abstract noun. Correction of such mistakes is done based on the knowledge about government models of predicates in the predicate lexicon.

Thus in addition to spell checking a new tool included into The AutoPat Analyzer performs two types of proofing, — grammar check and content check, the latter checks whether the phrase is in the right template

² AutoPat was tested by both native English speakers and non-native English (Russian and Japanese) speakers. Rules a) and b) are written to cover actual mistakes done by a non-native English speaker.

slot, checks prepositions, semantics of nouns, etc. The proofing procedures switch one after another, first spell-, and then grammar and content check.

Grammar and content check take unambiguously tagged case-role fillers output by the analyzer match them against domain specific set of pattern rules in the system knowledge base. In case a pattern is found the tool returns a diagnostic message or warning and the corresponding segment is marked. Any problems, which are found must be resolved by the user by interacting with the system. Examples of grammar (a) and content (b) checking rules are given below (see also user interface screenshots in Figures 3 and 4):

- a) IF T= $\sim$ Nsg{1} $\sim$ Npl{1} $\sim$ Qu{1} $\sim$ Num{1;"one"}
 THEN
 HINT $\sim$ Qu $\sim$ Num $\sim$ Nsg*Npl
- b) IF (T= $\sim$ NF{1}) AND (PURPOSE)
 THEN
 HINTMOVE "Consider moving to 'What' ",SUBJ
 HINTTEXT "Insert preposition"
 HINT +Prep(for) $\sim$ NF

where

the rule a) reads “ If a string of tags starts with one singular noun followed by one plural noun followed by one quantifier and ends in one numeral “one”, then reorder the string so as it starts with the quantifier followed by the numeral, followed by the former first noun in singular and ends in the second noun whose form is changed into noun, plural”; and

the rule b) reads “If the current string of tags starts with one noun of the semantic class “functional” and if this string is the filler of the case-role “purpose” then either move this filler to the case-role “what”- subject or insert the preposition “for”.

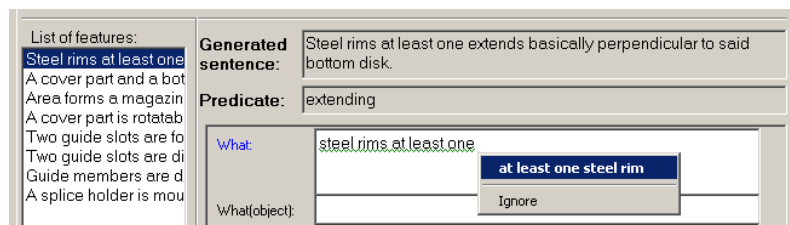


Fig. 3. A diagnostic message in the user interface underlined by the rule a) above.

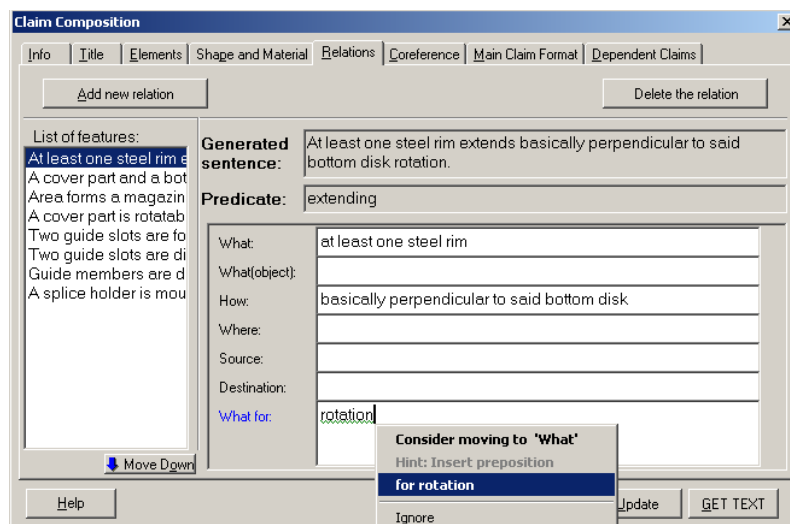


Fig. 4. An interface predicate template with content checker interactive messages underlined by the rule b) above.

The AutoPat proofing tool incorporates different diagnostic messages. Most of the messages suggest rewrites or move to a different slot, which the user can approve by a single mouseclick. Once the problems have been resolved, the predicate template is considered to be correctly filled and may be submitted to generator.

Conclusion

The current version of AutoPat, — a generator of patent claims, is a 32-bit Windows application developed to run in a number of operating environments: Windows 95/98/2000/NT.

In this paper we focused on the human input analysis and proofing components of the system, whose efficiency is conditioned by the *MEANING-TEXT THEORY* framework of the application. The analyser performs two tasks. It returns specific diagnostic messages for case-role fillers that require system and/or user intervention and produces an internal meaning representation for each quantum of the knowledge input by the user, thus building a claim draft to be input into the generator.

As is the case with all linguistic knowledge in the AutoPat generator proofing rules are separated from the program and can be brushed up by a linguist without extra programming effort.

Preliminary proofing tool test results show a reasonably small number of proofing failures, mainly due to the incompleteness of analysis rules that are being improved and augmented with larger involvement of predicate knowledge and extension of the tagging lexicon.

References

- Apresyan, Yu. D. 1974. *Leksicheskaya semantika: Sinonimicheskie sredstva yazyka* (Lexical Semantics: The Synonymic Resources of Language). Moscow: Nauka, 1974.
- Fillmore, Ch. J. 1970. Subjects, speakers and roles. *Synthese*. Vol 21. ¾. 1970. P. 251—274.
- Gnasa, M. and J. Woch. 2002. Architecture of a knowledge based interactive Information Retrieval System. <http://konvens2002.dfki.de/cd/pdf/12P-gnasa.pdf>
- Kittredge, R. and B. Lavoie. 1998. MeteoCogent: A Knowledge-Based Tool For Generating Weather Forecast Texts. In *Proceedings of the American Meteorological Society AI Conference (AMS-98)*, Phoenix, Arizona, 1998.
- Joshi A. and B. Srinivas. 1994. Disambiguation of Super Parts of Speech (or Supertags): Almost Parsing. <http://acl.ldc.upenn.edu/C/C94/C94-1024.pdf>
- Mel'čuk, I. 1981. Meaning-Text Models: A recent trend in Soviet linguistics. *Annual review of Antropology* 10. 1981. P. 27—62.
- Mel'čuk, I. 1992. *Lexicon: an overview*. In: Bright (ed.) *International Encyclopedia of Linguistics*, 2. Oxford, 1992.
- Nyberg E., T Mitamura, D. Svoboda, J. Ko, K. Baker, J. Micher. 2003. An Integrated system for Source language Checking, Analysis and Terminology management. *Proceedings of Machine Translation Summit IX, September*. New-Orleans, 2003.
- Reiter, E. B. 1994. Has a consensus natural language generation architecture appeared and is it psycholinguistically plausible? In *Proceedings of the 7th International Workshop on Natural Language Generation*. 1994. P. 163—170.
- Sheremetyeva S. and S. Nirenburg. 1996. Interactive Knowledge Elicitation in a Patent Expert's Workstation. *IEEE Computer*. Vol. 7. 1996.
- Sheremetyeva S. 2003. Towards Designing Natural Language Interfaces. *Proceedings of the 4th International Conference "Computational Linguistics and Intelligent Text Processing"* Mexico City, Mexico, February, 2003. P. 16—22.

An MTT-Based Parser for German

Vitali Specht

Intelligent Systems Institute, University of Stuttgart
vitali.specht@iis.uni-stuttgart.de

The paper describes an MTT-based parser for German. The parser analyses German sentences to the surface-syntactical level. The parsing process is divided into three stages. In the first stage, sentences are segmented into topological fields. In the second stage, nominal and prepositional phrases are recognized in the so far received fields, before, in the third stage, the full parse tree is built.

Синтаксический анализатор для немецкого языка на основе МСТ. В данной работе описывается программа для синтаксического анализа немецких предложений, основанная на модели «Смысл-Текст». Анализ осуществляется до поверхностно-синтаксического уровня и состоит из трёх этапов. На первом этапе предложение разбивается на топологические сегменты. На втором этапе проводится распознавание фраз внутри сегментов. На третьем этапе выстраивается синтаксическая структура предложения.

1. Introduction

The Meaning-Text model (MTM) has so far been applied mainly to text generation. The task of parsing has been less studied, cf. (Apresian et al., 2003; Nasr, 2000). In this paper, we present an MTT-based parser for German. The parser analyses German sentences from the surface-morphological (SMorph) to the surface-syntactical (SSynt) level. Our approach primarily differs from the «classical» Meaning-Text model in two points:

1) Following the proposal of Gerdes and Kahane (to appear), we use a topological representation according to the *Topological Field Model (TFM)* of German instead of the deep-morphological (DMorph) representation of MTT.

2) For the mapping between the representation levels, we use a network-based component instead of a rule-based one.

As pointed out by Gerdes and Kahane (to appear), there are two kinds of linguistic constituents: syntactic and topological ones. Syntactic constituents represent the syntactic structure of the sentence, which primarily

expresses the dependency relations between the nodes in the dependency tree. Topological constituents represent the topological structure of the sentence, which primarily expresses the order and the grouping of the words in the sentence. The syntactic structure is closer to the deep representation of the sentence (*meaning*), whereas the topological structure is closer to its surface representation (*text*). Thus, the topological structure can be used for the representation of the sentence on the deep-morphological (DMorph) level.

In our approach, we have adapted the constituency-based PEP-parser (Klatt, 2005) to the requirements of the MTT. The PEP-parser creates constituency structures of the sentences according to the X-bar-theory. The parsing is carried out by the application of deterministic transition networks, which caused the choice of the network-based formalism for the mapping between the representation levels in our approach.

The structure of the remainder of the paper is as follows. Section 2 briefly introduces the Topological Field Model (TFM) of German. Section 3 presents the analysing technique *Pattern-matching Easy-first Planning (PEP)*, which underlies the PEP-parser, as well as some modifications, that we have made to the parser, making it possible to produce MTT-conform SSynt-structures. Section 4 describes in more detail our approach to the syntactical analysis of German sentences. Section 5 gives some examples of surface-syntactic relations, which are used for the surface-syntactic representation of German sentences, and describes the work done so far. Finally, Section 6 draws a conclusion and gives an outlook for worthwhile extensions of the parser.

2. Topological Fields for Modelling Sentence Structure

Three different clause types are distinguished in German depending on the position of the finite verb:

- 1) verb-first-clause (V1) — as in yes/no-questions, cf. (1);
- 2) verb-second-clause (V2) — as in main clauses, cf. (2);
- 3) verb-last-clause (VL) — as in subordinated clauses, cf. (3).

- (1) **Hast** du das Buch bereits gelesen?
lit. 'Have you the book already read?'
- (2) Ich **habe** das Buch bereits gelesen.
lit. 'I have the book already read.'
- (3) Er hat gesagt, dass er das Buch bereits gelesen **hat**.
lit. 'He has said, that he the book already read has.'

However, all three clause types are based on the same structure, which is known as *Topological Field Model (TFM)* of German (Bech, 1955 [1983]; Höhle, 1986). According to this model, every German clause contains the so called *Satzklammer (sentence bracket, SK)*, consisting of the left (*left bracket, LK*) and the right (*right bracket, RK*) part. SK splits up the clause into three fields called *Vorfeld (top field, VF)*, *Mittelfeld (middle field, MF)* and *Nachfeld (rest field, NF)*. The fields are linearly ordered:

VF	LK	MF	RK	NF

Table 1 (cf. Wöllstein-Leisten et al., 1997:53) shows the lexical instantiations of these fields, depending on the three possible clause types. The **bold** lexical units are obligatory, the remaining ones — facultative. VF and NF can contain at most one constituent. LK contains either the finite verb (VFIN) or a subordinate conjunction. RK contains the entire verbal complex (VC) or a discontinuous VC without the finite verb, if the latter is already placed in LK. MF can contain an arbitrary number of constituents. The TFM is built recursively, i.e. VF, MF and NF can contain embedded clauses, which are again partitioned into the topological fields.

	VF	LK	MF	RK	NF
V1		VFIN	Constituents	VINFs	Constituent
V2	Constituent	VFIN	Constituents	VINFs	Constituent
VE		Conjunction	Constituents	VINFs VFIN	Constituent

Table 1. Lexical instantiations of the TFM-fields

TFM was extended by the so called *Satzrahmen (sentence frame, SR)*, consisting of two parts: *Satzanfangsrahmen (sentence top frame, SAR)* and *Satzenderahmen (sentence bottom frame, SER)*, cf. (Rehbein, 1992). The two fields contain conjunctions and punctuation marks. The SER of the preceding clause is at the same time the SAR of the succeeding clause. Topological structures for the sentences (1)—(3) are shown in table 2.

SAR	VF	LK	MF	RK	SER	NF
		Hast	du das Buch bereits	gelesen	?	
	Ich	habe	das Buch bereits	gelesen	.	
	Er	hat		gesagt	,	
,		dass	er das Buch bereits	gelesen hat	.	

Table 2. Examples of topological structures

3. The Analysing Technique PEP

We adapted the constituency-based PEP-parser (Klatt, 2005) to the requirements of the MTT. The PEP-parser is implemented in CommonLISP based on the analysis technique *Pattern-matching Easy-first Planning (PEP)* (Klatt, 1997; 2005). In contrast to mainstream parsing techniques, the PEP-parser does not build the parse tree systematically (e.g. from bottom to top). Rather, it tries to recognize those structure elements, that it can, making easy decisions first.

The analysis process is driven by traversing of transition networks (TN) deterministically. A TN is defined by several states, which are linked by directed arcs. Every arc is associated with a so-called network function (NWF). The two most important NWFs are: *nb* for recognizing adjacent elements like DAS and BUCH in (4) and *seq* for recognizing non-adjacent patterns with a left and right border element like HAT and GELESEN in (4).

- (4) Er hat das Buch bereits gelesen.
lit. *'He has the book already read.'*

It is further possible to assign to the arcs so-called syntactic operations, which are used for building new constructs. The PEP-parser already contains a set of predefined syntactic operations according to the X-bar-theory. In order to adapt the parser, we have defined a new set of syntactic operations. Each syntactic operation in our set corresponds to a surface-syntactic relation (SSyntRel) and connects two elements of the dependency tree (i.e. nodes or subtrees) with this SSyntRel.

For example, applying the NWF *nb* in (5) to the sentence in (4), the parser recognizes the nominal phrase (NP) 'DAS BUCH'. The NWF *nb* looks for a determiner (*m-cat det*) left adjacent to a noun (*m-cat n*). Both elements must agree in case, gender and number (*:agree (kasus genus num)*). If such pattern exists, the syntactic operation *det-merge* is called. This syntactic operation connects the elements with the SSyntRel *determinative*, creating a new subtree. The governing element (the head) of the relation is the right neighbour, i.e. the noun (*:head right*).

- (5) (nb ((m-cat det)) ((m-cat n))
:agree (kasus genus num)
:syn-op det-merge
:head right)

4. The Parsing Process

The parsing process consists of three main stages. In the first stage, sentences are segmented into topological fields. In the second stage, simple nominal and prepositional phrases are identified in the topological fields. In the third stage, the dependency structure of the entire sentence is built. In what follows, we describe these three stages in detail.

a. Stage 1: TFM-Segmentation

In stage 1, the parser recognizes SK for each clause and analyses the verbal complex in RK as well as the element in LK (which is possible a finite verb or a subordinate conjunction, see table 1). The stage consists of several steps, which are briefly described below.

In the first step, multi-word expressions¹ (MWE) are processed, cf. MWE 'FIX UND FERTIG' in (6).

- (6) Nach dem Spiel war er fix und fertig.
lit. 'After the game he was bushed.'

In the second step, word coordinations in LK and RK are detected. According to MTT, such coordinations are connected with *coordinative* and *coordinate-conjunctive* SSyntRels.

- (7) Er hat gesungen und gelacht.
lit. 'He has sung and laughed.'
[SINGEN_{ppf}—**coord**→UND—**coord-conj**→LACHEN_{ppf}]
- (8) Er singt und lacht.
lit. 'He sings and laughs.'
[SINGEN_{pres,3,sg}—**coord**→UND—**coord-conj**→LACHEN_{pres,3,sg}]

In the next step, the verbal complex in RK is recognized. According to (Mel'čuk and Pertsov, 1987), particular verbs within complex verb forms, like perfect or passive, have to be connected with the SSyntRel *auxiliary*. We distinguish in case of verbal complexes several different SSyntRels, e.g. *auxiliary*, *modal*, *infinitive*, in order to emphasize distinct syntactic phenomena. Some examples of these relations are given in (9)–(11).

¹ Multi-word expressions are phraseological units, which consist of several words, but semantically substitute a single word. Therefore, they are considered as a single unit in the rest of the analysing process.

- (9) Er sagt, dass er das Buch bereits gelesen hat.
lit. 'He says, that he the book already read has.'
[HABEN_{pres, 3, sg}—**aux**→LESEN_{ppf}]
- (10) Er behauptet, dass er fliegen kann.
lit. 'He claims, that he fly can.'
[KÖNNEN_{pres, 3, sg}—**modal**→FLIEGEN_{ppf}]
- (11) Er ist schwimmen gegangen.
lit. 'He ist [to] swim gone.'
[SEIN_{pres, 3, sg}—**aux**→GEHEN_{ppf}—**inf**→SCHWIMMEN_{inf}]

In the fourth step, the parser recognizes SK, i.e. the verbal complex, which has been recognized in the previous step, is attached to the element in LK. In case of the finite verb in LK, the same verbal SSyntRels as in the previous step are used. In case of a subordinate conjunction, we use the *subordinate-conjunctive* SSyntRel, cf. (12)—(13).

- (12) Er hat das Buch bereits gelesen.
lit. 'He has the book already read.'
[HABEN_{pres, 3, sg}—**aux**→LESEN_{ppf}]
- (13) Er sagt, dass er das Buch bereits gelesen hat.
lit. 'He says, that he the book already read has.'
[DASS—**sub-conj**→HABEN_{pres, 3, sg}]

After this step, each clause can be segmented into *non-recursive* topological fields. For example, two topological structures in (14) are created for the sentence in (13), but it is not known yet, that the second clause is placed in NF of the first one.

- (14) [_{VF} Er] [_{LK} sagt]
[_{LK} dass] [_{MF} er das Buch bereits] [_{RK} gelesen hat]

In the final step of this stage, government patterns of the lexical verbs are integrated.

b. Stage 2: Recognition of simple nominal and prepositional phrases

After the segmentation of the sentence into topological fields, simple nominal (NP) and prepositional (PP) phrases are identified in VF, MF and NF of each clause. Using each noun as the starting point of the search, the parser looks for its neighbour elements, which agree with the noun in case, number and gender. These elements are connected with the noun using phrasal SSyntRels such as *determinative*, *modificative*, *prepositional* etc.

The recognition of the NPs and PPs is carried out in three main steps. In the first step, we recognize structures, which consist of adjacent elements like ‘DIESES BUCH’ and ‘DES NACHBARN’ in (15). The result is the relations in (16).

- (15) Dieses Buch haben ihr die drei netten kleinen Kinder des Nachbarn geschenkt.

lit. ‘*This book have her the three nice little children of the neighbour presented.*’

- (16) [DIESES_{acc, neut, sg}—**det**→BUCH_{acc, neut, sg}]
 [DER_{gen, masc, sg}—**det**→NACHBAR_{gen, masc, sg}]

Phrases with attributive adjectives, like ‘DIE DREI NETTEN KLEINEN KINDER’ in (15), are recognized in the next step, cf. (17).

- (17) [KIND_{nom, neut, pl}—**det**→DIE_{nom, neut, pl}]
 [KIND_{nom, neut, pl}—**quant**→DREI]
 [KIND_{nom, neut, pl}—**modif**→NETT_{nom, neut, pl}]
 [KIND_{nom, neut, pl}—**modif**→KLEIN_{nom, neut, pl}]

Finally, we recognize phrases with possessive elements, attaching in sentence (15) the possessive element ‘DES NACHBARN’ to the head of the subtree ‘DIE DREI NETTEN KLEINEN KINDER’, i.e. KINDER.

c. Stage 3: The Construction of the Dependency Structure

The dependency structure of the entire sentence is created in two steps.

In the first step, the dependency structure for each clause is created. The NPs and PPs, which have been recognized in the stage 2, are attached to the verb according to the government pattern. We use actantial SSyntRels like *subjectival*, *direct-objectival*, *prepositional-objectival* etc., if these phrases are actants of the verb, or adverbial SSyntRels, if they are modifiers of the verb. Note, that in case of a verbal complex, the subject is attached to the finite verb, because both agree in case, number and person, whereas the objects are attached to the lexical verb.

In the second step, the dependency structure of the entire sentence is built, i.e. the subtrees of the clauses are integrated in the global structure.

5. The Surface-Syntactic Relations

We have chosen the names of the SSyntRels according to the classification of SSyntRels for English in (Mel’čuk and Pertsov, 1987). How-

ever, as pointed out in literature, cf. (Mel'čuk and Pertsov, 1987: 98—99; Iordanskaja and Mel'čuk, 2000), the actantial SSyntRels (i.e. *predicative*, *1st*, *2nd*, *3rd* and *4th-completive* and *agentive* SSyntRels) in this classification follow too closely the deep-syntactic relations (DSyntRels) and often correspond with them in one-to-one. In recent works in MTT, cf. (Iordanskaja and Mel'čuk, 2000; Nasr, 2000), as well as in this work, the *subjective*, *direct-objective*, *indirect-objective* etc. SSyntRels are used instead of the actantial SSyntRels proposed by (Mel'čuk and Pertsov, 1987), cf. (18).

- (18) Hans hat seiner Frau zum Geburtstag einen Wagen geschenkt.
lit. 'Hans *has to-his wife to the birthday a car presented.*'
[HABEN_{pres, 3, sg} — **subj** → HANS_{nom, mask, sg}]
[SCHENKEN_{ppf} — **dir-obj** → WAGEN_{acc, mask, sg}]
[SCHENKEN_{ppf} — **indir-obj** → FRAU_{dat, fem, sg}]
[SCHENKEN_{ppf} — **prep-obj** → ZU — **prepositional** → GEBURTSTAG_{dat, mask, sg}]

We added the SSyntRel *genitive-objective*, in order to describe the relations between verbs and their genitive complements, cf. (19). The *modal* SSyntRel describes the construction «modal verb + infinitive», cf. (20).

- (19) Hans bezichtigte die Frau des Diebstahls.
lit. 'Hans *accused the woman of theft.*'
[BEZICHTIGEN_{past, 3, sg} — **gen-obj** → DIEBSTAHL_{gen, masc, sg}]
(20) Sie müssen das nächste Spiel gewinnen.
lit. 'They *have to win the next game.*'
[MÜSSEN_{pres, 3, pl} — **modal** → GEWINNEN_{inf}]

To describe the relations to the complements of copula verbs such as SEIN (= TO BE) or SCHEINEN (= TO SEEM) in (21) as well as non-verbal predicates, e.g. parametric adjectives such as ALT (= OLD) in (22) or adverbs such as GENUG (= ENOUGH) in (23), we use the *completive* SSyntRel.

- (21) Sie war sprachlos und schien erschrocken.
lit. 'She *was mute and seemed scared.*'
[SEIN_{past, 3, sg} — **compl** → SPRACHLOS]
[SCHEINEN_{past, 3, sg} — **compl** → ERSCHROCKEN]
(22) Sein Sohn ist schon sieben Jahre alt.
lit. 'His *son is already seven years old.*'
[ALT_{past, 3, sg} — **compl** → JAHR_{nom, 3, pl} — **quant** → SIEBEN]
(23) Die Tasche ist leicht genug zum Tragen.
lit. 'The *bag is light enough to carry.*'
[GENUG — **compl** → LEICHT]

Further, we use the *zu-infinitiv* SSyntRel to describe the relation between the verb particle ZU (=TO) and the verb in an infinitive construction, cf. (24), instead of the *prepositional* SSyntRel proposed in (Mel'čuk and Pertsov, 1987).

- (24) Sie versprachen das Spiel zu gewinnen.
 lit. 'They *promised to win the game.*'
 [VERSPRECHEN_{past, 3, pl} — **dir-obj** → ZU — **zu-inf** → GEWINNEN_{inf}]

About 35 SSyntRels have so far been implemented for the surface-syntactic representation of German sentences. By the implementation, we first concentrated on common SSyntRels like actantial and phrasal ones. SSyntRels, that are rather uncommon, e.g. *subjective-copredicative* (cf. Mel'čuk and Pertsov, 1987: 102—104) are not implemented yet.

6. Conclusion and Future Work

The segmentation of the sentence into topological fields proved to be very helpful in respect of simplifying the parsing process. The segmentation allows us to recognize the lexical verb in an early stage of the parsing process. Together with the usage of a valency lexicon, this information heavily reduces the search space for the ongoing analysis. Processing complex sentences, we can easily segment each clause into TFM-fields and parse each field separately, combining them in a later step to the global syntactic structure. Further, we can make use of other TFM properties, such as the property, that only one constituent may occur in VF. This information is helpful, for example, for the identification of multi-word expressions. Some other benefits of the TFM-segmentation are described in (Klatt, 2004).

We are currently working on the implementation of the remaining SSyntRels. The next step to be done is the extension of the parser to the deep-syntactical (DSynt) level. The main problems to deal with in this step are recognition of lexical functions and co-reference resolution.

References

- Apresian, J., I. Boguslavsky, L. Iomdin, A. Lazursky, V. Sannikov, V. Sizov and L. Tsinman (2003). ETAP-3 Linguistic Processor: a Full-Fledged NLP Implementation of the MTT. In *First International Conference on Meaning-Text Theory*. Paris (France), Ecole Normale Supérieure, 2003. P. 279—288.

- Bech, G. (1955). Studien über das deutsche Verbum infinitum. Copenhagen, 2005. Reprinted Niemeyer, Tübingen, 1983.
- Gerdes, K. and S. Kahane (to appear). Phrasing it differently. In Leo Wanner (ed), *Recent works in the Meaning-Text theory in honor of I. Mel'čuk*.
- Höhle, T. (1986). Der Begriff Mittelfeld. Anmerkungen über die Theorie der topologischen Felder. In W. Weiss, Wiegand E. H. and M. Reis (eds), *Textlinguistik contra Stilistik/Wortschatz und Wörterbuch/Grammatische oder pragmatische Organisation von Rede*. Niemeyer, Tübingen, 1986
- Iordanskaja, L. and I. Mel'čuk (2000). The Notion of Surface-Syntactic Relation Revisited (Valence-Controlled Surface-Syntactic Relations in French). In L. L. Iomdin, L. P. Krysin (eds), *Slovo v tekste i v slovare. Sbornik statej k semidesjatiletiju Ju. D. Apresjana*, Moskva: Jazyki russkoj kul'tury, 2000. P. 391—433.
- Klatt, S. (1997). Pattern-matching Easy-first Planning. In Alice Drewery, Kruijff Geert-Jan M., and Richard Zuber (eds), *Proceedings of the 2nd ESSLLI Student Session*, Aix-en-Provence (France). 9th European Summer School in Logic, Language and Information, 1997.
- Klatt, S. (2004). Segmenting Real-Life Sentences into Topological Fields — or Better Parsing and Other NLP Tasks, In *Proceedings of the KONVENS 2004. 7. Konferenz zur Verarbeitung natürlicher Sprache*, Vienna (Austria), Ernst Buchberger, 2004.
- Klatt, S. (2005). Kombinierbare Textanalyseverfahren für die Korpusannotation und Informationsextraktion. Shaker Verlag, Aachen, 2005.
- Mel'čuk, I. and N. Pertsov (1987). Surface-syntax of English. John Benjamins Publishing Company, Amsterdam, Philadelphia, 1987.
- Nasr, A. (2000). A Generative Approach to Parsing in the Framework of the Meaning-Text Theory. In Alain Polguere and Leo Wanner (eds). *Selected Aspects of Dependency Theory*. Benjamins Academic Publishers, 2000.
- Rehbein, J. (1992). Zur Wortstellung in komplexen deutschen Sätzen. In Ludger Hoffmann (ed), *Deutsche Syntax: Ansichten und Aussichten*. Walter de Gruyter, Berlin, New York, 1992.
- Wöllstein-Leisten, A., A. Heilmann, P. Stepan, and S. Vikner (1997). Deutsche Satzstruktur: Grundlagen der syntaktischen Analyse. Stauffenburg, Tübingen, 1997.

Compiling a «Classical» Explanatory Combinatorial Lexicographic Description into a Relational Database

Jacques Steinlin Sylvain Kahane Alain Polguère

LaTTice — Université de Paris 7 OLST —
Université de Montréal,
jsteinli@linguist.jussieu.fr, sk@ccr.jussieu.fr alain.polguere@umontreal.ca

This paper deals with the processing of a formal electronic dictionary of French centered around the description of semantic derivations and collocations. The purpose of our project was to obtain a more flexible format for this dictionary, the DiCo, which has a traditional structure. By compiling the DiCo, we get an object model which, in turn, can be stored in an SQL database. We show how the object model allows us to more easily access the data and gives us a new perspective on the structure of dictionary. This latter can be completely reorganized following new axes like lexical function links.

1. Introduction

The present work is part of a lexicographic project that targets an electronic dictionary of French centered on semantic derivations and collocations. The dictionary is developed within the framework of the Explanatory Combinatorial Lexicology, the lexical module of the Meaning-Text Theory ((Mel'čuk, 1974), (Žolkovskij & Mel'čuk, 1965), (Mel'čuk *et al.*, 1995)) and the project is supervised by Igor Mel'čuk and Alain Polguère at the OLST-Université de Montréal. The electronic dictionary, the DiCo, is designed to serve as a generic lexicographic description from which one can produce i) «computable» databases such as the one presented here and ii) «general public» descriptions such as the *Lexique Actif du Français* or *LAF* (Polguère, 2000). The DiCo possesses a rather traditional structure, inspired by paper dictionaries, and adopted by numerous electronic dictionaries (TLFi (Dendien & Pierrel, 2003), eEH (Arregi *et al.*, 2003)). The entries of the dictionary are vocables (including idioms). Every vocable is a set of lexical units corresponding to its different

senses. Our goal was to propose for the DiCo a richer and more flexible architecture to improve the access to the data. On one hand, this architecture aims at improving the query possibilities of the DiCo for human users and at making the lexicographic information exploitable by natural language processing systems (especially in natural language generation). On the other hand, it should facilitate all the tasks of updating and development of the dictionary, that is the lexicographic tasks. For that reason, we chose to parse the existing dictionary, the DiCo, and then to feed a relational database via an object representation. In this paper, we present the result of this modeling, the DiCobjet, and show the advantages of such a structure from a theoretical point of view as well as from a practical one. This structuring can be compared with the one which is used in WordNet (Fellbaum, 1998). However, as we shall see, the network resulting from the analysis connects not only lexical items by means of lexical function relations, but also encodes all lexicographic information that is associated with each lexical unit (such as, for example, the subcategorization frame). We first introduce the DiCo as it is developed by lexicographers, then we describe its compiled structure and its representation. We end with a discussion of the theoretical and practical advantages obtained by our modelization.

2. Overview of the DiCo

As we said, the DiCo adopts a structure close to the structure of standard paper dictionaries, that is, a structuring made up of lexicographic entries. This structuring is required by the lexicographer, who wants to have a global vision of the bulk of information related to each lexical unit. The description inside an entry consists of nine main fields, each corresponding to a specific type of information, as we can see on Figure 1.

Field	Value
word	ENGUEULADE
lexicographic number	1
grammatical properties	nom, fém, «fam»
semantic label	communication langagière
semantic formula	~ DE L'individu X [VISANT L'individu Y] POUR LE fait Z
subcategorization frame table	X = I = de N, A-poss Y = II = — — Z = III = Prep-pour N Prep-pour = { _à propos de_, _au sujet de_, pour }, pour V-inf-passé

synonymy	{QSyn} «fam» savon; «soutenu» admonestation, remontrance, réprimande; blâme
lexical functions	{QAnti} compliment, félicitation; flatterie {VO} engueuler {Magn} belle, bonne, sacrée antepos < majeure postpos {Oper213} essayer [ART ~ _de la part de _ N=X Prep-pour N=Z] {Oper23} subir [ART ~ Prep-pour N=Z]
Phraseology	

Fig. 1. Partial description of the lexical unit ENGUEULADE₁
(= ‘argument’, ‘bawling out’)

The first two fields contain the name of the vocable and the lexicographic number of the lexical unit. The **grammatical properties** field records the part of speech of the lexical unit as well as other information relative to its cooccurrence: stylistic label (ex: «fam»), graphic variants, inflection constraints (ex: «no plural»), and so on. The **semantic label** field position the lexical unit in a semantic hierarchy used to account for the central semantic behavior of lexical units (for a detailed description of the DiCo semantic labeling, see (Polguère, 2003)). The **semantic formula** field enumerates the semantic actants of the lexical item and possibly supplies them with a semantic label. The **subcategorization frame table** describe how semantic actants are to be expressed in structures that are syntactically controlled by the lexical unit. **Synonymy** and **Lexical Functions** fields describe lexical function links controlled by the lexical unit. A lexical function’s name is written between braces and is followed by a value list. Lexical functions like **Magn** or **Oper_i** allow to describe the collocations in which the lexical unit occurs: *une belle engueulade*, *une sacrée engueulade* or *subir une engueulade*. Finally the **phraseology** field lists all the full idioms that formally contain the lexical unit. It is empty here, ENGUEULADE₁ having no associated full idioms. In practice, a relational mono-table database is used for the storage of the DiCo. Every entry is a record, each row corresponding to a field inside the record. Beyond this first level of structuring, we find a second level of structuring, which is syntactic. Every field is written according to some syntactic and typographic conventions. Thus, it becomes possible to identify automatically the keys of the description. We were able to design a compiler to translate automatically the DiCo into the DiCobjet. A similar work was made within the framework of the project Papillon for the translation of the DiCo in a XML format ((Lapalme & Sérasset, 2003)).

3. Object Modeling of the DiCo

3.1. General overview of the object modeling

As we can see, the fields described in the previous section contain lists of information items which we can split into fields and so on. This is the compiler purpose. The content of the DiCo is fed into the compiler as a tab-separated text. The compiler parses the file and produces a semantic representation of the DiCo, the DiCobjet¹. One of the objectives that we set ourselves in compiling the DiCo was to obtain the most detailed representation as possible. But in fact, the modeling only consists in bringing to light (by clarifying it) the structure of the DiCo. So, by examining again figure 1, we notice that the lexical function field contains a list of lexical functions. A lexical function is supplied with a value list linked to the lexical function formula (for example, **Oper**₂₁₃). Each of these values (a support verb in this case) can be organized in turn in fields: the value itself, its style label («soutenu»), its subcategorization frame, a constraint (for example `postpos` for an adjective), and so on.

The work of the compiler thus consists in identifying the atoms of information and organizing them in an adequate data structure. The data structure representing the DiCo is the DiCobjet (see the diagram, figure 2). We represented the classes (that is, the definition of objects) by boxes. The upper part of the box is reserved for the name of the class. The lower part enumerates the attributes which are of primitive type (integer, boolean, character strings). Other attributes, which have classes for value, are represented by the relations between the classes. These relations, although directed, are bidirectional. We call them aggregate relations. Such a relation means that the class being on the side of the diamond owns one or several classes as attributes. On the other side of the relation, we wrote the number of objects expected in the aggregation. Figure 2 must be read in the following way: a DiCobjet is a set (of size superior or equal to 0) of Word objects. The Word class is characterized by attributes (only the attribute «name» being compulsory) and a set of LexicalUnit objects (the set must include at least one element). A LexicalUnit is described in turn as a container of collections of complex objects: SemanticLabelFormula, a set of semantic actants (SemActantVar), a set of tables of subcategori-

¹ *Semantic representation* is a computer science concept. From the computer point of view, the input is just a character string. In compiling it, the character string is translated into objects which have a semantic value. For example, the input string {Magn} becomes a LFFormula object (and therefore it has some specific properties).



Fig. 2. DiCo Object Diagram

zation frame (SubCatFrameTable), a set of Phrasemes, one grammatical property (GramProp, which is a set of grammatical properties primes) and LFRubric. LFRubric, besides being defined by a name, includes a set of objects of type LF, this last class consisting of the grouping of LFFormulas (what we usually call a Lexical Function), of a Gloss (the paraphrase in natural language of the lexical function) and of a set of LFValue (the values returned by the lexical function applied to the described lexical unit). Finally, LFFormula is a tree structure combining some LFPrimitive. The class LFValue contains many attributes. Among these, the value strictly speaking (for example, *flatterie*) is a string. This modeling is not satisfactory as far as the value is in fact a more complex object and might be treated as a LexicalUnit. In the case of a complex lexical item like an expression, we give its syntactic composition with a dependency tree. We have to emphasize that the aggregate relation does not foresee the orientation of the link between objects. If SubCatFrameTable is clearly an attribute of a LexicalUnit, it is also possible to go back up from a given SubCatFrameTable to the LexicalItem objects which possess it. In the conclusion we will return to the repercussions of this property of the model on the conception we have of the dictionary.

3.2. Exploitation

The advantages of object modeling are numerous. It becomes possible to do sophisticated queries and to filter the results of these to keep only the relevant information (instead of the complete records). For example, we can retrieve the list of the lexical units having at least two actants whose first one is realized by the surface forms *de N*, *A-poss* or the lexical items having at once two semantic actants and one *Real@*. To allow this enhancement of the query power, it is necessary to add the possibility of making queries under other viewpoints than the one imposed by the structuring in words. We can consider, for example, the dictionary from the lexical functions point of view. The current version of the DiCo gives 143 different values for the lexical function **Oper₁** (Figure 3 gives a sample). We get back also 48 different lexical functions returning the value *donner* (see figure 4).

So, as it was requested by (Grossmann & Tutin, 2003), we can now try to discover regularities among the values since we can now group them by the semantic label of the lexical unit they are linked to. For example, we can see that many lexical units labeled with the semantic label *sentiment (feeling)* have the lexical function **IncepPredPlus** returning the value *augmenter (to increase)*. Similarly, we can compare values returned

q _{ti} y	Value	q _{ti} y	Value	q _{ti} y	Value	q _{ti} y	Value
64	avoir	14	pousser	1	aller	1	être pris
42	être	11	émettre	1	battre	1	jauger
27	faire	10	constituer	1	célébrer	1	occuper
20	posséder	9	_faire preuve_	1	comprendre	1	partir
17	ressentir	8	se trouver	1	couvrir	1	peupler
15	éprouver	(...)	(...)	1	être (présent)	1	proférer

 Fig. 3. **Oper**₁ values in decreasing frequency

q _{ti} y	LF	q _{ti} y	LF	q _{ti} y	LF
10	{CausFunc1}	3	{Liqu1Oper1}	2	{Labreal21}
7	{Oper12}	3	{Son}	2	{Liqu1Real1}
6	{Oper1}	2	{Caus1Func0}	2	{Real31-I}
5	{Oper2}	2	{Caus de nouveau Func1}	2	{Real31-II}
4	{Oper13}	2	{Caus1Func0}	2	{Realo-II}
4	{Real12}	2	{Fact1}	1	{PredAble2}
3	{Caus3Func0}	1	{Fact21-production}		

 Fig. 4. Lexical Functions returning the lexical unit *donner* in the current version of the DiCo

by the LF's **Oper**₁ and **Real**₁ or **Fact**₀ and **Real**₁. It is also possible to check whether a lexical unit is, at the same time, an argument of many LF's such as **Oper**₁ and **Real**₁².

Further applications could possibly be the use of the DiCobjet API to program procedures of planification for new entries based on related entries. Not only the trivial **QSyn** relation could be used, but also lexical units sharing the same semantic label.

3.3. Practical and theoretical contribution of object modeling

The DiCobjet was automatically extracted from the DiCo. This was possible thanks to the usage of strict conventions in the editing of the DiCo (only some minor adjustments were necessary to reduce the existing ambiguities). Notice that one of the main requirements of the lexicographers is to be able to continue to develop the DiCo with its initial format (or an equivalent one). Indeed, as we said earlier, the compilation of a dictionary cannot amount to a blind filling of predefined fields and the

² In the DiCo, 229 lexical units have **Oper**₁ values while 99 have **Real**₁ values; but only 22 have both.

lexicographer must be able to read a dictionary entry taken as a whole. As is expected, the object modeling of the dictionary increases the access to the data of the DiCo, by allowing varied queries. More crucially the result returned by these queries is not any more inevitably the entry, it can be a value of lexical function or a set of subcategorization frames. Besides the query and the exploitation, the modeling facilitates the revision of the DiCo by enabling an easy check of the homogeneity of every type of information. It also facilitates its development by the processing on the data and by allowing for example to combine the information of several entries to build the skeleton of a new entry. The object modeling of the dictionary increases not only the possibilities of data mining. The most important and least expected result from the project is that our vision of the dictionary changed. For example, at the surface level, the dictionary is no longer a flat collection of entries. The object model can be taken in fact by any end and the DiCo can thus be completely turned upside down. We can get into the DiCo by the FL object and produce for every lexical function an entry giving the lists of the values for every keyword. We have then a much more semantic vision of the dictionary, where we see how senses like the intensification, the causation, the realization, etc. are realized in very different ways. We do not have to deal any more with a dictionary of 2000 entries as in the beginning, but with a dictionary of several dozens thousand entries. The online dictionary can be reached at the following address: <http://www.olst.umontreal.ca/dicouebe/>.

References

- Arregi X., Arriola J., Artola X., Diaz de Ilarraza A., Garcia E., V. L., Sarasola K., Soroa A. & Uria L. (2003). Semiautomatic conversion of the euskal hitzegia basque dictionary to a queryable electronic form. *Traitement Automatique des langues, TAL*, 44 n°2, 107—124.
- Dendien J. & Pierrel J. M. (2003). Le trésor de la langue française informatisé, un exemple d'informatisation d'un dictionnaire de langue de référence. *Traitement Automatique des langues, TAL*, 44 n 2, 11—37.
- C. Fellbaum, Ed. (1998). *WordNet, An Electronic Lexical Database*. MIT Press.
- Grossmann F. & Tutin A. (2003). Quelques pistes pour le traitement des collocations. In F. Grossmann & A. Tutin, Eds., *Les collocations, analyse et traitements*. Revue Française de linguistique appliquée.
- Lapalme G. & Serasset G. (2003). Batch creation of Papillon entries from DiCo. In *Workshop Papillon*.
- Mel'čuk I. (1974). Opyt teorii lingvisticheskikh modelej «smysl ↔ tekst». *semantika, sintaksis. Nauka*.

- Mel'čuk I., Clas A. & Polguère A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Duculot.
- Polguère A. (2000). Towards a theoretically-motivated general public dictionary of semantic derivations and collocations for french. In *Proceedings of EURALEX 2000*.
- Polguère A. (2003). Étiquetage sémantique des lexies du DiCo. *Traitement Automatique des langues, TAL*, 44 n°2.
- Žolkovskij A. & Mel'čuk I. (1965). O vozmožnom metode i instrumentax semantičeskogo sinteza. *Naučno-texničeskaja informacija*, 6, 23—28.

Противопоставительная структура как средство семантической интерпретации отрицания

В. М. Труб

Институт украинского языка Национальной Академии наук Украины
volodya@gilan.uar.net

Contrastive structure as a means of semantic interpretation of negation. The paper deals with a variety of aspects of interaction of negation with other parts of the utterance. The contrastive structure like *not P, but Q* is considered to be a universal means of interpretation of negation in the function of negative evaluation of the truth of the utterance. The different types of implicit oppositions, peculiar to negative sentences as well as to the meanings of beginning and cessation are being examined.

В работе рассматриваются аспекты взаимодействия отрицания с другими компонентами высказывания. В качестве универсального средства семантической интерпретации отрицания как негативной оценки истинности предложения рассматривается противопоставительная структура вида 'не P, а Q'. Анализируются различные типы имплицитных противопоставлений, присущих отрицательным предложениям, а также значения начинательности и прекращения.

В процессе исследования традиционной логико-лингвистической проблемы отрицания обнаруживается целый ряд интересных его свойств. В данной работе будут рассмотрены некоторые аспекты взаимодействия отрицания с компонентами предложения, попадающими в сферу его действия.

1. Прежде всего отметим, что одна из важных особенностей отрицания иногда проявляется в различии идиоматической сочетаемости того же самого слова в утвердительном и отрицательном контексте. В качестве примера можно привести различия в обозначении семантического параметра Magn [6, 7] для ключевых слов *осадки* и *возражения*. Ср.: *Завтра в Москве без существенных осадков*, но плохо ? ...*без значительных (больших) осадков*. С другой стороны *Прошли большие (значительные...) осадки*, но плохо ? *Прошли суще-*

ственные осадки. Ср. также *Особых возражений (изменений) нет* при неприемлемости **У него возникли особые возражения*; **Произошли особые изменения* (надо: *У него возникли серьёзные (большие) возражения*; *Произошли серьёзные (большие, существенные...) изменения*).

Примечательно также, что некоторые слова под отрицанием меняют своё значение. Таким образом, разные значения некоторых омонимов оказываются в отношении дополнительной дистрибуции: одно значение реализуется только в утвердительном контексте, другое — только в отрицательном. Ср., например, глагол *удосужиться*. Фраза *Х удосужился Р* означает, что Х осуществил такое Р, которое характеризуется общепризнанной заведомо негативной оценкой. Напротив, *Х не удосужился (даже) Р* (ср. *Он не удосужился (даже) помыть посуду*) предполагает, что Р не имеет каких-либо негативных коннотаций, а негативно оценивается именно невыполнение Р.

Утвердительная форма безличного глагола *приходится* нормативно указывает на вынужденное принятие далеко не оптимального решения — ср: *Приходится действовать в одиночку (самостоятельно)*. Между тем его отрицательная форма, управляющая некоторыми препозитивными ментально-оценочными глаголами, указывает на то, что нет никаких оснований для мнения, которое обозначает соответствующий ментальный предикат: *На это рассчитывать (надеяться) не приходится* ≈ 'То, что есть какие-либо основания для того, чтобы на это рассчитывать (надеяться), неверно'.

2. Как известно, современная методология семантической интерпретации отрицания рассматривает его как негативную оценку истинности некоторой информации, которая уже так или иначе актуализирована в общем поле зрения говорящего и слушающего [11]. Тем самым имеется в виду, что в сферу действия отрицания попадает «готовая» пропозиция, теме которой уже сопоставлена определенная рема. Функция отрицания состоит в указании на «неправомерность сопоставления данной ремы соответствующей теме» [3] и в информационной коррекции — сопоставлении исходной теме другой, «правильной» ремы, степень конкретизации которой может варьироваться.

Традиционно принято различать нейтральное отрицание вида *не Р* и противопоставительное (контрастивное) — *не Р, а Q*, где компонент Q нормативно соотносится с «правильной» ремой, которой следует заменить первоначальную рему Р. Мы считаем, что преобразование, манифестируемое последней структурой, является универ-

сальным, т. е. что противопоставительное отрицание отражает сущность любого отрицания — ср. [10]. Тем самым структура *не Р, а Q* является более глубоким уровнем представления структуры *не Р*. Семантическое представление отрицательного предложения должно отражать ту реальную трансформацию, которой подвергается исходное значение Р под воздействием оператора ‘То, что Р, неверно’. Именно эта трансформация отличает отрицательное предложение от утвердительного, в котором определенная рема непосредственно сопоставляется компоненту, являющемуся темой.

В одних случаях сфера действия противопоставительного отрицания может совпадать со сферой действия нейтрального (ср. (1а) *Коля не спит* и (1б) *Коля не спит, а ходит*, где в (1а) и в (1б) в качестве темы выступает компонент ‘вид поведения Коли в обсуждаемом временном интервале’. В других случаях такое совпадение отсутствует — ср. (2) *Коля не спит, а дремлет*, темой которого является ‘степень релаксации организма Коли’.

Мы исходим из того, что на глубинном уровне нейтральное отрицание также имеет противопоставительную структуру, которая, в отличие от поверхностных контрастивных конструкций, является неизменной. Далее мы сосредоточимся на вычленении ряда типов этих глубинных противопоставлений.

1) Отрицание пропозиции, построенной на базе «параметрического» глагола — ср. [1, 3, 8]. В подобных случаях особенность отрицания состоит в том, что хотя глагол и сохраняет свое исходное значение, пропозиция, подвергаемая отрицанию, не тождественна по смыслу результату восприятия предложения, построенного на основе утвердительной формы соответствующего предиката. Так, в (3) *Мешок не весит 50 кг* объектом опровержения является мысль не о том, что мешок весит 50 кг, а о том, что он весит 50 кг или больше — ср. (4) *Вес мешка не менее (минимум) 50 кг; Вес мешка — от 50 кг*. Здесь уместно напомнить, что объектом отрицания является результат акта восприятия фраз типа (4). В данном случае имеется в виду, что опровергается уже готовое, сформированное представление о том, что вес мешка — не менее 50 кг. В то же время сам процесс восприятия той или иной утвердительной пропозиции соотносится с более ранним этапом — актом, в ходе которого теме (здесь — ‘вес мешка’) сопоставляется рема, задающая приблизительное значение данного параметра с четко заданной нижней границей (‘не менее 50 кг’). В этой связи нам представляется не вполне адекватной приведенная в работе [8, с. 11] трактовка, согласно которой исходным утвердительным предложением для (3) является (5)

Мешок весит\ 50 кг (где \ — знак нисходящей интонации). Дело в том, что значение (5), строго говоря, не тождественно (4), которое является исходным для (3). В самом деле, в отличие от (4), в (5) мысль о том, что вес мешка не менее 50 кг предполагается уже актуализированной в общем поле зрения говорящего и слушающего. При этом (5), так же как и (3), выполняет функцию истинностной оценки этой идеи, но оценки положительной: (5) $\approx$ 'То, что мешок весит не менее 50 кг, верно' (ср. [9, с. 367]). Поэтому если считать, что исходным утвердительным предложением для (3) является (5), то следует признать, что (3) негативно верифицирует не идею о том, что вес мешка не менее 50 кг, а ранее высказанную положительную верификацию этой идеи, возникшей еще раньше. В большей мере, чем (5), на роль исходного утвердительного для (3) претендует структура, в которой количественный показатель попадает в позицию контрастной темы [8] — ср. (6) *50 кг мешок весит*. В отличие от (5), (6) допускает, что к моменту его произнесения у слушающего могло и не быть предварительного предположения о минимальном весе мешка, хотя числовое значение данного параметра и было предметом его актуального интереса.

В качестве «правильной» ремы, которая сопоставляется теме 'вес мешка' в результате информационной коррекции, выступает остальная часть шкалы параметра «вес», т. е. та, которая не совпадает с участком шкалы, соотносящимся со значением Р (здесь — 50 кг или больше). Таким образом, (3) $\approx$ 'Вес мешка не 50 кг или больше, а меньше 50 кг'.

В работе [8] рассматриваются также отрицательные предложения, построенные на базе некоторых других глаголов с параметрическим показателем, которые допускают иную («увеличительную») трактовку — ср. (7) *Я не доберусь за полчаса* $\approx$ 'Я доберусь не за полчаса или меньше, а больше, чем за полчаса'. Как убедительно показано Е. В. Падучевой, противопоставление увеличительной vs. уменьшительной интерпретаций связано с различиями в модальностях, с которыми соотносятся глаголы с числовым показателем. Модальность способности / возможности имплицитно уменьшительную трактовку, тогда как модальность уступки / достаточности — увеличительную.

Кроме того, в данной работе сформулировано условие (прошедшее время), когда числовой компонент при глаголе становится показателем не приблизительного, а точного количества. Это блокирует возникновение каких-либо приращений (увеличительных или уменьшительных) как в исходном утвердительном, так и в отрицательном предложениях.

2) Случай, когда компонент Q глубинной противопоставительной структуры 'не Р, а Q' может быть «вычислен» с помощью оператора Anti [6, 7]. В подобных ситуациях значение, соответствующее компоненту Q, оказывается прямо противоположным значению компонента Р. Так, фраза (8) *Он не будет соблюдать правила* на глубинном уровне соотносится со структурой 'Он будет действовать не в соответствии с правилами, а вопреки правилам'.

3) Наиболее распространенный случай, когда степень конкретизации компонента Q остается недостаточно информативной; можно лишь утверждать, что значение Q определено на таком множестве значений произвольной природы, в состав которого не входит значение Р. Ср. (9) *Отца не было на пляже* $\approx$ 'В рассматриваемом временном интервале местопребывание отца было не пляж, а было определено на множестве мест за вычетом пляжа'. При этом имеется в виду, что говорящий не может дать более информативную конкретизацию параметра 'местопребывание отца'.

4) В случаях, когда в СД нейтрального отрицания попадают событийные предикаты, в качестве единственной альтернативы Р выступает Q вида 'продолжение состояния, которое нормативно прерывается каузативом, образующим значение данного предиката'. Ср. (10) *Иван не приехал* $\approx$ 'Ожидалось, что в момент T_i Иван с помощью транспортного средства окажется в Y — месте, в котором находится говорящий (или лицо, о котором идет речь). В T_i и далее Иван продолжал находиться не в Y, а за пределами Y'.

5) Случай, когда глубинная противопоставительная структура может быть представлена с помощью т. н. «компонентной» модели. Данный подход предусматривает, что в основе логико-семантического представления предложения (и не только отрицательного) лежит утверждение о тождестве разных типов обозначения одного и того же множества произвольной природы. Одно его обозначение может быть задано той или иной дескрипцией, а другое описывает то же множество через перечисление составляющих его компонентов, степень детализации которых может варьироваться.

При таком подходе репрезентация противопоставительной структуры 'не Р, а Q' предусматривает, что элементы Р и Q отражают разные представления о составе одного и того же множества. Так, отрицательные предложения с «локативным общефактическим» [2] значением глагола *быть* вида (11) *Отец не бывал на море* интерпретируются следующим образом: (11) $\approx$ 'Можно было полагать, что в состав множества М мест, в которых бывал отец, входят и морские побережья, т. е. что М тождественно $\{M_i \& \text{морские побережья}\}$.

Множество М тождественно не $\{M_i \& \text{морские побережья}\}$, а $\{M_i\}$ (т. е. морские побережья в это множество не входят).

В советских столовых самообслуживания раньше часто предлагалось блюдо «биштекс с яйцом» (т. е. биштекс, который сверху накрывался яичницей из одного яйца). Но при желании можно было заказать и биштекс без яичницы. Логико-семантическая репрезентация такого заказа (ср.: *Мне, пожалуйста, биштекс без яйца*) может быть представлена следующим образом: 'Блюдо, которое я заказываю, тождественно не $\{\text{биштекс} \& \text{яичница}\}$, а $\{\text{биштекс}\}$ '.

Ср. также возможное толкование для глагола *обходиться* (без Y-a): (12) *X обошёлся без Y-a* $\approx$ 'Общепринято считать, что в состав множества М средств, необходимых для осуществления Р, входит Y, т. е. что М тождественно $\{M_i \& Y\}$. X осуществил Р посредством не $\{M_i \& Y\}$, а посредством $\{M_i\}$ '.

Приведенные примеры отражают ситуации, в которых множество, соответствующее компоненту Q, образует «часть» множества, соотносящегося с компонентом Р. Но бывают и противоположные ситуации, описываемые поверхностной структурой *не только Р, но и Q* — ср.: (13) *Стрепетовой восхищалась не только Россия, но и вся Европа* $\approx$ 'Было естественно полагать, что зрительская аудитория, восторгавшаяся сценическим искусством Стрепетовой, исчерпывалась Россией. Зрительская аудитория, восторгавшаяся искусством Стрепетовой, была тождественна не $\{\text{Россия}\}$, а $\{\text{Россия} \& \text{Европа}\}$ '. Как видим, в подобных случаях противопоставительная часть Q включает в свой состав и опровергаемый компонент Р.

3. Как известно, предложения, построенные на базе предикатов с внутренним отрицанием вида 'не Р' синонимичны предложениям с нейтральным отрицанием, в которых смысл 'не Р' выражен эксплицитно. Ср: (14а) *Он нарушает правила* = (14б) *Он не соблюдает правила*; (15а) *Он остался в Москве* = (15б) *Он не уехал из Москвы*. Однако в трансформационном плане между такими предложениями есть и существенные различия. Как мы видели, фразы с поверхностной структурой *не Р* могут быть распространены во фразы с противопоставительным отрицанием (причем сфера действия противопоставительного отрицания может не обязательно совпадать со сферой действия нейтрального — ср. (15б) и, с другой стороны, (16) *Он не уехал из Москвы, а улетел*). Для предложений же типа (14а)—(15а) такое распространение невозможно. Как было показано в работах [3, 4], для предикатов с внутренним отрицанием сфера действия отрицания является неизменной и не может выйти за пределы их тол-

кования. В частности, она не способна распространяться на такие эксплицитные элементы предложения, которые формально могли бы претендовать на роль противопоставления Q. Ср. (17a) *Коля не спит, а ходит*, но (17б) **Коля бодрствует, а ходит*.

Возможны и «тавтологические» примеры типа:

(18a) *Он не спит, а бодрствует*; (18б) *Он бодрствует, а не спит*

(19a) *Он не бодрствует, а спит*; (19б) *Он спит, а не бодрствует*,

функция которых состоит в эмоциональном опровержении уверенности слушающего в истинности ситуации, обозначенной компонентом Р.

Прежде чем обосновать нормативность подобных примеров посредством их семантического разложения, отметим, что фразовая эквивалентность рассматриваемого типа иногда носит не столько синонимический, сколько квазисинонимический характер. В действительности в значении многих предикатов с внутренним отрицанием часто инкорпорирован свой, индивидуальный тип имплицитного противопоставления, не вполне совпадающий с глубинным противопоставлением в предложениях с эксплицитным нейтральным отрицанием. Так, приведенный выше пример (1a) *Коля не спит* неоднозначен. Он может указывать как на то, что Коля занят какой-нибудь активной деятельностью (без уточнения какой именно), так и на то, что Коля лег спать, но уснуть пока не удалось. В последнем случае речь идет о том, что в момент речи степень релаксации организма Коли оказалась, против ожидания, неудовлетворительной, тогда как в первом имеется в виду, что в обсуждаемом временном интервале Коля ведет себя не так, как можно было ожидать — не спит, а активен. Именно такое значение ‘X не спит, а активен’ (без уточнения вида активной деятельности) и имеет предикат *X бодрствует*.

Обоснуем теперь нормативность предложений типа («18a» — «19б») на примере анализа предложения (19a), заменив слово *бодрствует* его толкованием:

(19a) *Он не бодрствует, а спит* = ‘Он не (не спит, а активен), а спит’ ≈ ‘Он не не спит, а не активен, а спит’, что после снятия двойного отрицания дает: ‘Он спит, а не активен, а спит’. По правилу зачеркивания гаплогогически дублируемого компонента [1] ‘а спит’ получаем: (19a) ≈ ‘Он спит, а не активен’.

Отметим, что квазисинонимичность свойственна и соотношениям локативных значений в (20a) *X не присутствует в Y* и (20б) *X отсутствует в Y*. Во-первых, (20a) и (20б) характеризуются разными модальностями исходного представления о том, что X мог в опреде-

ленное время находиться (присутствовать) в Y-е. Так, (20а) допускает более широкий спектр модальностей соответствующей идеи. Помимо модальности долженствования в него могут входить, например, и такие как допущение, предположение о присутствии X-а в Y-е, то, что в рассматриваемый период времени присутствие X-а в Y-е не исключено и т. д. В то же время (20б) опирается на высоковероятное ожидание присутствия X-а в Y-е, поскольку нормативно его пребывание в Y-е является естественным, закономерным (например, потому что X работает в Y-е, Y является местом его постоянного проживания и т. д.).

Во-вторых, отсутствие X-а в Y-е обычно предполагает временность его «неприсутствия». Это свойство проявляется в наличии у *отсутствовать* (в его локативном значении) обязательной семантической валентности срока (ср. *Он два дня отсутствовал*), которой нет у его аналитического аналога *не присутствовать*.

Обратим внимание на денотативное сходство ситуаций, описываемых в (20а — 20б) и рассмотренном ранее (9) *Отца не было на пляже*. В обоих случаях реализуется локативное значение — указание о том, что лицо, о котором идет речь, не находится там, где этого можно было ожидать. Однако в (5) реальное местонахождение отца, которое не может быть уточнено в рамках предложения, продолжает оставаться в фокусе внимания говорящего. В то же время в (20а — 20б) оказывается важным лишь сам факт отсутствия лица X в месте Y (то, что X находится в «не Y»), а установление его реального местопребывания является нерелевантным.

4. В традиционно принятых толкованиях предикатов со значением начинательности обращает на себя внимание некоторая неестественность («неподготовленность») появления отрицания. Ср., например, толкование, приведенное в работе [5]: (21) *началось P* = '(а) до некоторого момента имело место не-Р; (б) после этого момента имеет место Р'. Нормативно идее о «не-Р» должно так или иначе предшествовать представление о Р. Между тем в (16) первое упоминание о Р появляется уже после упоминания о «не-Р».

Мы исходим из того, что идея начала (как и конца) всегда соотносится с таким этапом развития какой-либо ситуации, более ранняя стадия которой уже так или иначе находилась в поле зрения говорящего и слушающего. На глубинном уровне эти идеи могут быть эксплицированы посредством противопоставительной структуры '(уже) не Р, а Q', где Р соотносится с более ранней стадией развития ситуации, а Q — с ее новой стадией. При этом имеется в виду, что данная

эволюция протекает на временной оси, разбитой на равные отрезки, измеряемые в произвольных единицах (секундах, минутах, часах, днях... годах и т. д.). Если речь идет о начале чего-либо, то в поле зрения попадает то новое Q , которое в i -й момент пришло на смену прежнему P , сохранявшемуся включительно до момента $(i-1)$; если же говорится о прекращении чего-либо, то акцентируется отсутствие в момент i и последующие за ним того P , которое имело место ранее. При этом первый актанта глаголов типа *начинаться* (или второй актанта *начинать*) соответствует переменной Q , а первый актанта *заканчиваться* (или второй актанта *переставать*) соотносится с переменной P . Таким образом, при глаголах первого типа нормативно не может быть заполнена валентность предшествующего состояния, а при глаголах второго типа — валентность нового состояния. Ср. (22) *Он начал заниматься политикой* $\approx$ 'В интервале T_i и последующих о нем можно утверждать не то, что имело место в интервале T_{i-1} и ранее, а то, что он занимается политикой'.

(23) *Он перестал заниматься политикой* $\approx$ 'Презумпция: В интервале T_{i-1} и ранее он занимался политикой, т. е. прилагал n -ное количество усилий для занятий политикой. Ассерция: В интервале T_i и позднее количество усилий, прилагаемое им для занятий политикой, равно не n , а равно 0'.

С этих позиций некоторая неестественность толкований типа (21) может быть объяснена тем, что в них в сферу действия отрицания попадает компонент Q глубинной структуры 'не P , а Q ', тогда как нормативно в неё должен попадать компонент P — ведь Q указывает на ситуацию, которая реально имеет место «вместо» того P , которого можно было ожидать (точнее, продолжения которого можно было ожидать).

В работе [5] убедительно показано, что в отличие от предикатов с внутренним отрицанием, сфера действия которого не способна выходить за пределы толкования предиката, отрицание, входящее в «значение начинательности, выраженное внутрисловно, может захватывать в свою сферу действия обстоятельство, занимающее рematicкую позицию». Это свойство внутрисловного значения начинательности сближает его со свойством частицы *не*, которая, «будучи отдельным словом, имеет возможность включать в свою сферу действия обстоятельство». Последнее положение наглядно продемонстрировано на контрастирующих примерах:

(24а) *Он не будет соблюдать правила приличия ради тебя и*

(24б) *Он будет нарушать правила приличия ради тебя*

В отличие от (24б), в (24а) поверхностная частица *не* включает в свою сферу действия обстоятельство *ради тебя*.

Рассмотрим подробнее семантическую интерпретацию (24а). В отличие от (24б), (24а), в частности, опирается на прагматическую презумпцию о том, что «он» вообще имеет обыкновение систематически нарушать правила приличия. Кроме того, (24а) базируется на исходном предположении, что ради слушающего он мог бы хотя бы временно отказаться от этой привычки, т. е. можно было полагать, что во временной отрезок T_i он ради слушающего не будет нарушать правила приличия. Тогда ассертивная часть значения (24а) может быть представлена следующим образом: 'В T_i он будет не не нарушать правила приличия, а не соблюдать их' $\approx$ 'В T_i он будет не не не соблюдать правила приличия, а не соблюдать их' $\approx$ 'В T_i он будет не соблюдать правила приличия, а не соблюдать их' $\approx$ 'В T_i он будет действовать не в соответствии с правилами приличия, а вопреки им'.

Как видим, хотя в (24а) обстоятельство *ради тебя* и входит в сферу действия отрицания, оно, тем не менее, не связывается ни с компонентом Р, ни с компонентом Q противопоставительной структуры 'не Р, а Q'. Оно становится важным элементом исходного предположения, указывая на ограничение («ради тебя»), в рамках которого соответствующее ожидание становится допустимым.

А (24б), где внутреннее отрицание не способно никак воздействовать на обстоятельство *ради тебя*, является целиком ассертивным и не образует никаких исходных предположений: (24б) $\approx$ 'Ради тебя он будет действовать не в соответствии с правилами приличия, а вопреки им'.

В случаях же имплицитного (как и эксплицитного) выражения начинательности обстоятельство, занимающее рематическую позицию, попадает в сферу действия внутреннего отрицания «иначе» — оно образует собственно противопоставительную часть Q структуры 'не Р, а Q':

(25) *Окончив школу, юноша занялся политикой всерьёз* $\approx$ 'Окончив школу, юноша начал заниматься политикой не так, как раньше, а всерьёз'.

5. В контексте изучения свойств отрицания представляет также интерес многозначный глагол *оставаться*. Он отличается богатством своих «отрицательных» значений, которые маркируются различиями в его модели управления.

Так, структура *X остался в Y-е* (ср. приведенный выше пример (15а)) соотносится с предикатом с внутренним отрицанием 'X не по-

кинул Y' , значение которого опирается на презумпцию вида 'ожидалось, что X покинет Y' '.

Структура вида X *остался* Y -ом (с творительным падежом второго актанта) констатирует, что вопреки закономерным ожиданиям, что X мог в том или ином отношении измениться, X , тем не менее, сохраняет прежнее свойство — ср. (26) *Он остался оптимистом (остался прежним...)*. Ср. также: *Каким ты был, таким остался...* Иногда внутренний предикат 'измениться' может быть выражен эксплицитно: *Приговор оставлен без изменений*.

Особый интерес представляет структура X *остался без* Y -а, имеющая самое общее значение нежелательной утраты чего-либо. Часто она указывает на негативно оцениваемое прекращение обладания той или иной собственностью: (27) *Он остался без денег (без сапог, без квартиры, без телефона...)*.

Если же актант Y относится к той же категории, что и X (например, если X и Y являются лицами), то семантика предложения может быть описана через значение «прекращения», которое репрезентируется посредством рассмотренной выше компонентной модели. В подобных случаях соответствующее событие представляется как негативно оцениваемое сокращение состава некоторого важного для X -а множества лиц, организованного по определенному признаку — множества, в которое изначально входили X вместе с Y -ом. Так, речь может идти о множестве лиц, связанных определенными узами, например, о семье.

Ср. (28) X *остался без отца (без жены, без брата...)* $\approx$ 'Презумпция: В интервале T_{i-1} и ранее множество M лиц, важных для X -а, было тождественно $\{X \& (M_i) \& \text{отец } X\text{-а (жена } X\text{-а, брат } X\text{-а...)}\}$. Ассерция: В интервале T_i и далее множество M тождественно не $\{X \& (M_i) \& \text{отец } X\text{-а (жена } X\text{-а, брат } X\text{-а...)}\}$, а $\{X \& (M_i)\}$ '. (Здесь заключенное в круглые скобки обозначение M_i указывает на некоторое факультативное (не обязательное) подмножество множества M — имеется в виду, что M помимо X -а и Y -а может факультативно содержать и другие элементы). При этом ситуация осмысливается (концептуализируется) с позиции X -а — лица, для которого сокращение состава множества M является нежелательным.

В других случаях актант Y рассматриваемой структуры может указывать на какую-либо неотделимую часть тела X -а — ср. (29) X *остался без руки (ноги, почки, двух зубов...)*. Интерпретация подобных фраз предусматривает, что исходное множество M соотносится с множеством «комплектующих», нормативно образующих тело X -а. При этом в представлении исходного множества среди его

компонентов особо выделяется (высвечивается) нормальное или исходное количество именно тех органов, которые были подвергнуты «изъятию»: (30) *X остался без двух зубов* $\approx$ '... Ассерция: В интервале T_i и далее множество частей тела X-а тождественно не $\{(M_i) \& n \text{ зубов}\}$, а $\{(M_i \& (n-2) \text{ зуба})\}$ '.

Заметим также, что если описываемая ситуация утраты не воспринимается как нежелательная, то фраза становится то ли сомнительной, то ли приобретает юмористический оттенок: (31) *Он остался без аппендикса*.

Как представляется, рассмотренная здесь «компонентная» модель может оказаться одним из эффективных средств репрезентации значения лексических единиц и высказывания.

Литература

1. Апресян Ю. Д. Лексическая семантика: Синонимические средства языка. М.: Наука, 1974.
2. Апресян Ю. Д. Типы информации для поверхностно-семантического компонента модели «Смысл $\Leftrightarrow$ Текст» // Wiener Slawistischer Almanach. Sonderband I. Wien, 1980.
3. Богуславский И. М. Исследования по синтаксической семантике. М.: Наука, 1985.
4. Богуславский И. М. Сфера действия лексических единиц. М.: Школа «Языки русской культуры», 1996.
5. Богуславский И. М. Сфера действия начинательности и актуальное членение: втягивание ремы // Семиотика и информатика. Вып. 36. М.: Языки русской культуры; Рус. словари, 1998.
6. Жолковский А. К., Мельчук И. А. О семантическом синтезе // Проблемы кибернетики. № 19. М., 1967.
7. Мельчук И. А. Опыт теории моделей «Смысл $\Leftrightarrow$ Текст». М.: Восточная литература, 1974.
8. Падучева Е. В. О семантике просодических сдвигов и вкладе просодии в семантику предложения // Сер. лит-ры и языка. Т. 63. 2004. № 5.
9. Труб В. М. О некоторых коммуникативных аспектах интерпретации отрицания // Тезисы докладов и сообщений Школы-семинара «Семиотические аспекты формализации интеллектуальной деятельности» Кутаиси, 23—28 апреля 1985 г. Москва 1985. С. 365—368.
10. Труб В. М. К проблеме функционального сходства нейтрального и контрастивного отрицания // Логический анализ языка: Тождество и подобие. Сравнение и идентификация. Институт языкознания АН СССР. М., 1990.
11. Труб В. М. О коммуникативных аспектах отрицания как негативной оценки истинности // Вопр. языкознания. 1994. № 1.

Annotating Lexical Functions in Corpora: Showing Collocations in Context

Agnès Tutin

Laboratoire de Linguistique et Didactique des Langues Etrangères
et Maternelles (LIDILEM), Université Grenoble 3
agnes.tutin@u-grenoble3.fr

In this paper, we show how the DiCo database of French Collocations (Polguère 2000) can be adapted in order to annotate collocational information in corpora for pedagogical purposes. We first present syntactic information that should be added to the database for this application and show how simple finite-state transducers associated with metagraphs (lexicon-grammars) can be successfully used to annotate collocations.

Introduction

Collocations are part of the class of frozen expressions and present remarkable linguistic properties for they do not seem to be entirely frozen nor fully compositional. They eliminate the clear-cut distinction between lexicon and syntax, since a collocation is usually associated with a syntactic construction (Lewis 2000, Gross 1975, Mel'čuk *et al.* 1995). However, despite the existing detailed syntactic descriptions (Hausmann 1989; Lexical Functions of Mel'čuk, Clas & Polguère 1995), these generally fail to show the syntactic and morphological variations of collocations. For example, in the LF model, a prototypical LF such as Oper_i is described from the syntactic viewpoint as a light verb having the noun as a second actant. In both the French ECD (Mel'čuk *et al.* 1995) and in the DiCo (Polguère 2000), the value of the LF is usually described merely with its standard construction. Nevertheless, within the framework of second language acquisition/teaching, it is crucial to place the stress on the verbal collocations' most representative syntactic patterns so as to facilitate the acquisition of the frozen expressions together with their respective syntactic use. For instance, a collocation such as *fonder une analyse sur* ('to base analysis on') can be observed in the active voice but

also, and more often, in the passive voice (*l'analyse est fondée sur*) or in the passive pronominal construction (*l'analyse se fonde sur*). With the use of large corpora together with NLP robust tools may also indeed facilitate the acquisition of such syntactic properties and thus could also be integrated in a learner tool. In this paper, we experiment automatic collocation annotation by adapting Polguère's DiCo (Polguère 2000) combined with patterns described with finite state transducers (Silberztein 1999). We first present what kind of syntactic information should be added to the Lexical Function database for the NLP application. We implement this database in a "lexicon-grammar" model and finally we evaluate the results of the automatic annotation in the semantic field of emotions.

1. Using the DiCo database to encode the LF

In the present study, our aim is to build a corpus of collocations in the semantic field of emotions for learners of French as a first or second language... These annotated examples could be related to a lexical database and enable to introduce information about the relative frequency of collocations, the most used syntactic patterns, morphosyntactic properties, etc. At this first stage, we focused only on Verb-Noun patterns such as *faire peur* ('to frighten'), *nourrir de la haine* ('to feel hatred'), *surmonter ses craintes* (lit 'to overcome fears') to assess the feasibility of the method. The Verb-Noun syntactic pattern is the most frequent and the most complex, as there are numerous alternations which can be observed:

- passive voice: *la haine est surmontée* (lit. 'the hatred is overcome').
- pronominal passive constructions: *la peur se lit sur son visage* ('fear can be read on his face').
- pseudo-passive constructions: *une peur difficile à surmonter* ('a fear hard to overcome').
- etc.

Our hypothesis is that the present approach can be easily extended to other verbal collocations. Adjectival collocations are predictably easier to annotate since they are less prone to variations.

We used the DiCo¹ database developed by Alain Polguère and his colleagues (Polguère 2000; Steinlin *et al.* 2004) as a basis for the annota-

¹ The Dicouèbe database developed by Polguère and his colleagues from the DiCo database has a more regular structure and could have been used for this experiment, but it was not yet available when this study began.

tion of our corpus. We then converted the latter into an Excel-like table, as presented in table 1 (we have presented only the fields relevant to our study). We selected nouns belonging to the semantic field of emotions (semantic labels such as ‘sentiment’ (*feeling*), ‘sentiment négatif’ (*negative feeling*), ‘émotion’ (*emotion*), ‘sentiment positif’ (*positive feeling*)) with LFs corresponding to Verb-Noun patterns such as: *provoquer le regret, susciter le regret, faire l’admiration...* and obtained a subset of collocations².

Though the DiCo is probably the principal source and the most detailed database which exists for French collocations, there remain certain imprecisions concerning formal properties of collocates for NLP applications. To provide a long term solution dealing with collocations in an NLP context, one should focus in establishing a relation between the value (e.g. *susciter*, CausFunc value for REGRET II.1) and the lexical entry of the LF value. Hence, syntactic and semantic properties would be registered and handled with inheritance mechanisms (Tutin 1997). We shall not develop this idea any further. We shall list only the linguistic properties which must be added to the collocates (the LF values) for our annotation application (Table 1).

L e m m a s o f t h e L F v a l u e

In the DiCo database, some values are not lemmas but inflected forms. It is the case for adjectival values in particular. For example,

Magn(*anxiété*) [‘anxiety’] = *folle, absolue* ... [‘mad, absolute’]

This inflectional presentation is obviously useful for pedagogical purposes and everyday use and thus must be preserved. However, the lemma must be added systematically in order to analyse other inflected forms which occur in the texts. As for verbal values, the problem does not concern the latter as they are always presented as lemmas.

M o r p h o - s y n t a c t i c i n f o r m a t i o n o f t h e L F v a l u e

In the ECD and the DICO, information regarding the part of speech of the LF value remains implicit: a given LF applied to a specific part of speech produces a specific part of speech. For example, when **Magn** placed before a verb it produces an adverb, yet when it is applied to a noun it produces an adjective. It gets more complex when the value is not a single word but a phraseme or even an expression which cannot be considered as a lexical unit.

² 755 collocations.

Lexeme	N°	grammatical information	semantic label	gloss	Lexical Function	value	grammatical properties of the LF value
REGRET 'regret'	II.1	nom, masc	sentiment	[Qqch.] causer un R. chez X	CausFunc1	provoquer	— [ART ~ Loc-in N=X]
REGRET 'regret'	II.1	nom, masc	sentiment	[Qqch.] causer un R. chez X	CausFunc1	susciter	— [ART ~ Loc-in N=X]
HAINE 'hatred'		nom, fém	sentiment négatif	[Qqch.] causer '[Y] être visé par la H. de X'	CausFunc2	attirer	— [ART ~ sur N=Y]
HAINE 'hatred'		nom, fém	sentiment négatif	[Qqch.] causer '[Y] être visé par la H. de X'	CausFunc2	canaliser	— [ART ~ sur N=Y]
ÉMOI 'agitation'		nom, masc, surtout sing	émotion	[Qqch.] rendre l'É. de X apparent	CausManif	trahir	— [ART ~]
PEINE 'sorrow'	II.1	nom, fém	sentiment négatif	[Qqch.] causer non 'P. commence à exister en X'	CausNonl ncepFunc1	épargner	— [ART ~ à N=X]
ADMIRATION ('admiration')	I	nom, fém, pas de pl	sentiment positif	[Y] être l'objet de l'A. de X	Oper2	faire	— [l' ~ de N=X]

Table 1. Records extracted from the DiCo database

Eng. Magn(rain): cats and dogs.

Fr. Magn(*pleuvoir*): (colloquial) *comme vache qui pisse* [lit. ‘like a pissing cow’]

Moreover, phrasal and single units do not behave the same way from the syntactic viewpoint³. Phrasal units have to be detailed according to their syntactic behaviour as is the case for the pronominal status of verbs such as *s’accroître* (‘to increase’).

To deal with this, two fields should be introduced:

- **a general part of speech information** including the phrasal status of the value, e.g. adverb (*badly*), adverbial “phraseme” (Fr. *par monts et par vaux* (‘up hill and down dale’)), adverbial expression (e.g. not a lexical unit) (*comme vache qui pisse*).
- **a detailed subtype information of the phrasal unit**. We have not tackled this problem in the present paper and more research has to be carried out on this specific topic.

Syntactic information of the LF value

In the DiCo database, some grammatical information is available for the LF value. For example, for the lexeme REGRET II.1, the CausFunc₁ value has the following properties: __ [ART ~ Loc-in N=X]. It means that the verbal value is followed by a determiner and a complement introduced by a locative preposition (this is the first actant of the key-word). Yet, this grammatical information does not suffice, for it should be broken down further into several fields for it is too concise.

First of all, the **deep syntactic relation** between the value and the key-word should be indicated. Yet, this information remains implicit in the ECD model, for the syntactic relation is taken into account in the LF definition, and can easily be acquired automatically⁴. In addition, the **surface syntactic relation** should also be clearly indicated (with values such as Object, Prepositional, etc), with the specific preposition(s) used. Then, **constraints on determiners** should appear in a specific field, with a consistent description (definite or indefinite determiners, lack of determin-

³ For example, use of auxiliaries, use of modifiers, etc.

⁴ Here is the list of syntagmatic LFs where the key-word is the second actant of the verbal value: AntiReali, Caus(i)Manif, Caus(i)NonManif, CausContFunci, CausFacti, CausFunci, CausOperi, CausPredMinus, CausPredPlus, Conv21Manif, FinOperi, LiquFunci, Magn--temp.Oper12, NonPerm1Fact0, NonPerm1Manif, Operi, PredAblei, Reali.

ers, ...). The **number** of the value has also to be filled in a specific field. Constraints on other **actants** (depending on the verb or the key word) should also be described.

For our application, we just replaced the DiCo's grammatical information with several fields (highlighted in the table of the following section) that we detail in the following section⁵.

2. A lexicon-grammar based method to extract and annotate collocations

In order to annotate the collocations in our corpora, we use a shallow analysis based on finite state transducers, more particularly on a methodology inspired by Maurice Gross's lexicon-grammar (Gross 1975). In this method, data registered in an Excel-like table are related with metagraphs (metatransducers) which are transducers (Roche 1993, Silberstein 1999, Laporte 2005). The tool used for this task is INTEX⁶ (Silberstein 1993, Silberstein 1999). Each cell in the table is related to a transition in the metatransducer and a large finite state transducer is generated from these data. The table below shows examples of entries of nouns of emotion. Due to the limited amount of space, we have only presented a simplified version of the table.

We shall focus on the first entry in our table, namely the collocation *emporter l'admiration* ('to be the admiration of'). The value which can be analysed as the value of an IncepOper₂₃ LF, is a verb (F column) and is related with the key-word *admiration* with the help of a deep syntactic second actant relation (G column) realised as an object surface relation (H column). The noun cannot be introduced without a determiner (value "-“ in the I column) but with the help of a definite article ("le") (J column). The metagraph exploits the recorded data in the cells of the lexical entry and consequently produces a transducer which includes all the data recorded in the tables. For reasons of readability, the following is a simplified version of a metagraph showing how these values are used.

The first path, on the top of the graph, shows how a collocation such as *emporter l'admiration* can be analysed in the active voice and how it

⁵ We did not introduce constraints on the actant of the value, because this field was not essential in our application.

⁶ Another tool called UNITEX (very similar to Intex) implements the same functionality (Laporte 2005). This functionality was first implemented in Intex.

A = LEXEME	B = N°	C = lemma	C = semantic tag	D = LF	E = LF value	F = cat. value	G = Deep Synt. Rel	H = Surf. Synt. Rel.	I = O Det?	J = type of DET	L = sing	M = plur
ADMIRATION 'admiration'		admiration	sentiment positif	IncepOper23	emporter	V	II	OBJ	—	:le	+	—
HO NTE 'shame'	I.1	honte	sentiment positif	NonPerm1M anif	caler	V	II	OBJ	—	:det	+	+
HONTE 'shame'	I.1	honte	sentiment négatif	Oper1	avoir	V	II	OBJ	+	:det	+	—

Table 2. Examples of records (derived from the DiCo database) used by the metagraph

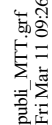


Fig. 1. The V-N collocation metagraph

has been annotated. The first transition (the variable @E) must be a verb in the table (column E) (the “< >” indicate a lemmatized form), possibly followed by a modifier such as an adverb or a prepositional phrase. The graph then assesses whether a determiner can be introduced and consequently chooses which one (column J), and then examines whether the noun chosen can be a plural or not (columns L and M). Thus, a collocation like *emporter l’admiration* can be analysed, but *emporter les admirations* will be rejected, since this collocation does not allow for a plural (column M).

The other paths on the graph are used for other syntactic constructions (relative clause, passive, etc.). Once the collocation is analysed, it is annotated with XML tags (see the last empty transition). Examples of annotated collocations are presented below.

"Et je vous jure qu'au fracas du coup de fusil que je n'attendais point, j'<colloc const="dir" colloc="avoir-angoisse_1" type_verbe="Oper1" type_nom="sentiment négatif" lemme_nom="angoisse">eus<colloc> une telle angoisse du cœur, de l'âme et du corps...

['I swear to you that by the crush of a rifle blow that I did not wait, I had much fear in heart, soul ad body']

C'était peut-être la première fois qu'un désir <colloc const="passif" colloc="manifeste_désir_1" type_verbe="verbalement Caus1Manif" type_nom="sentiment" lemme_nom="désir">manifesté</colloc> par le colonel eût obtenu l'approbation ...

['Perhaps it was the first time that the desire shown by the colonel got approval']

Vous avez expliqué l'intérêt que pouvait avoir Tomaso à menacer monsieur Barricini au nom d'un bandit redoutable, par le désir qu'il <colloc const="reldir" colloc="avoir-désir_1" type_verbe="Oper1" type_nom="sentiment" lemme_nom="désir">avait</colloc> de conserver à son frère Théodore ...

['You have shown the interest Tomaso could have in threatening Mr Barricini, in the name of a formidable gangster, for the mere pleasure that he had to keep for the sake of his brother Théodore.']

Fig. 2. Examples of annotated collocations

The verbal collocate is annotated with the help of the tag <collocate>. Several attributes give information about the verbal collocation:

- **const** indicates the type of syntactic construction : direct object (“dir”), passive construction (“passive”), relative clause (“rel”), ...
- **colloc** gives the lemmatized form of the collocation.
- **type_verbe** indicates the Lexical Function.
- **type_nom** gives the semantic type of the noun.
- **lemme_nom** gives the lemma of the noun.

These tags can be used for looking for specific examples in the corpus, in order to observe the syntactic or semantic behaviour of collocations.

3. Evaluation of the methodology and perspectives

Our evaluation was brought about through a set of novels and short stories of French classical authors, often studied at the level of “collège” (11—15 years of secondary school), and freely available. We would like to stress on the fact that our corpus should be used for pedagogical purposes, both for college pupils and for FFL students. The corpus includes around 233 000 words⁷.

Our lexicon-grammar is based on an extract of Polguère’s DiCo database of nouns of emotions⁸ and the corresponding transducers have been applied to the present corpus (see previous section). The output has been analysed with both the precision and the recall rates according to the data available. In general, the results are encouraging⁹.

⁷ The corpus includes the following texts:

- *Les Lettres de mon Moulin*, Alphonse Daudet.
- *Le petit chosé*, Alphonse Daudet.
- *La petite fadette*, George Sand.
- *Colomba*, Prosper Mérimée.
- *Le tour du Monde en 80 jours*, Jules Verne.

Most texts can be downloaded from the following URL : <http://abu.cnam.fr/>.

⁸ Here are the nouns included at the moment in the DiCo database: *admiration, angoisse, appréhension, aversion, coup de foudre, crainte, déception, dégoût, dépit, désir, effroi, émoi, extase, faveur, fierté, gratitude, haine, honte, hostilité, indignation, irritation, mécontentement, peine, rancune, regret, remords, répugnance, répulsion, ressentiment, sensibilité, vexation*.

⁹ the precision rate has been evaluated on 118 results.

The precision rate is quite satisfactory since we achieve 90% correct analyses. Several errors observed are due to polysemy and even the more sophisticated techniques, such as dependency parsers, could not easily avoid such errors. For example, the polysemous collocation *avoir de la peine* (lit. ‘to have sorrow’ or ‘to encounter a hardship’) is encountered in the text and it has been analysed erroneously.

Les policemen avaient beaucoup de peine à contenir le populaire, et à mesure que s’avançait l’heure à laquelle devait arriver Phileas Fogg...
(*Le tour du monde en 80 jours*, Jules Verne)

[‘The policemen struggled to contain the masses and as the hour approached in which Phileas Fogg was expected to arrive’]

The recall rate is approximately 86,2 %. Most errors are due to complex syntactic patterns such as coordination, interrogative structure or gapping, which have not been accounted for in our graphs, but could be, as shown in the following example:

-- Mais pourquoi as-tu une si grosse peine ?
[‘but why are you so troubled ?’]

Our method, despite its simplicity, provides satisfactory results, and could probably be easily extended to other collocational patterns. In the near future, we plan to explore the issue further following two directions of research:

- Firstly, to compare our pilot method to more ambitious methods using dependency grammars. A first experiment of dependency grammars using Xerox XIP dependency parser (Aït-Mokhtar *et al.* 2002) has been attempted in our team (Haddara 2004). However, it did not show a significant pre-eminence over finite-state techniques, even if the lexical data used were rather contrasting.
- Secondly, to extend this technique to other collocational patterns and provide a fully annotated corpus with the 50 most frequent nouns of emotion, to be exploited for pedagogical purposes.

Acknowledgements

Special thanks to Alain Polguère who offered me the DicO database for this experiment.

References

- Aït-Mokhtar S., Chanod J-P., Roux C. (2002), Robustness beyond shallowness: incremental dependency parsing. *Special issue of the NLE Journal*. 2002.
- Bolshakov I. A., Gelbukh A. Heuristics-based replenishment of collocation databases. In: E. Ranchhold, N. J. Mamede (Eds.) *Advances in Natural Language Processing* (PorTAL-2002). Lecture Notes in Computer Science, № 2389, Springer-Verlag. P. 25—32.
- Evert S., Heid U., Spranger K. (2004) Identifying Morphosyntactic Preferences in Collocations, in *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*, Lisbonne, 2004.
- Gross M. (1975) *Méthode en syntaxe*. Paris: Hermann, 1975.
- Haddara M. (2004), *Extraction automatique des collocations verbales du lexique transdisciplinaire des écrits scientifiques par une approche de type dépendance: étude de faisabilité*, Mémoire de DEA, Université Grenoble3-Stendhal, 2004.
- Laporte E. (2005) Graphes paramétrés et lexique-grammaire, *Interface lexique-grammaire et lexiques syntaxiques et sémantiques* (journée ATALA du 12 mars 2005) 2005.
- Lewis, M. (2000) (ed.) *Teaching Collocations*. Boston: LTP, 2000.
- Mel'čuk I (1998). Collocations and Lexical Functions. In A. P. Cowie (ed.), *Phraseology. Theory, Analysis and Applications*. Oxford: Clarendon Press, 1998.
- Mel'čuk I., Clas A., Polguère A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Louvain: Duculot, 1995.
- Polguère A. (2000), Towards a Theoretically-Motivated General Public Dictionary of Semantic Derivations and Collocations for French. In *Proceedings of EURALEX 2000*, Stuttgart, 2000.
- Roche E. (1993). Une représentation par automate fini des textes et des propriétés transformationnelles des verbes. In *Lingvisticae Investigationes* XVII: 1 (p. 189—222). John Benjamins Publishing Company: Amsterdam/Philadelphia, 1993.
- Silberztein (1999) «INTEX: a Finite State Transducer toolbox», in *Theoretical Computer Science* #231: 1, Elsevier Science, 1999.
- Steinlin J., Kahane S., Polguère A., El Ghali A. (2004), De l'article lexicographique à la modélisation objet du dictionnaire et des liens lexicaux, *Actes d'Euralex*, Lorient, 6 au 10 juillet 2004.
- Tutin, Agnès (2004), Pour une modélisation dynamique des collocations dans les textes, *Actes d'Euralex*, Lorient, 6 au 10 juillet 2004.

Connective Particles and Theory of Valence

Elena Uryson

Institute of the Russian Language, RAS
x-uryson@mtu-net.ru

Connective particles like *тем не менее* have two semantic actants. The problem is that strict linguistic rules are sufficient for specifying the expression of one of the actants, but regulate the expression of the other only partially. One cannot represent this case in the context of any formal syntactic model. The treatment of such connective particles offers a clearer view of syntactic scopes of a lexeme.

Фразовые частицы и теория валентностей. На материале лексики *тем не менее* продемонстрирована ситуация, когда слово имеет два семантических актанта, причем к поверхностному выражению каждого из них предъявляются определенные лингвистические требования. Однако лишь один из актантов полностью характеризуется правилами. Для определения второго требуется интеграция лингвистических правил и процедур обработки понятийной сети текста.

The classic theory of valence used in MTT is intended for the representation of semantic and syntactic actants of a predicate in the belief that both the predicate and its actants are expressed within the same sentence [Mel'čuk 1974; 1995; Apresjan 1974]. As we shall see later, the theory of scopes [Boguslavskij 1996], which covers a great deal more facts than the classic theory, is also confined to analysing a single sentence. It is of interest to consider connectors that express semantic links between sentences or larger text fragments from the standpoint of these theories.

I will review the Russian connective particle *тем не менее* 'still, yet'. My goal is to show that connective particles like *тем не менее* constitute a particular group of predicates. They have two semantic valences, and frequently one of them is non instantiated within the same sentence: the meaning in question can only be found in the preceding text. For the theory of valence adopted in MTT it means that there is no syntactic valence to match the semantic valence. However, the expression of both semantic

actants of a particle like *тем не менее* is regulated by strict rules. The problem is that these linguistic rules are sufficient for specifying the expression of one of the actants, but not quite sufficient for describing the expression of the other: they regulate its expression only partially. One cannot represent this case in the context of any formal syntactic model. Using metaphorical language, one can say that a particle like *тем не менее* has a real syntactic valence and a rudimentary one. The treatment of such connective particles offers a clearer view of syntactic scopes of a lexeme.

The particle *тем не менее* has attracted considerable interest of linguists. The meaning of this word has been described in detail by V. Apresjan [V. Apresjan 1999; 2004]. I will concentrate on the valences and the actants of this lexeme. For the sake of clarity I will contrast the particle *тем не менее* with the subordinate conjunction *хотя* ‘although’: these connectors have similar meanings but different syntactic properties.

- (1) a. *Было уже очень поздно (P). Тем не менее он продолжал работать (Q). VS.*
 b. *Хотя было уже очень поздно (P), он продолжал работать (Q).*
 [a. It was late (P), still he went on working (Q). VS. b. Although it was late (P), he went on working (Q)].
- (2) a. *Игрок брал не более пяти-шести нот, лениво тянул их, не стараясь связать в мотив (P), но тем не менее в его пiske слышалось что-то суровое и чрезвычайно тоскливое (Q) (A. P. Chekhov). VS.*
 b. *Хотя игрок брал не более пяти-шести нот и лениво тянул их, не стараясь связать в мотив (P), в его пiske слышалось что-то суровое и чрезвычайно тоскливое (Q).*
 [a. The player produced no more than five or six notes, dragged them out languidly with no attempt at forming a tune (P), and yet [*lit.*: but yet] there was something harsh and extremely dreary in the sound of the piping (Q). VS. b. Although the player produced no more than five or six notes, dragged them out languidly with no attempt at forming a tune (P), there was something harsh and extremely dreary in the sound of the piping (Q).]
- (3) a. *Да, это внезапное появление могло испугать кого угодно (P), и тем не менее оно являлось большой радостью (Q) (M. Bulgakov). VS.*
 b. *Хотя это внезапное появление могло испугать кого угодно (P), оно являлось большой радостью (Q).*

[a. Yes, this sudden appearance could frighten anyone (P) and yet it was a great joy for everybody (Q). *VS.* b. Although this sudden appearance could frighten anyone (P), it was a great joy for everybody (Q).]

Both the particle *тем не менее* and the conjunction *хотя* describe a situation with two participants (= semantic actants) P and Q, each of them in its turn being a situation. The actant P is the situation that usually prevents the existence of a situation like Q. The actant Q is the situation that in the given case takes place in spite of P. So the words under discussion have two semantic valences. Let us consider their syntactic properties.

A syntactic valence of a lexeme is, by definition, “the ability of a predicate lexeme to syntactically subordinate a word, a group of words or a clause that matches its semantic actant” [Апресян 2004: XXV]. Let us assume that the conjunction *хотя* subordinates both the main and the dependent clauses. The main clause denotes the actant P [the hindering situation, “the obstacle”], and the dependent clause expresses the actant Q [the “key” situation that takes place despite the obstacle]. Thus the subordinate conjunction *хотя* has two syntactic valences, both of them are obligatory and are instantiated within the same sentence. Each syntactic actant matches a certain semantic actant. The syntactic actants (= the expression of the semantic actants) can be entirely described by formal syntactic rules.

How many syntactic valences and syntactic actants does the word *тем не менее* have?

In sentences (1)–(3) the particle *тем не менее* is the first word of the clause, but it is not its obligatory position. Cf.

- (1) с. *Он тем не менее продолжал работать; Он продолжал тем не менее работать; etc.*

But in any case the word *тем не менее* is located within the sentence denoting the situation Q [the “key” situation that takes place in spite of the obstacle]. Let us assume that the word *тем не менее* subordinates this sentence. Thus the particle *тем не менее* has the syntactic valence for the expression of Q or, more properly, for the syntactic actant denoting the “key” situation Q. This syntactic actant is the sentence in which the particle *тем не менее* is located.

As for the actant P [the hindering situation, “the obstacle”], it can be denoted by quite large text fragments. Cf. examples (4)–(5) extracted from the texts where the description of “the obstacle” (in fact, of various hindering situations) takes several pages.

- (4) *Тем не менее костюм от редакции я получил* (S. Dovlatov).
[Still they gave me a new suit.]
- (5) *Тем не менее друг к другу они чувствовали настоящую приязнь и разговаривали обо всем совершенно свободно* (Yu. Dombrovskij).
[Yet they were friends.]

The definition of syntactic valences and syntactic actants does not apply in this case. The definition requires that syntactic actants be syntactic subordinates of the predicate lexeme. So syntactic actants and the predicate must be located within the same sentence.

One can regard the examples above as nontypical or infrequent and concentrate on the case when the actant P [“the obstacle”] is expressed in the adjacent sentence, cf. (1)—(3), or within the same sentence.

- (6) *Подавленный всем случившимся, он тем не менее продолжает петь свою арию* [P — ‘он подавлен всем случившимся’].
[Lit.: Being depressed [P] he nevertheless is singing his aria [Q].]
- (7) *Мало что понимая по-немецки, Иван тем не менее догадывался, как издевается над всем виденным Георг Майзель (А. Азольский)* [P — ‘Иван мало что понимал по-немецки’].
[Lit.: Understanding German very little [P] Ivan nevertheless guessed that George Maisel was mocking at everything around [Q].]
- (8) *В этом городе, где уже столько лет вытравляется дух приключения, оно тем не менее живет, ползет по улицам, лепится к окнам* (В. Аксенов) [P — ‘в этом городе уже столько лет вытравляется дух приключения’].
[In this city where... [P] it is nevertheless ... [Q].]

In these examples the semantic actant Q [the “key” situation] is expressed by the verb phrase in which the word *тем не менее* is located. It is reasonable to assume that *тем не менее* subordinates this verb phrase. So this phrase is the syntactic actant Q of the particle *тем не менее*. The semantic actant P [“the obstacle”] can be denoted by a variety of means: by a subordinate clause as in (8) or by an (adverbial) participle as in (6)—(7), etc. The point is that whatever agreements we assume it is impossible to regard these fragments as syntactic subordinates of the word *тем не менее*. Thus this word has no syntactic actant denoting the semantic actant P [“the obstacle”].

Let us apply to this case the theory of scopes [Boguslavskij 1996] which allows one to describe the linguistic properties of particles and its semantic actants expressed by a variety of irregular means.

In general terms, a syntactic scope of a lexeme can be expressed by any fragment of the syntactic representation of a sentence. The only requirement imposed upon this fragment is that its meaning corresponds to a certain semantic actant of the lexeme under discussion [Boguslavskij 1996: 43]. One could think that the definition of this kind perfectly fits our case, and the word *тем не менее* has a syntactic scope corresponding to the semantic actant P [“the obstacle”].

It should be emphasized that in actual practice syntactic scopes are specified by strict rules that can be used in linguistic algorithms. As for various expressions of the semantic actant P [“the obstacle”] of the word *тем не менее*, it is impossible to describe them by rules of this kind. This distinction between the cases described by I. Boguslavskij and the word *тем не менее* is of fundamental importance. The point is that the notion of syntactic scopes is an extension of the traditional notion of syntactic actants that was introduced for the representation of verbal complements. These complements have two characteristic properties: (a) they are immediate syntactic dependents of a verb; (b) each of the complement has specific morphological, syntactic or morphosyntactic features, determined by lexical properties of the subordinating verb or by the meaning of the complement [Apresjan 1979]. If one substitutes the notion of syntactic actants for the more powerful notion of syntactic scopes and at the same time drops any requirements to surface representation of a semantic actant, one should abandon the very idea of a linguistic model of an MTT type, and in particular the idea of parsing. The notion of syntactic scopes needs a slight correction: it should be mentioned that a syntactic scope of a lexeme can be described by strict linguistic rules.

The classic theory of valence and in practice its extension - the theory of scopes envisages the only case when a semantic actant has a well defined set of surface realisations, and it is possible to reveal the corresponding syntactic scope by a formal procedure. There are no such procedures in the case of the word *тем не менее*.

In a situation like this, one may say that the semantic actant under discussion is inexpressible (= the lexeme has no syntactic valence (no syntactic actant or syntactic scope) corresponding to the given semantic actant).

So, the semantic actant P [“the obstacle”] of the word *тем не менее* is inexpressible. This lexeme has only one syntactic valence, its only syntactic actant denotes the semantic actant Q [the “key” situation]. Because of this, *тем не менее* is a quasiconversive of the conjunction *хотя* [Uryson 2003].

Valence properties of the word *тем не менее* are determined by its part of speech attribution. It is not a subordinate conjunction like *хотя*,

but a connective particle. Notice that in (2)—(3) *тем не менее* is adjacent to a coordinate conjunction *но* 'but' or *и* 'and'.

The connective *тем не менее* demonstrates a new type of inexpressible semantic actants. Let us compare this particle with the verb *промахнуться* 'to miss when shooting', which is a classic example of a predicate with an inexpressible semantic actant [Mel'cuk 1974: 135; Apresjan 1974: 148].

The verb *промахнуться* 'to miss' describes the situation when the subject shoots (or throws something) at a target but does not hit it. So, the target is a participant (= a semantic actant) of this situation. The Russian verb *промахнуться*, unlike its English equivalent *to miss* has no syntactic valence corresponding to this actant. Cf. ungrammatical sentences ^{??}*Охотник промахнулся по тигру* <**в тигра*> (lit. 'the hunter missed at the tiger'). It is needless to say that in real texts the target is usually denoted by a certain word. Cf. *Охотник выстрелил в тигра, но промахнулся* 'The hunter shot at the tiger but missed'; *По выражению лица охотника мы поняли, что он опять промахнулся. А стрелял он, как выяснилось, в лису.* 'We understood from the hunter's look that he had missed again. It turned out that he had been shooting at a fox'. The point is that this word has no syntactic relation to the verb *промахнуться* and is impossible to identify by any formal procedure. One could think that the word *тем не менее* is similar to this verb. However, there are fundamental differences between the two predicates.

Elena V. Paducheva has demonstrated (see her article in this volume) that the semantic actant 'target' of the Russian verb *промахнуться* is inexpressible because it appears in the presupposition part of the lexicographic definition of the verb but not in its assertion part. In this respect the verb *промахнуться* does not differ from Russian verbs that denote «liquidation of the result of the action» [Apresjan 1974: 147]. Cf. *разбинтовывать (руку)* 'take the bandage or something like this off the arm', *разворачивать (покупку)* 'take the paper or something like this off what has been bought', *развязывать (ящик)* 'put the string or something like this off the box', *разгружать* 'unload', etc. Jury D. Apresjan noticed that these verbs cannot be defined without mentioning a certain X which A takes off B. Thus, these verbs have a semantic valence for this object, whilst the corresponding semantic actant is inexpressible [Apresjan 1974: 147]. It is easy to see that the inexpressible semantic actant of these verbs also appears only in the presupposition part of the lexicographic definition of the verbs. Naturally, there is no need to denote a presupposed actant in the text. Thus, these texts are grammatical: *Не знаю, зачем он стрелял, знаю только, что промахнулся* 'I don't know why he shot, I only know that he missed'; *Чем он зарабатывает?* —

Вагоны разгружает ‘How does he earn his living? — He unloads railway cars’; etc. (It should be noticed that in English the rules for specifying inexpressible semantic actants are different from those for Russian.).

As for the lexeme *тем не менее*, its semantic actant P [“the obstacle”] must be mentioned in the text; cf. examples (4)—(5), which require a pre-text describing “the obstacle”. Moreover, some formal requirements are imposed on the fragment denoting the semantic actant P: (A) This fragment must be located in the pre-text; (B) It must include a rheme (no matter whether a primary one, cf. (1)—(3), or a secondary one, cf. (5)—(7)). In some sense, the word *тем не менее* has a rudimentary syntactic valence for the semantic actant P [“the obstacle”] — one can specify certain syntactic requirements for the denotation of this semantic actant, but they are too scanty. (Perhaps these requirements are similar to the requirements to the antecedent of a pronoun like *he* or *she*.)

The same rudimentary syntactic valence is present in quasi-synonyms of the word *тем не менее*: these are particles *все-таки*, *все же*, *все равно*. (Their subtle semantics was described in [Paillard 1992; V. Apresjan 2004].)

To identify the semantic actant P [“the obstacle”] of the word *тем не менее* in a text, one needs both linguistic and encyclopedic information, results of syntactic and communicative analysis of the text, and knowledge network processing, as well.

This work was supported by RGNF (grant № 05-04-04190a), OIFN RAN (Programme “Slavic history, languages, and literature”, p. 4. 15), and grant NSH-1576.2003.6.

References

- Apresjan 1974 — Ю. Д. Апресян. Лексическая семантика. М., 1974.
 Apresjan 1979 — Ю. Д. Апресян. К понятию глагольного управления // Wiener Slawistischer Almanach. Bd. 3. Wien, 1979.
 Apresjan 2004 — Ю. Д. Апресян. Лингвистическая терминология словаря // Новый объяснительный словарь синонимов русского языка / Под общ. руков. акад. Ю. Д. Апресяна. 2-е изд., испр. и доп. М., 2004.
 Boguslavskij 1996 — И. М. Богуславский. Сфера действия лексических единиц. М., 1996.
 V. Apresjan 1999 — В. Ю. Апресян. Уступительность в языке и слова со значением уступки // Вопр. Языкознания. 1999. № 5.
 V. Apresjan 2000 — В. Ю. Апресян. Словарная статья «ВСЕ-ТАКИ 1» // Новый объяснительный словарь синонимов русского языка. Второй выпуск / Под общ. руков. акад. Ю. Д. Апресяна. М., 2000.

-
- Mel'čuk 1974 — *И. А. Мельчук*. Опыт теории моделей «Смысл ↔ Текст». М., 1974.
- Paillard 1992 — *D. Paillard*. *Vse-taki et vse že*: une question de point de vue // Linguistique et slavistique / Ed. M. Guiraud-Weber et Ch. Zaremba. T. 1. Aix-en-Provence — Paris, 1992.
- Uryson 2003 — *Е. В. Урысон*. Семантическая и валентная структура слов с уступительным значением // Русский язык в научном освещении. № 6.

What Type of Entity is a Lexical Function?

Leo Wanner

Institució Catalana de Recerca
i Estudis Avançats
Pompeu Fabra University
leo.wanner@icrea.es

Margarita Alonso

Facultade de Filoloxía
Universidade da Coruña
lxalonso@udc.es

Lexical functions are currently used as (1) a typology of co-occurrence and derivation relations between lexical units (LUs) in the dictionary; (2) deep LUs at the deep-syntactic level of the MTT-model, and (3) a lexicalization device during the transition between the deep-syntactic and surface-syntactic levels. This ambiguity is an obstacle, especially for the formalization and implementation of the MTT model. To avoid it, we propose a differentiated notation for LFs.

1. Introduction

It is well-known that lexical functions (LFs) as introduced in the Explanatory Combinatorial Lexicology (Mel'čuk et al. 1995; Mel'čuk, 1996) are a suitable means for the description of lexical co-occurrence and semantic derivation relations between two lexical units (LUs). LFs satisfy three requirements that are indispensable for such a means:

- they provide a semantically-oriented typology that allows for the association of an abstract meaning to the lexico-semantically restricted relation between the two LUs;
- they describe the syntax and the actantial structure of the unit formed by the two LUs in the case of syntagmatic relations, and
- they encode the functional dependency (and thus asymmetry) between the two LUs.

Nonetheless, LFs have often been criticized. The criticisms come from both outside and inside the Meaning-Text Theory (MTT) community. Researchers who examined LFs for use outside MTT usually complain that the LF-formalism is not “user-friendly” (too complex, too opaque, too

difficult to apply); cf., e.g., van der Wouden (1992) and Robin (1990). While MTT-researchers do not have problems with the use of LFs, some of them point out the flaws in the formal foundations of LFs (see, especially Kahane and Polguère, 2001, Alonso and Tutin, 1996 and Alonso Ramos, forthcoming) — for instance with respect to the distribution of the indices in LFs. Consider, for example, Oper_1 vs. Func_1 : in Oper_1 , ‘1’ stands for the first deep-syntactic actant, DSyntA I , of the LF-value, while in Func_1 , it stands for the second deep-syntactic actant, DSyntA II , of the LF-value.¹ This is inconsistent — even if the inconsistency can be readily explained: the information on the diathesis of a collocate verb is encoded by both the name of the LF and the actantial subscripts added to the LF-name.

In this article, we argue that some of the problems related to the use and interpretation of LFs as well as to the underlying formalism are due to the ambiguous use of the term “lexical function” and its notation. As is well-known, the terminology is very important in any comprehensive linguistic theory and thus in particular in MTT. Therefore, we think that the time has come to take the terminology of LFs seriously and to carry on the debate on the nature of LFs already opened by Kahane (2003). The reflections presented below aim to be a meta-linguistic exercise that contribute to this debate. But despite their meta-linguistic background, they are not to be seen as purely theoretical. Rather, they imply consequences for the definition and implementation of grammar formalisms in an MTM, for instance, for text generation.

In the next section (Section 2), we discuss the different uses of the term “lexical function”. In Section 3, we propose a modified notation for LFs at the DSynt -level of an MTM and a lexicographic description for LFs as lexical units. In Section 4, we summarize our argumentation and discuss briefly the consequences that our proposal bears for the formal description of an MTM and its implementation for purposes of Natural Language Processing (NLP).

2. The Ambiguity in the Use of Lexical Functions

According to the definition by I. Mel’čuk (cf., e.g., Mel’čuk 2001: 32), a lexical function f is defined as “a function that associates with a specific lexical unit (L) which is the argument or “keyword” of f , a set of

¹ We presuppose a certain familiarity of the reader with MTT and MTT-terminology. For a basic introduction, see, e.g., (Mel’čuk, 1988).

synonymous lexical expressions, $\{L_1, L_2, \dots\}$, the value of f ; cf. $\text{Magn}(\text{HATRED}) = \{\text{deep}, \text{profound}, \dots\}$. In a more rigid mathematical notation, we can write this definition as $f: V \rightarrow P^V$ (with ‘V’ being the vocabulary of a natural language L and ‘ P^V ’ being the power set of the vocabulary of L ; see also, e.g., (Mel’čuk, 1996) for the argumentation that LFs are functions in the mathematical sense). Nonetheless, a closer look at the exploitation of lexical functions in MTT reveals three different uses of LFs:

- as a means to encode lexico-semantically restricted co-occurrence or derivation in the Explanatory Combinatorial Dictionary (ECD);
- as a symbol (name) interpreted as a “deep lexical unit” at the deep-syntactic (DSynt) level of an MTM;
- as a “lexicalization” function during the transition between the deep-syntactic level and the surface-syntactic level of an MTM.

It is especially confusing if a specific LF-notation such as “ $\text{Magn}(\text{LOVE})$ ” in a given structure receives several of the above interpretations — depending on the purpose of the author. Let us discuss each of the uses in turn.

2.1. LFs as a Means to Encode Co-occurrence and Derivation

LFs were conceived as a universal (i.e., language-independent) device for encoding lexico-semantically restricted co-occurrence and semantic derivation in a lexical resource. Thus, in the lexicographic entry for the noun LOVE in an ECD, we will find the declaration of the value sets for all LFs for which the target domain is not empty (in the mathematical notation from above: $P^V \neq \emptyset$), given LOVE as argument:

LOVE

Gener:	feeling, emotion
Magn:	bottomless, strong, overwhelming, deep, ...
Ver:	true, pure, unconditional, real, genuine, ...
AntiVer:	selfish, ...
Oper ₁ :	feel, have
...	...

which is an alternative notation for the functional notation $\text{Magn}(\text{LOVE}) = \{\text{bottomless}, \text{strong}, \dots\}$, $\text{Ver}(\text{LOVE}) = \{\text{true}, \dots\}$, $\text{AntiVer}(\text{LOVE}) = \{\text{selfish}\}$, and so on. In contrast to the functional notation, this notation emphasizes the relational nature of the information encoded — as do all conventional collocation and synonym dictionaries.

2.2. LFs as Deep-Lexical Units

In the multi-stratal MTT-model, LFs form part of the node alphabet at the deep-syntactic level of representation (DSyntR). In the MTT-terminology, LFs are “deep lexical units” (Mel’čuk 1988, 1996: 85). In this sense, an LF *f* cannot be interpreted as a function. Rather, it must be interpreted as a name which belongs into the lexicon of any language.² As a lexical unit, it can occupy a node in the DSyntR of given sentence in the same way as any other lexical unit (LU) L_x .³ Consider in Figure 1 a simple DSynt-structure that contains two collocations encoded by two LFs in the traditional notation.

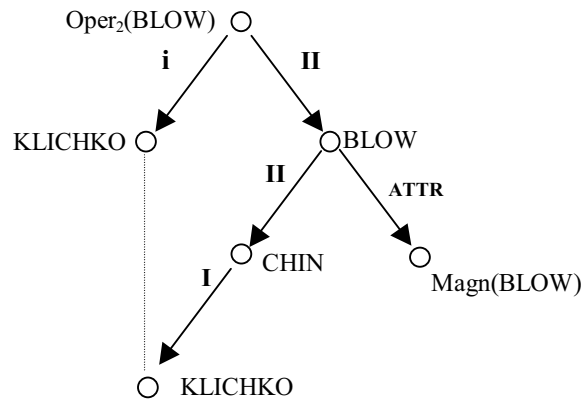


Fig. 1. DSyntS for *Klichko took a hard blow to the chin*.

Although at the DSynt-level of an MTM, LFs are claimed to be lexical units, the figure shows that LFs are encoded using functional notation; cf. introduction to Section 2. However, this notation implies that $\text{Oper}_2(\text{BLOW})$ is equivalent to $\{[to] \text{ take, get, receive, suffer, ...}\}$ and $\text{Magn}(\text{BLOW})$ is equivalent to $\{hard, severe, heavy, ... \}$. In other words, we are entitled to substitute ‘ $\text{Oper}_2(\text{BLOW})$ ’ as $\{[to] \text{ take, ...}\}$ and ‘ $\text{Magn}(\text{BLOW})$ ’ as $\{hard, ... \}$. This interpretation obviously leads to undesired consequences:

² Any speaker knows consciously or unconsciously that there are restricted lexical choices. That is to say, a speaker of English is aware of the incorrectness of the expression “to make a walk” (cf. in French, *faire une promenade*), even if he does not know the notion of support verb nor the LF Oper which encodes it. In that sense, we can say that lexical functions form part of the lexicon of a language (Polguère 2003).

³ The difference between L_x and an LF is that an LF is a *generalized LU*: the function has not been applied yet and its value is unknown, at this level.

- some nodes of a DSynt-structure are labelled by sets of LUs, which is in clear contradiction to the definition of DSynt-structures in an MTM: in a DSyntR, a set of collocates is not a valid node label, only an LU is. Therefore, an LF-node should be interpreted as an LU — a generalized and a special LU, but still an LU.
- the values of LFs are introduced at the DSynt-level while they are assumed in MTM to be introduced at the SSynt-level⁴;
- a DSynt-SSynt “interface” grammar as proposed, e.g., by (Kahane and Lareau, forthcoming), must be able to handle set-lexeme correspondences.

In order to avoid these consequences, we must thus use a “trick” used in functional programming languages such as Lisp: at the Deep-syntactic level, ‘Oper₂(BLOW)’ must be interpreted as a name (rather than as a function), i.e., a string of characters “Oper₂(BLOW)”⁵ whose “evaluation” is postponed to a later stage of processing. However, such an interpretation is, strictly speaking, not of linguistic nature: (a) the string “Oper₂(BLOW)” is not a valid linguistic sign, (b) the notion of a “postponed evaluation” of a name as a function does not form part of the description of a linguistic representation since it concerns the procedural instructions on the time and strategy of the interpretation of the entities of the representation.

An alternative, occasionally used, LF-notation at the DSynt-level is the name of LF without the argument thereof: ‘Oper₂’ instead of ‘Oper₂(BLOW)’, ‘Magn’ instead of ‘Magn(BLOW)’, etc. This notation is consistent with the assumption that LFs are (deep) LUs.⁶ Furthermore, it avoids the necessity of the distinction between a name and a function and thus also the problems that result from this distinction.

2.3. Lexical Functions as Lexicalisation Functions

Finally, LFs are used as lexicalization functions during the transition from the DSynt-level to the SSynt-level of an MTM in that each LF *f* is applied to the keyword L to which it is associated in the deep-syntactic

⁴ Note, however, that some authors argue for the introduction of LF-values at the DSynt-level; cf., e.g., (Iordanskaja et al., 1996).

⁵ When we say that “Oper₂(BLOW)” is a string of characters, we mean that it is considered as a whole as single label (without that BLOW and Oper₂ are identified as meaningful parts of it).

⁶ Recall Zholkovsky’s paraphrasing of one of Pushkin’s poems in the “LF-language”; cf. (Mel’čuk, 1996).

structure. However, since $f(L)$ returns a SET of values and a node of the SSyntS can only be a single lexeme, $f(L)$ alone is not a valid lexicalization device. Rather, $f(L)$ must be either embedded into a lexical choice procedure that selects one lexeme from the value set of $f(L)$ in accordance with predetermined context restrictions obtained from elsewhere or $f(L)$ must be redefined as a singularity function which returns each time it is applied to a keyword a single lexeme from the value set (the lexeme being different from the lexeme provided by the application before). In the current implementations, the lexicalization of LFs is usually specified in lexical choice procedures that make part of DSynt-SSynt transition rules.

If at the DSynt-level the functional notation of an LF, $f(L)$, is treated as a character string (cf. the discussion in the previous subsection) it must be “interpreted” (in the sense of a program language interpreter) as a function to obtain the values of the LF in question⁷.

2.4. Discussion of the Three Usages of the LFs

Obviously, for our interpretation of the three usages of the notion of LF as identified above counterarguments may be encountered. Or, if they are accepted, the relevance of their distinction may be questioned. The following argumentation could be put forward:

- A. In the ECD, the LF-value sets are declared for a given keyword, while during the transition between the DSynt-level and the SSynt-level of the MTM the LFs are applied. That is, there is no need to distinguish between usage 2.1 and usage 2.3.

From the discussion above it became clear that there are two principally different notions of LF: (i) LF as a deep LU and (ii) LF as a functional notation of co-occurrence and derivation relations. In this sense, usages 2.1 and 2.3 draw upon the same notion — the notion (ii). However, there is still a significant difference in the use of this notion in 2.1 and 2.3 that justifies our distinction. The difference is not only of practical, but also of theoretical nature, presupposing a change of the notation of the LF for usage 2.3. Thus, the purpose of applying an LF during the DSynt-SSynt transition is to obtain a valid unit of the SSynt-node alphabet. As already argued above, the fact that an LF provides a SET of units (and sets are not admitted as node labels at the SSynt-level) enforces a two step transition procedure:

⁷ In Lisp, the system function “eval(uate)” would be applied: (eval(‘Func₁(HOPE)’)).

- (1) application of the LF to the keyword in question (the result being invalid as a node label)
- (2) selection of one of the units from the set obtained in (1) (the result being a valid unit of the SSynt-alphabet)

In order to be able to define a formally uniform DSynt-SSynt transition model, it is desirable to avoid the intermediate “invalid” structure, which enforces to incorporate additional processing into otherwise purely declarative DSynt-SSynt node correspondences. This can be done by defining LFs during the DSynt-SSynt transition as *singularity* functions (see below).

That is, it is not only that in 2.1, the values of LFs are declared for the keyword in question and in 2.3 these LFs are applied. As we have seen, both usages differ also in the interpretation of what is expected from the application of an LF.

- B.** The ambiguity between the interpretation of an LF as a deep LU and as a function (i.e., the distinction between usage 2.2 and usage 2.3) is a “sloppiness” very common in formal descriptions and usually does not lead to misinterpretations. Moreover, the context helps to identify the correct interpretation.

This argumentation is correct in a framework in which (a) the value of the function in question is a valid unit of the alphabet over which the linguistic structures are defined; (b) the human reader is the only addressee of the description (and can thus arrive at the correct interpretation drawing upon the context). Both (a) and (b) do not apply to MTT:

- (i) as pointed out in 2.2, the values of LFs are not valid units of the DSynt-alphabet, and since for $f(X) = Y$, $f(X)$ is absolutely equivalent to Y , the functional notation $f(X)$ as a label of a DSynt-node is simply erroneous in the context of a DSynt-structure — as is the string notation “ $f(X)$ ”, which would require that “ $f(X)$ ” makes part of the DSynt-alphabet;
- (ii) the goal of MTT is a formally sound linguistic model which can readily be “translated” into an implementation.

3. Towards a Differentiated Notation of LFs

Based on the considerations presented in the previous section, we propose a differentiated notation of LFs (i) in the (ECD) as functions encoding co-occurrence and derivation relations in which a given LU is involved (usage 2.1), (ii) as deep lexical units at the DSynt-level of an

MTM (usage 2.2), and (iii) as lexicalization functions during the DSynt-SSynt transition.

Our suggestion takes further Polguère's (2003) proposal on the interpretation of LFs as "meta-lexical units".⁸ Given that a deep LU possesses the status of an LU, we assume that it receives an entry of its own in the ECD with a lexicographic description of its meaning, its syntactic combinatorics and its realization.

The difference between an LF interpreted as a deep LU and a free LU is that in contrast to a free LU, an LF has no associated lexical signs (e.g., the free LU [to] RAIN has as signs its inflected forms: *rains*, *rained*, etc.), and, therefore, no signifiers. The lexical signs of an LF are accessible only after the application of this LF to a specific keyword and the selection of one of the lexemes from the value set — e.g., the inflected forms of *feel* for **Oper₁**(LOVE). Preliminarily, the entries for **Magn** and **Oper₁** can be specified as follows (to emphasize the distinction between an LF as a deep LU and as a function, we add to the deep LUs the superscript 'DL'):

Magn^{DL}	
sem:	'intense very a lot -1 → L' ⁹
lex:	Magn^{DL} ↔ ♦ Magn (L)
Oper₁^{DL}	
sem:	∅
lex:	Oper₁^{DL} ↔ ♦ Oper₁ (L)
government:	V[I = X, II = L]

Note that **Magn^{DL}**, **Oper₁^{DL}**, etc. can be replaced by any other meaningful name. The symbol '♦' stands here for the "singularity" function that delivers one single element from a set (above, it delivers a single "surface-

⁸ Polguère (2003: 125) chooses the term meta-lexical unit ("metalexie", in French), because he conceives it as an entity which subsumes several lexical units — the result of the application of the LF.

⁹ As we see, the meaning of **Magn** is very vague; it can be formulated by a family of semantic glosses formulated in a controlled natural meta-language. However, when we speak about the meaning of LF, it is important to be clear about whether we mean the interpretation of the LF as a deep LU or as an element of the LF-value. For instance, the semantic content of **Magn** is vague (being paraphrased, roughly, as 'intense'). However, the semantic content of a given collocate is not necessarily vague. Thus, we can specify precisely what the semantic interaction between the base **ENEMY** and the collocate *bitter* is for **Magn**(**ENEMY**) = {*bitter*, ...}; see also (Alonso Ramos, 2003).

syntactic LU” from the value set $\text{Magn}(L)$ and the value set $\text{Oper}_1(L)$, respectively.

During the DSynt-SSynt transition, rather simple grammar rules ensure the proper functioning of LFs as a lexicalization device. For attributive LFs (including Magn^{DL}), the corresponding rule looks as follows:

```

context:    ?X -ATTR → ?Y
leftside:   ?Y
condition:  ?Y.type = deepLU
rightside:  λL lexicalize(LEXICON::?Y.lex(L)) (?X)

```

That is, we lexicalize a deep LU bound to the variable ?Y (e.g., Magn^{DL}) in that we access the lexical entry for this LU in the lexicon ‘LEXICON’ and retrieve the LU’s lexical sign “lex” equivalence. The “lex”-equivalence is a composed function made up of the singularity function and an LF (interpreted as function!) that is applied to the keyword ?X. The λ -calculus is a convenient instrument to unify ?X with ‘L’ from the lexical entry for the LU. The application of the LF to ?X provides us with a set of values of this LF; the singularity function picks one of them. With the function ‘lexicalize’, we ensure that from the different elements the most adequate one is chosen.

Grammar rules that capture the transition of verbal LFs have an analogous structure.

4. Conclusions: The Consequences for NLP

In this paper, we argued that LFs have three different uses in MTT and that these three uses need to be clearly distinguished. As a consequence, we proposed to extend the ECD by entries for LFs as deep lexical units, specifying in these entries all information that is necessary for a comprehensive and declarative description of an LU of this type.

The proposed differentiated notation of LFs has several advantages for the formalization and the implementation of the MTM:

- i) it allows for a formally “clean” definition of the DSynt-alphabet, as well as of the correspondence between any DSynt-LU and its lexical signs,
- ii) it allows for a clear declarative description of the different uses of the LFs,
- iii) it supports the implementation since a uniform handling of all DSynt-LUs is crucial for an NLP-grammar; in some generation

- grammars that are based on MTT, lexical entries for LFs are already used;
- iv) it makes the lexicalization task in generation more transparent (although more intelligence must be encoded in the singularity function). Furthermore, such a differentiation helps avoid ambiguities and misunderstandings among researchers working in MTT.

Our proposal is in line with Kahane (2003) as far as the distinction between usage 2.2 and 2.1/2.3 is concerned. Kahane does not single out usage 2.3. However, we think that this is indispensable for a formally clean elaboration of the MTM.

Acknowledgements

The work described in this paper was supported by the research grants BFF2002-04226-C03-01 (MCyT-FEDER) and PGDIT02PXIB30501PR and PGDITPXIC101401PN (Xunta de Galicia). We would like to thank Sylvain Kahane and the three anonymous reviewers for insightful comments, which helped to improve the final version of the paper.

Bibliography

- Alonso Ramos, M. (forthcoming), "Towards the Synthesis of Support Verb Constructions: Distribution of Syntactic Actants between the Verb and the Noun, L. Wanner (ed.) "Selected Lexical and Grammatical Issues in the Meaning-Text Theory. In Honour of Igor Mel'cuk".
- Alonso Ramos, M. (2003), "La nature des collocations: leur statut en tant qu'unités lexicales", in F. Grossmann et A. Tutin (eds.), *Les collocations : analyse et traitement, Travaux et Recherches en Linguistique appliquée*, Amsterdam, Editions De Werelt, 2003. P. 45—60.
- Alonso Ramos, M. and A. Tutin (1996), "A Classification and Description of Lexical Functions for the Analysis of their Combinations", in L. Wanner (ed.), *Lexical Functions in Lexicography and Natural Language Processing*, Amsterdam, Benjamins, 1996. P. 147—167.
- Kahane, S. 2003. "On the status of the deep-syntactic structure". In Proceedings of the First International Conference on the Meaning-Text Theory. Paris, 2003.
- Kahane, S. and F. Lareau (forthcoming), "Grammaire d'Unification Sens-Texte: modularité et polarisation", to appear in the Proceedings of TALN 2005, Dourdan, France.
- Kahane, S. and A. Polguère (2001), "Formal Foundation of Lexical Function", *Proceedings of Collocation: Computational Extraction, Analysis and Ex-*

- plaitation*, 39th Annual Meeting and 10th Conference of the European Chapter of the Association for Computational Linguistics, Toulouse, 2001. P. 8—15.
- Mel'čuk, I. A. (1998), "Collocations and Lexical Functions", in Cowie, A. P. (ed.) *Phraseology. Theory, Analysis, and Applications*, Oxford, Clarendon Press, 1998. P. 23—53
- Mel'čuk, I. A. (1996), "Lexical Functions: A Tool for the Description of Lexical Relations in a Lexicon", in L. Wanner (ed.), *Lexical Functions in Lexicography and Natural Language Processing*, Amsterdam, Benjamins, 1996. P. 37—102.
- Mel'čuk, I. A. (1988), *Dependency Syntax: Theory and Practice*, Albany, SUNY Press, 1988.
- Mel'čuk, I. A., A. Clas and A. Polguère (1995), *Introduction à la lexicologie explicative et combinatoire*, Louvain-la-Neuve, Duculot, 1995.
- Polguère, A. (2003), "Collocations et fonctions lexicales: pour un modèle d'apprentissage", in F. Grossmann et A. Tutin (eds.), *Les collocations : analyse et traitement, Travaux et Recherches en Linguistique appliquée*, Amsterdam, Editions De Werelt, 2003. P. 117—133
- Robin, 1990, "Lexical Choice". Technical Report. Columbia University, NY, 1990.
- van der Wouden, Ton (1992), "Prolegomena for a multilingual description of collocations", in *EURALEX'92 Proceedings I—II*, University of Tampere, 1992. P. 449—456.

The Concept of Diatheses and Voices in Early XXI Century¹

V. S. Xrakovskij

St. Petersburg Institute for Linguistic Studies, Russian Academy of Sciences
khrakov@vh2769.spb.edu

The paper reviews the development of the Diatheses and Voices Concept during the 35 years of its existence (1970—2005). The primary focus is on the basic postulates of the concept which, in the author's opinion, are not always observed in practical language descriptions. The aim of this publication is to give a detailed analysis of such deviations and outline ways to avoid them.

Концепция диатез и залогов в начале XXI века. В докладе рассматривается развитие Теории диатез и залогов за 35 лет ее существования (1970—2005). Основное внимание уделяется базовым постулатам теории, которые, по мнению автора, не всегда соблюдаются при описании материала конкретных языков. Цель публикации состоит в том, чтобы проанализировать случаи нарушения базовых постулатов теории и указать на некоторые пути их преодоления.

The fascination of science lies in the fact that one can never dot all the *i*'s in it. The development of the diatheses and voices concept during the first 35 years of its existence is one good example. One should simply compare the first papers [Xolodovič 1970] and [Melčuk, Xolodovič 1970] which formulated this concept and presented a seven-variant calculus of two-actant verb diatheses with the latest publications on this subject: [Melčuk 1998], in which the calculus of two-actant verb diatheses has 11 variants, and [Melčuk 2004: 293], in which the diatheses calculus has 20 theoretically possible variants, including 12 logically possible variants.

In his foreword to [Melčuk 1999: XIV], the author described the 25-year development of the Meaning-Text Theory (MTT) and noted a significant contribution made to the MTT by the diatheses and voices con-

¹ This text was prepared with financial aid from the RF President's Grant No. NS-2325.2003.6 provided to support Russia's leading schools of thought.

cept. This contribution is evident, since the concept provides a detailed analysis of all theoretically possible types of correlations between the semantic and deep-syntactic actants of the verb which, represented together in the word government model, make an important MTT component.

Many sciences have known situations when the basic postulates of a theory were not observed in practical descriptions made in the framework of that theory. In my view, this is the case with the papers [Melčuk 2004] and [Melčuk 1998]. The primary aim of this report is to identify such deviations and outline ways to avoid them.

In the first place, I have to note that I fully agree with the voice and diathesis interpretation as proposed in [Melčuk 2004: 287], where he says that “the notion underlying the concept of voice is the correspondence between the Semantic [=Sem-] and the Deep-Syntactic [=DSynt-] actants of verb forms, i.e. what we call diathesis”. In principle, I’m also prepared to agree with the definition of voice, by which “voice is a word-inflexion category, whose grammemes mark such changes of the basic verb diathesis that do not affect the propositional meaning of a given verb, i.e. do not change its meaning” [Melčuk 2004: 289]. However, I have certain doubts with respect to the three operations mentioned in the same paper and used for producing derivative diatheses from the basic diathesis. They are: (1) transposition (=permutation) of Deep-Syntactic actants (=DSyntAs) of the verb, (2) neutralization (=suppression) of the verb’s DSyntA, and (3) referential identification of two SemAs of the verb. The first operation is directly linked to DSyntAs numbering, and I do not always agree with the DSyntAs numbering in specific diatheses as proposed in the reviewed paper. Meanwhile, I believe this issue to be of crucial importance, since the overall form of a diathesis and voice calculus depends on the type of DSyntAs numbering used.

In [Melčuk 1999: 136], the DSyntAs numbering in Russian is defined by the following three criteria: (a) textual marking: all things being equal, the Nominative Case takes precedence (and, therefore, the Subject is invariably M_1), followed by the Accusative, Genitive, Dative, and Instrumental Cases that are further followed by the Infinitive and prepositional phrases; (b) possibility of omission: the smaller is the possibility of its omission from the text, the smaller is the actant number; and (c) “natural” linear order of the corresponding surface objects with respect to the governing word: the closer is the actant to the governing word, the smaller is its number. The paper also notes that actant numbering is “rather technical in character” and that the above criteria may be controversial, in which case “the decision is made on a more or less arbitrary basis”. While accepting the above criteria, I’d like to point out that all of them are not

semantic but formal and surface criteria. Apart from that, I believe that within the framework of the diathesis and voice theory, numbering is not of a technical, but of substantive nature, since, for example, a given actant numbering in a passive construction reflects our conception of the essence of passive modification.

Now let us consider some examples from [Melčuk 2004: 291], the DSyntAs numbering in which seems to need revision:

- (1a) *Masha [= I] poluchila prikaz [= II] ot Viti [= III] =*
lit. ‘Masha[**I**] got an order [**II**] from Vitia [**III**]’
- (1b) *Prikaz [= I] prišol Mashe [= II] ot Viti [= III] =*
lit. ‘The order [**I**] came to Masha [**II**] from Vitia [**III**]’
- (1c) *Prikaz [= I] byl polučen Mashej [= II] ot Viti [= III]*
lit. ‘The order [**I**] was received by Masha [**II**] from Vitia [**III**]’.

As I. A. Melčuk points out, all three examples are semantically identical. However, one should note that the examples (1a) and (1c) represent correlative active and passive constructions, and their combined analysis is quite appropriate in a paper on diathesis and voice. At the same time, there arises a question about the role of example (1b) (which, by the way, is not entirely natural): why did the author need this example? This example appears to be crucial for the author as a means to explain the DSynt number **II** given to the agentive object *Mashej* in (1c). I. A. Melčuk insists that “its semantic role, identical to the role of the object *Mashe* in (1b), guarantees it the DSynt number **II** (and makes it converse with respect to DSyntAs **I**)”. However, as noted above, the DSyntA numeration follows formal and surface criteria, which means that the reference to the role identity of the agentive object *Mashej* in the passive sentence (1c) and the object *Mashe* in the non-correlative active sentence (1b) is inappropriate and hits beside the mark. I. A. Melčuk needed this reasoning as an argument in support of his thesis that agentive objects are converse with respect to DSyntAs **I**, which means that they must be assigned the number DSyntA **II**. However, in compliance with the adopted criteria for DSyntA numeration (at least since the agentive object *Mashej* can be omitted from the text easier than all other actants), the *ot Viti* object in (1c) must be assigned the number DSyntA **II**, the agentive object *Mashej* must be assigned the number DSyntA **III**, and the example must take the following form:

- (1c) *Prikaz [= I] byl polučen Mashej [= III] ot Viti [= II]*².

² Actually, it is not quite clear why, after analyzing a prototypical case (the basic diathesis of the two-actant verb *comb*), in order to explicate the principle of DSyntA numbering, the author considers examples, the verbs of which do not

The DSyntA numbering pattern in passive constructions of three-actant transitive verbs also needs adjustment. In [Melčuk 1998] and [Melčuk 2004: 302] the following English examples are provided:

(2a) Active

Adam [= I] gave the apple [= II] to Eve [= III]

(2b) Direct full promotional passive

The apple [= I] was given to Eve [= III] by Adam [= II]

(2c) Oblique full promotional passive

Eve [= I] was given the apple [= II] by Adam [= III].

I have already analyzed these examples in [Xrakovskij 2000: 472—473], and now I only need to repeat the arguments put forward there. While I have no objections to the DSyntAs numbering in the basic active sentence, the second DSyntA status assigned to the agentive object *by Adam* in a sentence with a direct full promotional passive is arguable. The point is that the observed properties of agentive objects in English (the marking with the preposition *by*, the optional status, and the sentence-final position) perceptibly differ from the observed properties of the direct object in the basic active sentence (absence of prepositions, obligatory character, and the second position in the sentence). In addition, the agentive object *by Adam*, assigned the second DSyntA status, seems to have a less important role in the sentence than the object *to Eve*, assigned the third DSyntA status both in the passive and in the basic active sentence. Proceeding from the fact that the DSyntA numbering in both the basic and the derivative diathesis equally depends on the observed properties of the corresponding components, the construction of a direct full promotional passive should take the following form:

(2b') *The apple [= I] was given to Eve [= II] by Adam [= III].*

Correspondingly, the passive construction diathesis is obtained from the basic diathesis through the following operations: 1) the second SemA moves to the first DSyntA position, 2) the third SemA moves to the second DSyntA position, and 3) the first SemA moves to the third DSyntA position. Number III is given to the agentive object that did not appear in

denote situations but function as auxiliaries linking the name of the situation with their SemAs. In (1a) and (1c) the verb is Oper₂, while in (1b) it is Func₀. The point is that until a number of issues involved with describing such verbs' voice relationships are resolved (see [Melchuk 1999: 104]), it would be more appropriate to explain the principle of DSyntA numeration by using examples with standard verbs denoting situations.

the basic diathesis. Thus in this case the derivative diathesis is obtained from the basic diathesis not through simple conversion but through a more complex procedure.

The diathesis in the oblique promotional passive is non-typical in that the SemA moves to the first DSyntA position not from the second, but from the third DSyntA position. Although I have no objections to the DSyntA numbering in this construction, I have to note that the same agentive object is assigned two different DSyntA numbers (**II** and **III**, respectively) in two passive constructions, while its observed properties are absolutely identical in both instances. Most likely, this means that in both passive constructions the agentive object has the same DSyntA III status. Notably, the actual choice of the DSyntA status of an agentive object is determined by the presumption of the converse character of the basic and derivative diatheses: in direct full promotional passive constructions, the SemA moves from the direct object position to the first DSyntA position, while the agentive object is correspondingly assigned the second DSyntA status; whereas in oblique full promotional passive constructions, the SemA moves to the first DSyntA position from the indirect object position, with the agentive object correspondingly assigned the third DSyntA status. Meanwhile, the observed properties of the agentive object which are absolutely identical in both passive constructions and which, in compliance with the basic criteria, require that the agentive object should have the same syntactic status are, in fact, ignored.

In view of the above, I'd like to dwell in more detail on the presumption that the 'full passive' voice grammeme marks a DSyntA permutation, through which the benchmark (basic) diathesis is transformed into a conversational diathesis, cf. (1a) and (1c), (2a), (2b) and (2c). By the definition proposed in [Melčuk 1998: 247], "two diatheses are labeled *converse* diatheses if one of them can be built from the other through permutation of syntactic actants with respect to semantic actants, i.e. through changing the correspondence between syntactic and semantic actants". In itself, this definition does not raise any objections but, as I have already noted on many occasions (e.g., see [Xrakovskij 1974a: 31—33], [Xrakovskij 1974b], [Xrakovskij 2000]), **a diathesis marked with a 'full passive' grammeme generally cannot be conversational with respect to the benchmark (basic) diathesis.** My main argument illustrated above (see my comments to examples (1) and (2)) is as follows. In all derivative (passive) diatheses, in which it appears, the SemA occupying the first DSyntA position (= the subject) in the basic diathesis takes the position of the same specific DSyntA (agentive/subjectal object) **that never appears in the benchmark (basic) diathesis!** In the hierarchically ordered DSyntA

chain, this actant stands in the last position after all actants that can appear in the benchmark (basic) diathesis. If this assertion is true as I hope it is, the modification of the benchmark (basic) diathesis into any derivative passive diathesis, in which this DSyntA appears, takes place partly through DSyntA permutation and partly through substitution of a new DSyntA that corresponds to the SemA occupying the first DSyntA position in the benchmark (basic) diathesis. This new DSyntA, which is given the highest number, invariably represents the least important object. Irrespective of the specific number it is assigned, this DSyntA (=agentive/subjectal object) in all cases shows the same syntactic properties that distinctly differ from the syntactic properties of the DSyntAs of the benchmark (basic) diathesis and, correspondingly, of the active construction, the diathesis of which coincides with the basic diathesis. By the way, an adequate actant numbering in passive diatheses was proposed in the initial variant of the voices and diathesis concept, see [Xolodovič 1970: 12; 1979: 283; Melčuk, Xolodovič 1970]. The deviation from the adequate numbering was seemingly prompted by an erroneous presumption of the converse character of the active and full passive diatheses.

Now I'd like to address the third operation, the referential identification of two SemAs which "means that the corresponding variables are equated: $X = Y$, and this single SemA is expressed by a single DSyntA" [Melčuk 2004: 290]. I. A. Melčuk believes that the patterns of the correlative non-reflexive and reflexive constructions (3a) *Masha* (= **I**) *odevajet Vitiu* (= **II**) $\Rightarrow$ (3b) *Masha* (= **I**) *odevajetsia* lit. 'Masha dresses Vitia' $\Rightarrow$ 'Masha dresses herself' have the following forms:

X	Y	$\Rightarrow$	X = Y
I	II		I

I disagree with this interpretation in that it postulates only one SemA. If one postulates a single SemA instead of two in the basic diathesis, this calls for a conclusion that the propositional meaning of the verb is changed. However, the propositional meaning of the verb does not change, since the derivative diathesis retains the same two SemAs (= two different semantic roles) as in the basic diathesis. The only difference is that in contrast to the basic diathesis, in which each SemA has a corresponding referent and a DSyntA of its own, in the derivative diathesis both SemAs correspond to the same referent and to one DSyntA. It is exactly in this way that the diathesis of the reflexive construction (3b) *Masha* (= **I**) *odevajetsia* differs from the diathesis of the basic construction (3a) *Masha* (= **I**) *odevajet Vitiu* (= **II**):

X	Y	$\Rightarrow$	X	Y
R₁	R₂		R	
I	II		I	

If one takes a not especially exacting approach, one can maintain that the correlative reflexive and non-reflexive Russian verbs *odevat'* 'dress someone' and *odevat'sia* 'dress oneself' describe the same state of affairs. The only difference is that the non-reflexive verb denotes 'acting on another person', while the reflexive verb denotes 'acting on oneself', which involves identification of the referents.

By the way, apart from the reflexive grammeme, referent and syntactic-level modifications are also marked by the reciprocal grammeme which is excluded from the list of voices in [Melčuk 1998: 174], because "it adds another, fairly complex component to the lexical meaning of the verb: = ... 'one person acting on another person simultaneously, jointly participating in the action". If I'm not mistaken, in [Melčuk 2004] the reciprocal grammeme is not mentioned at all. However, one cannot simply ignore the reciprocal grammeme if only because in many languages this meaning is marked in the same way as reflexive meaning. One might believe that any reciprocal verb, e.g. Russian *celovat'sia* 'kiss each other' (as well as any reflexive verb), denotes the same state of affairs as the correlative basic verb (in this case, *celovat'* 'kiss'), while the semantic modifications distinguishing a reciprocal verb from its basic verb make the substance of reciprocal meaning. The reciprocal verb *celovat'sia* 'kiss each other' differs from the non-reciprocal verb *celovat'* 'kiss' in that within its diathesis, two referents simultaneously play two roles of the agent and the patient each, although each of the two referents correlates with two different actants. On the other hand, it is not actually clear if reciprocal meaning may be classified as that of a voice. Without going into detail, I can only note here that some considerations that both reflexive and reciprocal meanings have nothing to do with voice may be found in [Xrakovskij 1978; 1980; 1981; Xrakovskij 1982].

Another issue addressed in detail in [Melčuk 2004] is that of the homogeneity of voice grammemes. In short, the issue is whether the 12 logically possible voice grammemes belong to one category, or whether they are basically heterogeneous and must be classified with several independent categories. The baseline postulate is that voice is a single category, and it serves as a basis for the proposed calculus of the voice grammemes of two-actant verbs which comprises 20 theoretically possi-

ble and 12 logically possible variants. However, in his later analysis, taking into account that individual voice grammemes may appear together in one word form and providing corresponding examples, I. A. Melčuk distinguishes four equivalent independent voice categories: **Voice 1** (active — direct passive — oblique passive), **Voice 2** (neutral — permutative 2/3), **Voice 3** (non-suppressive — subjectless suppressive — objectless suppressive), and **Voice 4** (non-reflexive — direct reflexive — oblique reflexive).

Without going into the issue of the legitimacy of distinguishing specifically these categories, I'd like to address I. A. Melčuk's conclusion based on them. He maintains that theoretically we should recognize their existence, since "a formal pattern of this type makes it convenient to compare specific voices in specific languages" [Melčuk 2004: 306], and one cannot but agree with this. However, in practice he departs from the theory and, among other things, proposes to describe the voice in French as a category including the passive and the reflexive on the ground that their grammemes do not combine in the word form. To be sure, such a description is quite possible and economical, but it hides those theoretical differences between the passive and the reflexive which one cannot ignore when comparing the voice system of French with the voice systems of other languages. At the same time, there is another issue that is difficult to understand. If in theory we are obliged to accept the existence of four voice categories, then how should we regard the proposed definition of a single voice and the calculus of the voice grammemes of two-actant verbs based on that definition? Is not it a product of theory as well? But this product has no explanatory power if our theory recognizes the existence of four voice categories instead of a single voice.

Summing up the above, I'd like to emphasize that the intent of my critical comments is in no way destructive. I only believe that one cannot move ahead without an objective review of what has been achieved. In the 35 years of its existence, the diathesis and voices concept of I. A. Melčuk and A. A. Xolodovič has played an important role in the grammar theory development and, in particular, in the development of the MTT. It has proposed an evidently more adequate interpretation of voices compared to the traditional views and introduced theoretical facts that had never been considered in connection with voice issues. The papers targeted by my present comments, [Melčuk 2004] and [Melčuk 1998], also represent an important contribution to the theory of diathesis and voices. The goal of my comments is, on the one hand, to specify the basic postulates of the concept and on the other hand, to show the necessity to observe these postulates in specific material descriptions.

Literature

- Melčuk I. A. *Kurs obščej morfologii*. [General Morphology]. Vol. II. Moscow; Vienna, 1998.
- Melčuk I. A. *Opyt teorii lingvističeskix modelej "Smysl ⇔ Tekst"*. [A tentative theory of linguistic models "Meaning ⇔ Text"]. Moscow, 1999.
- Melčuk I. A. *Opredelenije kategorii zaloga i isčislenije vozmožnyx zalogov: 30 let spustia // 40 let Sankt-Peterburgskoj lingvističeskoj škole* [Towards a definition of voice and a calculus of possible voices: '30 years after' // 40th anniversary of the St. Petersburg School of Linguistics]. Moscow, 2004.
- Melčuk I. A., Xolodovič A. A. *K teorii grammatičeskogo zaloga* [Towards a grammatical voice theory] // *Narody Azii i Afriki*, 1970. № 4.
- Xolodovič A. A. *Zalog I: Opredelenije. Isčislenije* // *Kategorija zaloga: Materialy konferencii*. [Voice I: Definition. Calculus // Voice Category: Conference proceedings]. Leningrad, 1970.
- Xolodovič A. A. *Problemy grammatičeskoj teorii*. [Issues of grammar theory]. Leningrad, 1979.
- Xrakovskij V. S. *Konstrukcii passivnogo zaloga (operedelenije i isčislenije)* // *Kategorija zaloga: Materialy konferencii* [Passive voice constructions (definition and calculus) // Voice Category: Conference proceedings]. Leningrad, 1970.
- Xrakovskij V. S. *Passivnyje konstrukcii* // *Tipologija passivnyx konstrukcij: Diatezy i zalogi*. [Passive constructions // Typology of passive constructions: Diathesis and voice]. Leningrad, 1974a.
- Xrakovskij V. S. *Problemy sinonimii i konversivnosti passivnyx i aktivnyx konstrukcij* // *Universalii i tipologičeskije issledovanija: Meščaninovskije čtenija*. [Issues of synonymy and conversion of passive and active constructions // Universals and typological studies: Meščaninov Conference]. Moscow, 1974b.
- Xrakovskij V. S. *Zalog i refleksiv* // *Problemy teorii grammatičeskogo zaloga*. [Voice and reflexive meaning // Issues of the grammatical voice theory]. Leningrad, 1978.
- Xrakovskij V. S. *Nekotoryje problemy derivacii vzaimnyx konstrukcij* // *Semantičeskije aspekty slova i predloženiya* [Some issues of reciprocal construction derivation // Semantic aspects of the word and sentence]. Perm, 1980.
- Xrakovskij V. S. *Diateza i referentnost'* // *Zalogovyje konstrukcii v raznostrukturnyx jazykax*. [Diathesis and referentiality // Voice constructions in differently structured languages] Leningrad, 1981.
- Xrakovskij V. S. *Diatezy i zalogi (tridcat' let spustia)* // *Slovo v tekste i slovare*. [Diathesis and voice (thirty years after) // The word in text and in the dictionary]. Moscow, 2000.
- Khrakovsky V. S. *Diathesis* // *Acta Linguistica Academiae Scientiarum Hungaricae*. 1982. Vol. 29 (3—4).

Интонация иллокуций. Грамматика и словарь ¹

Т. Е. Янко

Институт языкознания РАН

tanya_yanko@list.ru

На материале русского языка выделены элементы семантического представления иллокуций и показано, что часть смыслов сочетается с другими смыслами, образуя иллокутивную грамматику, а часть смыслов не вступает в композиции с другими и требует фиксации в словаре. Рассмотрены интонационные элементы плана выражения иллокуций и выделен набор признаков, релевантных для различения интонационных единиц.

Intonation of illocutions. The grammar and the dictionary. Units of the semantic representation of Russian illocutions have been described. It has been shown that certain types of meanings can be combined with other meanings. They form, therefore, the illocutionary grammar, while some other types of meanings cannot interact with other meanings and should be kept in the dictionary. The intonation of illocutions and its relevant parameters have been viewed.

Просодические транскрипции не всегда нацелены на описание интонации с точки зрения выражаемых значений. Например, так называемый автосегментный подход Дж. Б. Пьерхамберт (Pierrehumbert 1980), завоевавший фонетический мир, предназначен для чисто фонетической интонационной транскрипции. Однако и в работах по интонации, где проблема связи формы и значения обсуждается (см. подход Е. А. Брызгуновой (Русская грамматика 1982: 103—118), работы Кодзасов 1996: 181—199; Кодзасов, Кривнова 2001: 392—398)), описание такой связи не всегда выдержано в той степени системности, которой эта связь заслуживает. Встает задача установления со-

¹ Работа выполнена при поддержке Программы фундаментальных исследований «История, языки и литературы славянских народов в мировом социокультурном контексте».

ответствия между множеством смыслов, которые можно выразить интонационно, и поверхностным выражением этих смыслов.

Известно, что интонация прежде всего связана с выражением иллокутивных целей говорящего, таких как сделать сообщение, задать вопрос, пригрозить, пожаловаться, позвать. И если обратиться к работам по анализу плана содержания, то окажется, что авторы гораздо более охотно апеллируют к сегментным средствам выражения иллокутивных смыслов, чем к интонации. К таким работам мы относим, в частности, книгу И. А. Мельчука, посвященную коммуникативной структуре предложения Mel'chuk 2001. Данные просодии там фактически не используются. Тем самым наиболее существенный пласт плана выражения структуры коммуникативных актов в полной мере не учитывается. Между тем, он способен подсказать, какие значения язык различает, а какие — рассматривает как одно значение, какие значения системно сочетаются друг с другом, а какие — используют застывшие идиоматичные средства. Кроме того, существенной типологической чертой русского языка оказывается то, что два основных типа речевых актов — сообщение и *да-нет*-вопрос — различаются только интонационно. И наконец, в русском и в других языках существует большое количество уникальных типов речевых актов, в которых коммуникативная цель говорящего передается интонацией.

Таким образом, традиционно, когда речь идет об интонации, план выражения и план содержания исследуются в существенной степени независимо. Цель данной работы — продвинуться по пути установления соответствия между иллокутивными смыслами и их интонационной манифестацией, а также представить основные иллокуции языка и интонационные средства их выражения как систему и отделить системные средства выражения иллокутивных значений от уникальных, или словарных. Ниже рассматриваются иллокутивные цели говорящих и другие смыслы, сопутствующие иллокутивным: контраст, эмпфаза, незавершенность текста. Основной объект анализа — русский язык.

В разделе 1 предлагается подход к типологии речевых актов в отношении к средствам выражения. В разделе 2 рассматривается множество иллокутивных смыслов, которые в русском языке образуют структуру. В разделе 3 приводятся примеры иллокуций, которые составляют иллокутивный лексикон. В разделе 4 по результатам, полученным в разделах 2 и 3, решается одна из задач описания иллокуций — выделение набора признаков, релевантных для различения типов речевых актов и отнесения их либо к типу системных, либо к типу уникальных.

1. Типы речевых актов в отношении к средствам выражения

Анализ русского материала показывает, что имеет смысл выделять три типа иллокутивных актов: *системные иллокуции*, такие как сообщение и вопрос, для которых характерна бинарная структура (тема vs. рема, собственно вопросительный vs. несобственно вопросительный компонент вопроса), способность к линейно-акцентным преобразованиям, связанным с приращением смысла, и сочетаемость с модифицирующими значениями (контрастом, эмфазой, верификацией); *частные, или уникальные, иллокуции* (интонация безрезультатной деятельности, интонация ментальных состояний, интонация обоснования, зов невидимого адресата), для которых характерна монолитная структура, неспособность к преобразованиям и к сочетаемости с модифицирующими значениями; *несобственно языковые иллокутивные средства*, которые носители языка называют приказным тоном, грозным голосом, просительной интонацией, металлом или слезами в голосе, жалобными нотками. Системные иллокуции и средства их выражения, преимущественно интонационные, составляют иллокутивную грамматику, т. к. удастся показать, что в русском языке сочетаемость коммуникативных компонентов с контрастом и эмфазой образует структуру, имеет регулярные средства выражения и, соответственно, может быть исследована операциональными методами. Частные иллокуции составляют иллокутивный лексикон, который исследуется инвентаризационными методами. Несобственно языковые интонационные средства здесь не рассматриваются.

2. Иллокутивные и другие коммуникативные смыслы.

Иллокутивная грамматика

Коммуникативное (или актуальное, по Матезиусу) членение предложения мы понимаем как отражение иллокутивного намерения говорящего. Речевой акт не элементарен. Он состоит из формирующего речевой акт, или собственно иллокутивного, компонента, такого как рема в повествовательном предложении, и несобственно иллокутивного коммуникативного компонента, который не формирует речевой акт, а только предворяет его совершение, такого как тема. Большинство речевых актов представляют собой единство собственно иллокутивного и несобственно иллокутивного компонента. Некоторые типы речевых актов неизменно атомарны, например обращение, у которого несобственно иллокутивный компонент отсутствует.

Кроме собственно иллокутивных и несобственно иллокутивных значений, выделяется ряд модифицирующих коммуникативных значений, таких как контраст и эмфаза, которые не формируют речевые акты, а только модифицируют компоненты речевых актов внутри одного типа. Модифицирующие значения соединяются со значениями, создающими речевые акты, с образованием контрастной ремы (*Во **вторник** он придет*)², контрастной темы (***Воскресную** прогулку пришлось отложить окончательно*), контрастного императива, контрастных компонентов вопроса. Так, в предложении *А какое оборудование завод Продмаш производит **серийно**?* контрастный несобственно вопросительный компонент вопроса словоформа *серийно*: здесь контраст говорит о том, что говорящий задает вопрос не о том оборудовании, которое завод производит в порядке опытного образца, а только о том, которое производится в промышленном режиме, т. е. серийно.

Семантика контраста состоит в соотношении выделенного элемента с известным говорящему и слушающему множеством, из которого делается выбор. Когда мы говорим ***воскресную** прогулку* — это значит, что прогулок, как минимум, две. Кроме того, при контрасте возникает идея о соотношении текущего речевого акта с некоторым высказанным ранее положением, с которым говорящим либо солидаризуется (*Да-да, это **Вася***), либо не соглашается (*Это **Вася**, а не Петя*).

Контраст сочетается как с ремой, так и с темой. Контраст соединяется с темой и ремой совершенно автономно, т. е. выделяются предложения с контрастными темами и простыми — не-контрастными — ремами, а также предложения с контрастными ремами и простыми темами, и предложения с контрастными темами и одновременно с контрастными ремами. Такая независимость контрастных тем от контрастных рем служит дополнительным аргументом в пользу того, что в повествовательном предложении следует вычленять два коммуникативных компонента — тему и рему.

Эмфазу мы понимаем как выражение говорящим сильных чувств по поводу ненормативных явлений жизни. Так, когда мы говорим *У тебя уже есть сын?!*, это значит, что говорящий считает, что слушающий еще чересчур молод для того чтобы иметь сына. Здесь значение эмфазы сочетается с собственно вопросительным (собственно иллокутивным) компонентом вопроса. Сочетаемость контраста и эмфазы с компонентами сообщения — темами и ремами — иллюстрирует Таблица. Заголовками строк Таблицы служат имена компо-

² Жирным шрифтом здесь помечен акцентоноситель контрастных компонентов.

нентов повествовательного предложения, заголовками столбцов — модифицирующие значения. На пересечении строк и столбцов приводятся русские примеры на взаимную сочетаемость этих значений в пределах одного коммуникативного компонента. В русском языке темы, ремы, компоненты *да-нет*-вопросов, а также композиции тем и рем с контрастом и эмфазой имеют тональные средства выражения, отражающие структуру плана содержания. Упрощая положение дел — ввиду ограничений, накладываемых форматом данной статьи, — отметим, что сочетания движений тона, маркирующих собственно иллокутивный и несобственно иллокутивный компонент, в контексте контраста и эмфазы последовательно преобразуются в соответствующие контрастные (которые отличаются бóльшим перепадом частот, чем простые, и большей интенсивностью) и эмфатические (которые отличаются своего рода «искривлением» тона, при котором к восходящему тону добавляется нисходящее начало, а к нисходящему — восходящее).

Таблица

	Контраст	Эмфаза
Рема	<i>Это Вася, (а не Петя)</i> (контрастная рема)	<i>Сто крон я за нее отдала!</i> (эмфатическая рема)
Тема	<i>Его новая книга лучше, чем старая</i> (контрастная тема)	<i>Такую огромную котлету мне ни за что не съесть!</i> (эмфатическая тема)

Таблица охватывает сочетаемость коммуникативных компонентов с контрастом и эмфазой только в пределах повествовательного предложения. Для вопросов и императивов характерна аналогичная картина, но масштаб данной статьи не позволяет нам увеличить в этой таблице число строк и столбцов. Продолжение *Таблицы* вниз охватило бы компоненты вопросов и императивов, а расширение вправо — позволило бы ввести ввести в круг рассматриваемых модифицирующих значений верификацию, или оценку истинности высказывания в целом, о верификации см. Адамец 1978: 101—103. Заметим, что эмфаза не столь универсально применима к компонентам речевых актов, как контраст. Так, сочетаемость эмфазы с темой затруднена. Попутно отметим, что несобственно языковые иллокутивные средства — так называемые «тона» и «голоса» (ср. *умоляющий тон*, *грозный голос*) — как и контраст, тоже композиционно накладываются на текст, но они соединяются не с темами и ремами по-

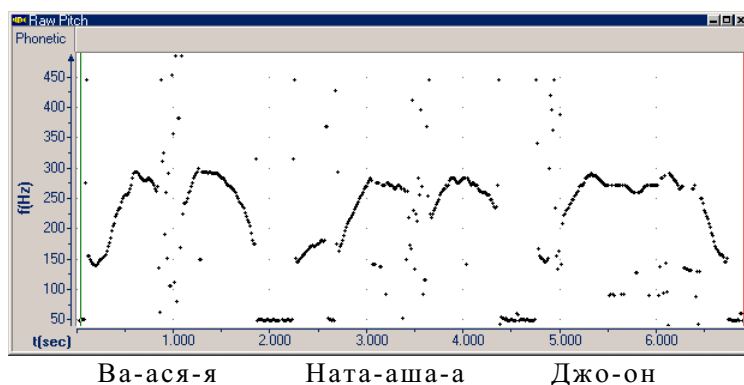
компонентно, а аморфно «растекаются» по большим фрагментам текста: *Ну дедушка, ну миленький, ну дай, пожалуйста* (пример заимствован из книги Н. Д. Светозаровой 1982: 23, 55, см. также Кодзасов, Кривнова 1977: 209 и там же литература по теме).

В данном разделе мы коротко обрисовали элементы семантического и фонтического представлений в применении к структуре речевого акта.

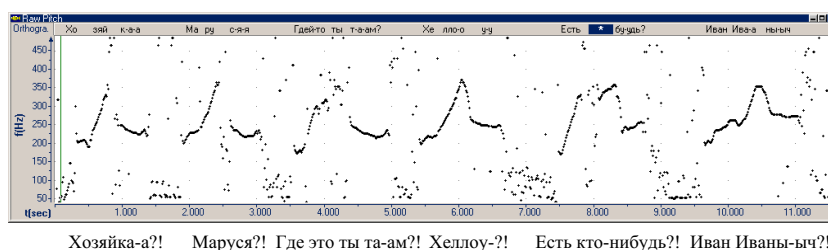
3. Иллокутивный лексикон

Перейдем к уникальным иллокутивным типам, которые хранятся в памяти говорящих в готовом виде. Их не удастся описать в виде композиций более дробных семантических признаков. Эти иллокуции весьма многочисленны и могут заслонять от исследователя системный характер основных русских иллокуций. Задача состоит в том, чтобы отделить один тип от другого и включить единицы, которые могут быть описаны операционально, в грамматику языка.

Для представления здесь уникальных иллокуций приведем несколько примеров. Наиболее типичные представители иллокутивных идиом — это вокативные контуры. Первый из анализируемых интонационных типов обслуживает ситуации существенной удаленности говорящего от слушающего. Тonoграмма показывает, что независимо от сегментного материала, на имени адресата фиксируется одна и та же акцентная кривая, которая занимает примерно одинаковое время звучания независимо от длины слова. Перед нами интонационный тип, который ложится на сегментный материал подобно музыке («as when chanted» по выражению Pierrehumbert 1980: 87—88). Обратимся к тонограмме.



Другой вокативный контур следует тому же принципу «выпевания». Он практически не зависит от структуры сегментного материала, в частности, от распределения ударных и безударных слогов. Он представлен в ситуации зова адресата, когда слушающий не знает, находится ли адресат в пределах слышимости. Говорящий обращается с призывом к невидимому адресату в пустой или темной комнате, в прихожей дома, когда неизвестно, на месте ли хозяева, к спящему, не будучи полностью уверенным в том, что спящий его слышит. Рассмотрим тонограмму.



Здесь мы наблюдаем ровный тон, за которым следует существенный подъем, тяготеющий к ударному слогу (в этом выражается некоторая зависимость данной просодической структуры от структуры сегментного материала). Затем, независимо от сегментного наполнения, наблюдается падение тона. Оно может приходиться и на изначально ударные, и на безударные слоги. Завершается контур растянутым, ровным и достаточно высоким тоном. Терминальный тон также не зависит от сегментного наполнения: несущий материал может быть получен за счет растяжения последнего гласного звука.

В следующем разделе при обсуждении признаков, релевантных для отнесения интонационной структуры к грамматике/лексикону, будет представлен еще один идиоматичный интонационный контур, который мы называем интонацией ментальной деятельности. Это припоминание (*Там тогда еще **знако-омы** собрались*); недоумение (*Куда-то мои **очки-и** запропастились*); передача чужой речи (*Тетя сказала: надо чего-то там **уко-олы** делать*); погружение в мечты (*А я **колба-аски** купила*).

4. Признаки, релевантные для идентификации интонационных единиц

В дальнейшем под основными единицами интонации мы будем понимать двусторонние единицы, для которых релевантен ряд при-

знаков, таких как способность быть носителем некоторого значения, фиксация на акцентоносителе, который выбирается согласно определенным правилам, движение тона на ударном слоге акцентоносителя и заударных слогах, если они есть, растяжки, аллегро и др. признаки, о которых речь ниже. Эти единицы мы называем акцентами. Они близки интонационным конструкциям Е. А. Брызгуновой.

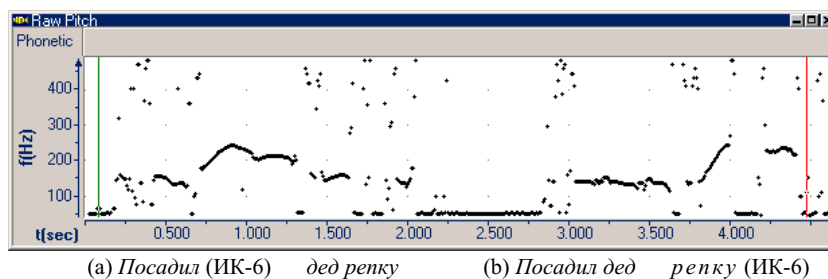
Для представления коммуникативно релевантных движений тона в грамматике и в словаре мы предлагаем выделять следующие признаки:

- значение (коммуникативная, или иллокутивная, функция);
- вхождение в одно- vs. двукомпонентные конструкции (ср. предложение *Дед посадил репку*, включающее тему *дед* и рему *посадил репку*, и предложение *Джон умер*, включающее только одну рему);
- привязка коммуникативно релевантного акцентного пика к определенному акцентоносителю (ср., с одной стороны, пример *Грачи* прилетели с акцентоносителем словоформой *грачи* и, с другой стороны, рассмотренную выше группу примеров с нечувствительным к акцентоносителю вокативным конструктором (*Хозяйка-а?!; Есть кто-нибудь?!;*);
- следование определенной системе правил выбора акцентоносителя;
- тональные параметры (форма частотной кривой, интенсивность, длительность, уровень, крутизна, или совершаемость движения тона в единицу времени, тайминг, или зависимость движения тона от распределения ударных и безударных слогов акцентоносителя и его контекста, и сопутствующие признаки, такие как паузация);
- отношение к сегментному материалу (в зависимости от структуры сегментного материала, в частности при отсутствии заударных слогов, акцент может подвергаться компрессии, как ИК-4 (по Брызгуновой), когда весь рисунок кривой основного тона без потерь фиксируется на единственном слоге акцентоносителя, а может усекается, как ИК-3, с потерей падения, которое фиксируется на заударных при их наличии);
- способность сочетаться с показателями модифицирующих коммуникативных значений (контраста, эмфазы) и дискурсивными показателями незавершенности.

Выделенные признаки позволяют относить акцент или склонять чашу весов в пользу отнесения акцента к типам «единица грамматики» или «единица словаря». Так, например, акценты, способные сочетаться с показателями модифицирующих значений и показателями

незавершенности, принадлежат грамматике, а единицы, не способные входить в композиции, принадлежат словарю; единицы, подверженные усечению в отсутствие заударных слогов, принадлежат грамматике, а единицы, подверженные компрессии, скорее, принадлежат словарю. Кроме того, набор выделенных признаков может быть использован в качестве схемы словарной статьи при описании иллокутивно-интонационной идиомы в иллокутивном словаре. Масштаб данной статьи не позволяет нам остановиться на всех признаках и проиллюстрировать их примерами. Обратимся к некоторым из них.

Относительно параметров «соотнесенность vs. несоотнесенность акцентного пика с определенным акцентоносителем» и «следование определенной системе правил выбора акцентоносителя» отметим, что во многих работах по интонации носитель акцента при анализе эмпирически берется из текста, и вопроса, почему акцент фиксируется именно на этой словоформе, а не на другой, не возникает. Однако различие между двумя акцентными контурами с разной семантикой может быть связано только с принципом выбора носителя соответствующего акцента. Приведем пример. В предложении из сказки (а) *Посадил дед репку* препозитивный глагол несет акцент типа ИК-6 по Брызгуновой, при котором на ударном слоге фиксируется подъем, а на заударных — относительно ровный высокий тон вплоть до словоформы *репку*, на которой фиксируется конечное падение: *Посадил* (ИК-6) *дед репку*. Это типичный линейно-акцентный контур рассказа о прошлом. Он используется в сказках, анекдотах (*Возвращается Петька из разведки*), занимательных историях (*Пришел больной к врачу; Прихожу я вчера домой*). Если же ИК-6 фиксируется на словоформе *репку* — (b) *Посадил дед репку* (ИК-6) — перед нами акцентный контур ментальной деятельности, введенный в рассмотрение в предыдущем разделе. Конечного падения у него нет. Этот контур уместен, например, в мечтах о будущем: *Посадил дед ре-епку-у. Осенью пареной ре-епкой полакомимся*. Тоннограмма иллюстрирует различие в локализации акцентного пика типа ИК-6 в минимальной паре предложений (а) и (b).



На примере просодии ментальной деятельности продемонстрируем зависимость выбора акцентоносителя от синтаксической структуры предложения. Рассмотрим просодию ментальной деятельности применительно к вопросам. Носителем акцента в таком вопросе служит вопросительное слово, ср. минимальную пару (с) и (d):

(с) *Куда-а* (ИК-6) *мои очки запропастились?*;

(d) *Куда-то* мои *очки-и* (ИК-6) *запропастились*.

С точки зрения лексико-синтаксической структуры различие между (с) и (d) только в частице *-то*, присутствующей в повествовательном предложении. Но переход от повествовательного предложения к вопросу, практически с тем же значением, немедленно меняет акцентоноситель. Что касается косвенных вопросов, они могут следовать обоим моделям выбора акцентоносителя. Если глагол пропозициональной установки служит в предложении вершиной, акцентоноситель в придаточном выбирается по правилам для повествовательных предложений (о выборе акцентоносителя в темах и ремах повествовательных предложений см. Янко 2001: 190):

(e) *Не знаю, куда* мои *очки-и* (ИК-6) *запропастились*.

Если же глагол рассматривается как вводное слово, акцентоносителем становится вопросительное слово:

(f) *Куда-а* (ИК-6) *мои очки запропастились, не знаю*.

* * *

На материале русского языка выделены элементы семантического представления иллокутивных смыслов и показано, что часть смыслов композиционально сочетается с другими смыслами, образуя иллокутивную грамматику, а часть не вступает в композиции с другими смыслами и требует индивидуальной фиксации в иллокутивном словаре. Рассмотрены интонационные элементы плана выражения иллокутивных смыслов и выделен набор признаков, релевантных для различения интонационных единиц русского языка.

Литература

Адамец 1978 — Адамец П. Образование предложений из пропозиций в русском языке. Acta Universitatis Carolinae Philologica. Monographia LXIX. Praha, 1978.

- Кодзасов 1996 — *Кодзасов С. В.* Законы фразовой акцентуации // Просодический строй русской речи. М., 1996. С. 181—199.
- Кодзасов, Кривнова 1977 — *Кодзасов С. В., Кривнова О. Ф.* Фонетические возможности гортани и их использование в русской речи // Проблемы теоретической и экспериментальной лингвистики. М., 1977. С. 180—209.
- Кодзасов, Кривнова 2001 — *Кодзасов С. В., Кривнова О. Ф.* Система русских интонационных единиц // Общая фонетика. М., 2001. С. 392—398.
- Русская грамматика. Т. 1. М.: Наука, 1982.
- Светозарова 1982 — *Светозарова Н. Д.* Интонационная система русского языка. Л.: Изд-во Ленинградского ун-та, 1982.
- Янко 2001 — *Янко Т. Е.* Коммуникативные стратегии русской речи. М.: Языки славянской культуры, 2001. 384 с.
- Mel'chuk 2001 — *Mel'chuk I.* Communicative Organization in Natural Language. John Benjamins Publishing Company. Amsterdam/Philadelphia, 2001.
- Pierrehumbert 1980 — *Pierrehumbert J. B.* The Phonology and Phonetics of English Intonation. PhD thesis. MIT. 1980.

Resemblances between Meaning $\Leftrightarrow$ Text Theory and Functional Generative Description

Zdeněk Žabokrtský

Institute of Formal and Applied Linguistics,
Charles University
zabokrtsky@cgl.mff.cuni.cz

Functional Generative Description (FGD) is a stratificational dependency-based approach to natural language description, which has been developed by Petr Sgall and his collaborators in Prague since 1960's. Although FGD bears surprisingly many resemblances with the Meaning-Text Theory, to our knowledge there is no reasonably detailed comparative study available so far. In this paper, we try to point out at least the basic obvious parallels (and also differences), which — in our opinion — remain generally unknown.

Сходства между моделью «Смысл $\Leftrightarrow$ Текст» и функциональным порождающим описанием. Функциональное порождающее описание (ФПО) представляет собой многоуровневый зависимостный подход к описанию естественного языка; оно разрабатывается в Праге Петром Сгаллом и его сотрудниками с 60-ых годов. Хотя ФПО обнаруживает удивительное сходство с моделью Смысл $\Leftrightarrow$ Текст, пока, насколько нам известно, не существует достаточно подробного сопоставительного анализа этих двух моделей. В настоящей статье мы пытаемся представить основные сходства и различия между двумя моделями.

1. Introduction

Nowadays, one can find a few comparisons of MTT and Chomskian streams in the literature, for instance [Nakhimovsky 1990] or [Bolshakov and Gelbukh 2000], but hardly anything has been written about the relation between MTT and FGD, although they have so much in common. Just for the beginning:

- both approaches originated almost at the same time ([Žolkovskij and Mel'čuk 1965] for MTT, [Sgall 1967] for FGD);
- both approaches share the same roots (or, at least, the similar background) in European structural linguistics;

- both approaches are stratificational;
- both approaches are dependency-oriented — they use dependency trees as the representation backbone on both syntactic levels;
- although both approaches proved to be useful and successful in large-scale implementations (machine translation system ETAP for MTT, [Apresian et al.2003]; Prague Dependency Treebank (PDT)¹ for FGD, [Hajič et al. 2001]), they are still outside the main stream of computational linguistics, which is governed by phrase-structure models.

We are not aware of any larger text comparing MTT and FGD — in the literature we found only a few marginal remarks (e.g. in [Sgall 1997]). We believe that there is hardly any third approach sharing so much with MTT and that is why we would like to draw attention to selected similarities between MTT and FGD. In the presented comparison (which is still preliminary and far from exhaustive) we will compare MTT and FGD from the following viewpoints:

- in Section 2 we show the similarity of FGD and MTT using the list of MTT features enumerated in [Bolshakov and Gelbukh 2000];
- in Section 3 we compare the systems of levels used in MTT and FGD;
- in Section 4 we compare deep-syntactic level of MTT with tectogrammatical level of FGD.

2. «Peculiarities of the MTM»

We borrowed the title of this section from [Bolshakov and Gelbukh 2000], where the authors present some properties of MTT as peculiarities of the MTM [Meaning-Text Model], which it retains as compared with other linguistic models”. In the following paragraphs, we go through most of these properties and show that in fact they could be ascribed to FGD too:

«Orientation to synthesis». Both approaches put originally emphasis on language synthesis (generation) rather than on analysis. However, in their implementations they proved to be able to work also in the opposite direction (building trees above a raw corpus in case of PDT, analyzing a sentence in the input language in case of ETAP)

«Multilevel character of the model». In both approaches, language description is organized into a set of linearly ordered levels (Section 3).

¹ FGD served as the theoretical framework for PDT. However, there are many subtle differences between the annotation scheme of PDT and the original FGD.

«Distinguishing deep and surface syntactical representations». FGD distinguishes deep and surface syntactical levels in a very similar way to MTT.

«Independence between the composition of words and their order in a sentence». As in other dependency formalisms, FGD clearly distinguishes between dependency and linear ordering of words. The nodes of the dependency tree at the tectogrammatical level of FGD are still ordered, but this ordering is used for representation of information structure instead of surface order.

«Accounting of communicative structure of text». Communicative structure (or information structure, in the Praguian terminology) is a part of the tectogrammatical level of FGD.

«Orientation to languages of a type different from English». FGD was originally developed for Czech, but English was often used for illustration of some aspects too. Recently, tectogrammatical representation of English is being elaborated in detail ([Cinková 2004]) and also experiments with tectogrammatical representation of German have been performed; the layered annotation scenario of PDT (morphological, analytical and tectogrammatical level) derived from FGD was also used for building a treebank of Arabic ([Hajič et al. 2004]).

«Labeling syntactic relations between words». Again, similarly as other dependency formalism, FGD does label the arcs of dependency trees (on the tectogrammatical level of FGD, these labels are called functors).

«Keeping traditions and terminology of classical linguistics». In this aspect, FGD obviously followed the tradition of the Prague School.

3. System of levels

Both in MTT and FGD, the relation between meaning and text/sound is defined via a sequence of correspondences between adjacent levels ([Sgall et al. 1986], page 7: ‘among others, the Prague School elaborated it [the distinction between the meaning and sound patterns] as a sequence of well-motivated levels articulating the complex relation between meaning and sound.’). In Praguian perspective, such correspondences are interpreted as a relation between form and function.

There are seven levels in MTT and five levels in the original FGD. Their rough alignment depicted in Figure 1 should mirror the following observations:

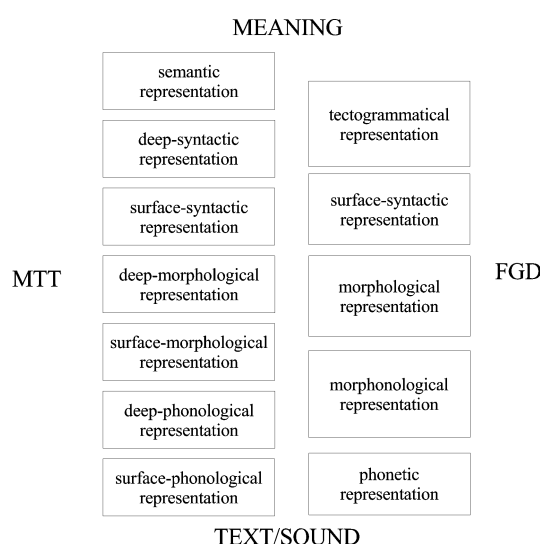


Fig. 1. Levels of utterance representation in MTT and in FGD.

- In FGD, there is no direct counterpart of what is called semantic representation in MTT². No notions comparable the notions of lexical function or semantic primitives were elaborated within FGD so far. However, there is a certain overlap of the MTT-semantic level with FGD-tectogrammatical level, namely they both contain the representation of the information structure (resp. communicative structure).
- tectogrammatical level matches almost perfectly the deep syntactic level of MTT, some details will be discussed in 4. But again, there is a certain overlap between surface syntactic level of FGD and deep syntactic level of MTT, because deep-syntactic relations in MTT ([Mel'čuk 1988]. P. 64—65) seem to be more surface-oriented than functors in tectogrammatical level.
- Surface syntactic level of MTT almost perfectly matches surface syntactic level in PDT³: both representations have a form of a de-

² However, the necessity of a higher level was declared in FGD from the very beginning, see [Sgall 1967], p. 56.

³ We have not found an explicit description of the FGD surface syntax level in literature, that is why we compare MTT only to PDT at this point. Moreover, in the recent versions of FGD (roughly since 1990's), surface syntax is not treated as an autonomous level of language description.

pendency tree where nodes correspond to the individual words of the sentence and the arcs are labeled with names of surface syntactic relations.

As the time went by, purely probabilistic models rather than linguistically motivated solutions were more successful in the field of spoken language processing, which is probably the reason why modern implementations of both approaches do not focus on the lowest levels of language description any more. We will not make any comments on the lower levels here.

As for the highest levels, both approaches state in a similar way where their understanding of (linguistic) meaning ends:

- [Mel'čuk 1988], p. 44: «'Meaning' ...refers only to information conveyed by language; in other words, it is whatever can be extracted from or put into an utterance solely on the basis of linguistic skills, without recourse to encyclopedic knowledge, logic, pragmatics or other extralinguistic abilities», p. 46: «...the correspondence between semantic representations and cognitive representations ⟨...⟩ goes beyond the scope of MTT».
- [Sgall et al. 1986], p. 13: «the concept of meaning ⟨...⟩ has to be distinguished from that of (cognitive, ontological) content (or factual knowledge)... The analysis of meaning ⟨...⟩ cannot be replaced by an analysis of 'extralinguistic' reality itself».

Their attitude to semantically or logically defective sentences also seems to be the same (perhaps in allusion to the discussion around Chomsky's sleeping green ideas):

- [Sgall et al. 1986], p. 106: «A sentence meeting the restrictions of strict subcategorization, but not the selectional restrictions, seems to have a meaning (or even several): witness the fact that it can be translated into other languages... Thus, it is not advisable to 'asterisk away' sentences that merely haven't found any occasion of use, thanks to our image of the world».
- [Mel'čuk 1988], p. 47: «the analysis of meaning itself goes beyond the scope of MTT: it does not distinguish "normal" meanings from absurdities, contradictions or trivialities. Discovering that something is stupid or absurd or detecting contradictions is by no means a linguistic task».

4. Deep-syntactic level vs. Tectogrammatical level

In the following list we would like to note several similarities between the tectogrammatical level in FGD and deep-syntactic level in MTT (a

simplified sample of tectogrammatical tree representation is depicted in Figure 2; the example comes from Prague Dependency Treebank 2.0,⁴ which will be released in the near future):

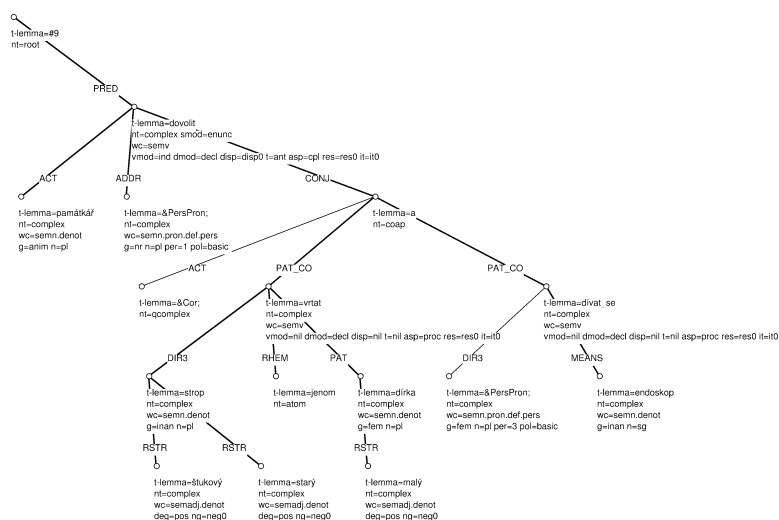


Fig. 2. Simplified tectogrammatical tree structure of the sentence «Do starých štukových stropů nám památkáři dovolili vrtat jenom malé dírký a dívat se do nich endoskopem» from PDT 2.0. (The conservationists allowed us to drill only small holes into the old parquet ceilings and to look into them with an endoscope).

- the skeleton of both representations is formed by dependency tree (unordered in MTT, ordered according to information structure in FGD);
- only semantically full lexemes (autosemantic words) do have nodes of their own (semantically empty lexemes/synsemantic words, such as prepositions, subordinating conjunctions,
- auxiliary verbs etc. are introduced only in the surface-syntactic structure);
- each lexeme is associated with appropriate semantically full grammemes (grammatemes in FGD terminology); grammemes imposed only by government and agreement are excluded; both approaches work with fictitious lexemes (needed e.g. to encode semantically non-empty syntactic constructions);

⁴ See <http://ufal.mff.cuni.cz/pdt2.0>.

- dependency tree is accompanied with (non-tree) grammatical coreferential relations, together forming dag (directed acyclic graph).⁵

Occasionally, the question of convergence of the Western main stream (computational) linguistics with the dependency-based approaches is discussed. In our opinion, it can be nicely illustrated on the recent developments in the field of treebanking. In [Kingsbury and Palmer 2002], a layer of predicate-argument structures was added to English Penn Treebank — they started by marking clause nuclei composed of verbal predicates and their arguments. In fact, these structures can already be viewed as small (still rather flat and isolated) dependency trees, headed by (potentially complex) verbal forms. Later, they added also ‘modifiers of event variables’, thus broadening the nuclei of dependency trees with what FGD would call free modifiers (e.g. [Babko-Malaya et al. 2004]). Then also the argument structures for instances of common nouns were added ([Meyers et al. 2004]). Finally, the isolated islands containing small dependency trees were connected with relations corresponding to subordinate and coordinating conjunctions [Mitsakaki et al. 2004], thus forming a deeper structure, extremely similar to deep-syntactic (resp. tectogrammatical) dependency trees available in MTT and FGD decades ago.

5. Final Remarks

Besides many similarities between FGD and MTT, one can find also several points in which the two approaches differ:

- Compared to MTT, FGD invested less energy into lexicological and lexicographical issues. Although for instance an electronic valency lexicon based on FGD is being intensively built now ([Straňáková-Lopatková and Žabokrtský 2002]), no studies fully comparable to [Mel’čuk and Žolkovskij 1984] were created in the context of FGD.
- MTT and FGD elaborated valency theories (e.g. [Panevová 1980] for FGD), but in spite of the fact that they both refer back to [Tessnière 1959], their criteria for distinguishing actants from modifiers substantially differ.
- Both approaches prefer different representation of coordination constructions — FGD represents coordination by adding a new dimension to the tectogrammatical tree (by using different type of

⁵ In PDT, also textual coreference is added into the tectogrammatical annotation.

«bracketing»),⁶ whereas in MTT the coordination members are attached one below another.

Finally, it should be mentioned that this paper is written from a humble perspective of a computer scientist interested in Natural Language Processing, with a deep respect to, but with only a limited insight into the two compared approaches. Due to this fact, some parallels outlined in this paper might be inaccurate, which is of course completely our own fault.

Acknowledgements

The presented paper would not be written without the financial support of the project MSM0021620838.

References

- [Apresian et al. 2003] Jurij Apresian, Igor Boguslavsky, Leonid Iomdin, Alexander Lazursky, Vladimir Sannikov, Victor Sizov, and Leonid Tsinman. 2003. ETAP-3 Linguistic Processor: a Full-Fledged NLP Implementation of the MTT. MTT 2003. First International Conference on Meaning-Text Theory. Paris, 2003.
- [Babko-Malaya et al. 2004] O. Babko-Malaya, M. Palmer, N. Xue, A. Joshi, and S. Kulick. 2004. Proposition Bank II: Delving Deeper. In A. Meyers, editor, HLT-NAACL 2004 Workshop: Frontiers in Corpus Annotation. P. 17—23. Boston, Massachusetts, USA, May 2—7. Association for Computational Linguistics.
- [Bolshakov and Gelbukh 2000] I. A. Bolshakov and A. F. Gelbukh. 2000. The Meaning-Text Model: Thirty Years After. International Forum on Information and Documentation.
- [Cinková 2004] Silvie Cinková. 2004. Manuál pro tektogramatickou anotaci angličtiny. Technical report, UFAL/CKL MFF UK, Prague, 2004.
- [Hajič et al. 2001] Jan Hajič, Eva Hajičová, Petr Pajas, Jarmila Panevová, Petr Sgall, and Barbora Vidová-Hladká. 2001. Prague Dependency Treebank 1.0 (Final Production Label). CDROM CAT: LDC2001T10, ISBN 1-58563-212-0.
- [Hajič et al. 2004] Jan Hajič, Otakar Smrž, Petr Zemánek, Jan Šnaidauf, and Emanuel Beška. 2004. Prague Arabic Dependency Treebank: Development in Data and Tools. In Proceedings of the NEMLAR International Conference on Arabic Language Resources and Tools, pages 110—117, September 22—23, 2004.

⁶ Coordination is treated differently in PDT 2.0: as it is illustrated in Figure 2, there is a special coordination node (functor CONJ), the children of which are the members of coordination.

- [Kingsbury and Palmer 2002] Paul Kingsbury and Martha Palmer. 2002. From Treebank to PropBank. In *Proceedings of Third International Conference on Language Resources and Evaluation*.
- [Mel'čuk and Žolkovskij 1984] Igor A. Mel'čuk and Alexander K. Žolkovskij. 1984. *Explanatory Combinatorial Dictionary of Modern Russian*. Wiener Slawistischer Almanach, Vienna.
- [Mel'čuk 1988] Igor A. Mel'čuk. 1988. *Dependency Syntax: Theory and Practice*. State University of New York Press.
- [Meyers et al. 2004] A. Meyers, R. Reeves, C. Macleod, R. Szekely, V. Zielinska, B. Young, and R. Grishman. 2004. The NomBank Project: An Interim Report. In A. Meyers, editor, *HLT-NAACL 2004 Workshop: Frontiers in Corpus Annotation*. P. 24—31, Boston, Massachusetts, USA, May 2—7. Association for Computational Linguistics.
- [Miltasakaki et al. 2004] E. Miltasakaki, A. Joshi, R. Prasad, and B. Webber. 2004. Annotating Discourse Connectives and Their Arguments. In A. Meyers, editor, *HLT-NAACL 2004 Workshop: Frontiers in Corpus Annotation*. P. 9—16, Boston, Massachusetts, USA, May 2—7. Association for Computational Linguistics.
- [Nakhimovsky 1990] Alexander Nakhimovsky. 1990. *Word Meaning and Syntactic Structure: Some Comparative Notes*. In *Meaning-Text Theory: Linguistics, Lexicography, and Implications*. University of Ottawa Press.
- [Panevová 1980] Jarmila Panevová. 1980. *Formy a funkce ve stavbě české věty*. Academia, Praha.
- [Sgall et al. 1986] Petr Sgall, Eva Hajičová, and Jarmila Panevová. 1986. *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. D. Reidel Publishing Company, Dordrecht.
- [Sgall1967] Petr Sgall. 1967. *Generativní popis jazyka a česká deklinace*. Academia, Prague, Czech Republic.
- [Sgall 1997] Petr Sgall. 1997. Valency and underlying structure. An Alternative View on Dependency. In L. Wanner, editor, *Recent Trends in Meaning-Text Theory*, pages 149—166. Amsterdam/Philadelphia: Benjamins.
- [Straňáková-Lopatková and Žabokrtský 2002] Markéta Straňáková-Lopatková and Zdeněk Žabokrtský. 2002. Valency Dictionary of Czech Verbs: Complex Tectogrammatical Annotation. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002)*. Vol. 3. P. 949—956.
- [Tesnière 1959] Lucien Tesnière. 1959. *Éléments de syntaxe structurale*. Paris.
- [Žolkovskij and Mel'čuk 1965] Aleksandr Žolkovskij and Igor Mel'čuk. 1965. *O vozmožnom metode i instrumentax semantičeskogo sinteza*. Naučno-texničeskaja informacija.

Сведения об авторах

Margarita Alonso Ramos is Professor of Linguistics at the Faculty of Philology, University of La Coruña. She received a doctoral degree from the Universidad Nacional de Educación a Distancia (UNED), Madrid and a doctoral degree from the University of Montreal, both in linguistics. Dr. Alonso Ramos's research fields include Explanatory Combinatorial Lexicology and Lexicography, lexical semantics, the relation between lexicon and grammar, and didactical aspects in language teaching. She has been extensively working on the problems related to the notion of lexical functions, publishing in 2004 a book on the Oper/Func/Labor-family of LFs (the so-called support verb constructions). Currently, she is leading a project on the compilation of a collocation dictionary of Spanish, which is based on the principles of the Explanatory Combinatorial Lexicology.

Маргарита Алонсо Рамос — профессор лингвистики факультета филологии Университета Ля Корунья. Она получила докторскую степень по лингвистике в Национальном университете дистанционного образования (UNED) в Мадриде и такую же степень в Монреальском университете. Области научных интересов д-ра Алонсо Рамос включают толково-комбинаторную лексикологию и лексикографию, отношение между словарем и грамматикой, а также дидактические аспекты обучения языку. Она много работала над проблемами, связанными с понятием лексических функций, и в 2004 году опубликовала книгу о семействе ЛФ Oper/Func/Labor (так называемых конструкциях с глаголами поддержки). В настоящее время руководит проектом создания словаря сочетаний испанского языка, основанного на принципах толково-комбинаторного словаря.

Dominic Anctil, Département de didactique, OLST — Département de linguistique et traduction, Université de Montréal.

After completing a Bachelor's degree in jazz piano at the Université de Montréal, Dominic Anctil started linguistics studies in 2002. Fascinated by this new perspective on languages, he has become passionate about semantics and lexicology. Member of the Observatoire de Linguistique Sens-Texte (OLST) since 2003, he has worked on different lexicographical projects based on the Meaning-Text Theory. Interested particularly by linguistic applications in language teaching, Dominic Anctil is now finishing a Master's degree in French teaching. His multiple experiences as a research auxiliary and teaching assistant for French and semantics courses reinforced his desire to become a university professor. Lexicology and vocabulary teaching constitute his primary fields of interest in research.

Доминик Анктиль, Отделение дидактики и Обсерватория лингвистики «Смысл — Текст» отделения лингвистики и перевода Монреальского уни-

верситета. Получив степень бакалавра искусств по классу джазового фортепьяно в Монреальском университете, Доминик в 2002 году начал изучать лингвистику. Вдохновленный новым взглядом на природу языка, он страстно увлекся семантикой и лексикологией. Будучи членом Обсерватории лингвистики «Смысл — Текст» с 2003 года, он работал в различных лексикографических проектах, основанных на модели «Смысл — Текст». Особый интерес у Доминика Анктиля вызывает применение лингвистики к преподаванию языка. Опыт лаборанта-исследователя и преподавателя-ассистента по французскому языку и семантике усилили в Доминике стремление стать университетским профессором. Основной областью его научных интересов является обучение студентов лексикологии и лексике.

Apresjan Juri. Born: 02.02.1930, Moscow, Russia. Education: Moscow Foreign Languages Institute, English Department. Candidate's degree: 1958, English Philology.

Doctor's degree: 1984, General Linguistics. Titles: Full member (academician) of the Russian Academy of Sciences. Doctor honoris causa of the Moscow University. Doctor honoris causa of the Sophia University. Awards: Humboldt's award, Germany; the Vladimir Dal Gold medal of the Russian Academy of Sciences. Affiliations: principal researcher of the Institute for information transmission problems of RAS; head of theoretical semantics section of the Vinogradov Russian Language Institute of RAS. Research Domains: Syntax, semantics, lexicography, formal models of language, machine translation. Major publications: 11 dictionaries (in collaboration), 10 monographs (4 of them in collaboration), over 260 papers.

Юрий Дереникович Апресян родился 2 февраля 1930 г. в Москве. Окончил английское отделение Московского института иностранных языков. В 1958 г. защитил кандидатскую диссертацию по английской филологии, а в 1984 г. — докторскую диссертацию по общему языкознанию. Действительный член Российской академии наук, почетный доктор Московского государственного университета, почетный доктор Софийского университета. Награды: премия им. Гумбольдта, Германия; Золотая медаль им. Владимира Даля Российской академии наук. Главный научный сотрудник Института проблем передачи информации РАН, заведующий сектором теоретической семантики Института русского языка им. В. В. Виноградова РАН.

Научные интересы: синтаксис, семантика, лексикография, формальные модели языка, машинный перевод. Основные публикации: 11 словарей (в соавторстве), 10 монографий (из них 4 в соавторстве), свыше 260 статей.

Olga Babina is an assistant professor, Department of Linguistics and Cross-Cultural Communication, and a post-graduate student in computational linguistics, Southern Ural State University (Russia). Recipient of grants for post-graduate students from Federal Education Agency and Ministry of Education and Science and from the Chelyabinsk Region Government. Current research is de-

voted to designing and acquiring the knowledge for natural language processing, such as generation and information retrieval in technical domains.

Ольга Бабина — доцент Отделения языкознания и межкультурной коммуникации Южноуральского государственного университета и аспирантка этого же университета по специальности «компьютерная лингвистика». Получатель грантов для аспирантов от Федерального агентства образования и Министерства образования и науки, а также от Челябинской областной администрации. В настоящее время занимается исследованиями в области конструирования и сбора знаний для автоматической обработки текстов, в частности, для порождения текстов и информационного поиска в технических предметных областях.

Lucie Barque received a master degree in computational linguistics and aims for the Ph.D in formal linguistics, all from the Université Paris 7. Her work, directed by Sylvain Kahane (Modico, Université Paris 10) and Alain Polguère (OLST, Université de Montréal), consists in proposing a typology of polysemic relations in French.

Люси Барк получила степень магистра в области компьютерной лингвистики в Университете Париж 7 и сейчас работает над диссертацией по формальной лингвистике на соискание степени доктора философии в том же университете. Ее работа, которой руководят Сильвен Каан (Modico, Университет Париж 10) и Ален Польгер (OLST, Монреальский университет) состоит в разработке типологии отношений многозначности во французском языке.

Pascale Bélanger received a Master of Linguistics from the Département de Linguistique et Traduction of the Université de Montréal in 2003. Her master's thesis presented an MTT-based examination of condensation processes over a corpus of French author-written abstracts dealing with botanical research. She is currently working as a linguist and a team leader at Nstein Technologies Inc. (Montréal), a public company providing software solutions for analyzing vast amounts of multilingual unstructured data. Her interest area is computational linguistics and natural language processing, especially summarization, categorization and automatic translation processes of textual material.

Паскаль Беланже получила степень магистра лингвистики в Отделении лингвистики и перевода Монреальского университета в 2003 году. Ее диссертация представляла собой выполненное в русле теории «Смысл — Текст» процессов сжатия текста на корпусе авторских аннотаций статей на французском языке, посвященных исследованиям в области ботаники. В настоящее время работает в качестве лингвиста и руководителя группы в акционерной компании Nstein. Technologies Inc (Монреаль), специализирующейся на компьютерных системах анализа обширных массивов многоязычных неструктурированных данных. Область ее научных интересов — компью-

терная лингвистика и автоматическая обработка текстов, в особенности аннотирование, категоризация и автоматический перевод текстового материала.

Boguslavskaya Olga Jurjevna. Studied linguistics at the Moscow State University, Department of Philology (1967—1972). Ph.D. in Structural and Applied Linguistics & Caucasian Linguistics from the Moscow State University (1989). In 1972—1990 worked in the Department of Philology of the Moscow State University. Researcher. Research in linguistic processors for natural language communication, Russian syntax and semantics, description of unwritten Daghestanian languages, syntactic typology. From 1990 worked in the sector of Theoretical Semantics at the Russian Language Institute of the Russian Academy of Sciences, Moscow. Development of New Russian Explanatory Dictionary of Synonyms. Research in theoretical linguistics and lexicography.

Publications: over 80 in Russian, 2 in English, 1 in Japanese, including 3 books (written with co-authors), published in 1982, 1986 and 1995; more than 50 dictionary entries for the New Russian Explanatory Dictionary of Synonyms (1997, 2000, 2003, 2004).

Богуславская Ольга Юрьевна. Окончила отделение структурной и прикладной лингвистики филологического факультета МГУ им. М. В. Ломоносова. Защитила кандидатскую диссертацию по двум специальностям: структурная и прикладная лингвистика и кавказское языкознание в МГУ в 1989 г. С 1972 г. по 1990 г. работала научным сотрудником на кафедре структурной и прикладной лингвистики филологического ф-та МГУ. Занималась разработкой лингвистических процессоров, русским синтаксисом и семантикой, синтаксической типологией, полевой лингвистикой (дагестанские языки). С 1990 г. работает в секторе теоретической семантики Института русского языка им. В. В. Виноградова, ст. научный сотрудник. Занимается теоретической лингвистикой и лексикографией. Имеет более 80 научных работ. Является соавтором трех монографий, опубликованных в 1982, 1986 и 1995 гг., написала более 50 словарных статей для Нового объяснительного словаря синонимов русского языка. Готовит монографию, посвященную причинно-следственной и целевой лексике (совместно с И. Б. Левонтиной).

Igor Boguslavsky graduated from the Department of Structural and Applied Linguistics of the Moscow State University in 1972. From 1973 to 1985 he worked in the “Informelectro” institute in the Ju. Apresjan’s laboratory. Since 1985 until now he works in the Computational linguistics laboratory of the Institute for Information Transmission Problems, RAS (from 1995 he is the head of this lab). From 2003 he works at the Madrid Polytechnic University within the Ramon y Cajal program. He has been a visiting professor in Gothenburg, Vienna and Konstanz. His research interests center on semantics and syntax in their interaction, scope theory, and computational linguistics. His publications include monographs “Research in syntactic semantics. Scope of logical words” (1985), “Scope of lexical items” (1996), “Linguistic description in the ETAP-2 machine

translation system” (1989, with co-authors), “Linguistic processor for complex information systems” (1992, with co-authors).

Богуславский И. М. закончил Отделение структурной и прикладной лингвистики Московского государственного университета в 1972 г. С 1973 по 1985 г. он работал в институте «Информэлектро» в лаборатории Ю. Д. Апресяна. С 1985 г. по настоящее время — в лаборатории Компьютерной лингвистики Института проблем передачи информации РАН (с 1995 г. — заведующий этой лабораторией). С 2003 г. работает в Мадридском политехническом университете в рамках программы Ramon y Cajal. Он преподавал как приглашенный профессор в Гетеборге, Вене и Констанце. Его основные научные интересы — семантика и синтаксис в их взаимодействии, теория сфер действия, компьютерная лингвистика. Автор книг «Исследования по синтаксической семантике. Сферы действия логических слов» (1985) и «Сфера действия лексических единиц» (1996). Соавтор книг «Лингвистическое обеспечение в системе ЭТАП-2» (1989) и «Лингвистический процессор для сложных информационных систем» (1992).

Krystyna Bojalkowska, born 1978, Master of Arts in Polish Philology, Post-graduate student of linguistics at Institute of Polish Language at Nicolaus Copernicus University in Toruń, Poland.

Research interests: grammar, syntax.

Кристина Боялковска родилась в 1978 г. Магистр польской филологии. Аспирантка Института польского языка Университета им. Николая Коперника (Торунь, Польша). Исследовательские интересы: грамматика, синтаксис.

Bolshakov Igor Alexeevich was born in 1934 in Moscow. Graduated from Physical Faculty of Moscow State University (1956), Doctor of Sciences in Technology (1966), Professor (1968), Laureate of State Prize of USSR (1988). He worked in the enterprises of radioelectronic, electronic, and electrical engineering industries, as well as at the institutes of the Russian Academy of Sciences. Lecturer of Moscow Institute of Electronic Technologies, Moscow State University, and Russian State University of the Humanities. Author of the largest database of Russian collocations. More than 250 publications on statistical processing of signals, applied statistics, theoretical and computational linguistics, rendered by more than 1000 citations. The recent years he is a researching professor at the National Polytechnic Institute, Mexico City, the highest excellence level member of the Mexican National System of Researchers.

Большаков Игорь Алексеевич родился в 1934 г. в Москве. Окончил физический факультет МГУ (1956), доктор технических наук (1966), профессор (1968), лауреат Государственной премии (1988). Работал на предприятиях радиотехнической, электронной и электротехнической промышленности (1956—1981) и в академических институтах (1981—1998). Препо-

давал в МИЭТе, МГУ и РГГУ. Создатель крупнейшей электронной базы словосочетаний русского языка. Имеет более 250 печатных трудов по статистической обработке сигналов, прикладной статистике, теоретической и компьютерной лингвистике, получивших более 1000 ссылок. В последние годы работает профессором-исследователем в Национальном Политехническом Институте (г. Мехико), член (высшего уровня отличия) Национальной Системы Исследователей Мексики.

Alexander Dikovsky, born 09.08.1945 in Leningrad (St.-Petersburg), Russia, is the author of more than 100 works in Formal Grammars Theory, Logic Programming, Nonmonotonic Reasoning, Behavior of Distributed Systems, Formal Syntax, Semantics and Grammatical Induction.

Professor Dikovsky received his Ph.D. at the Steklov Institute of Mathematics (Leningrad Branch) and his Doctoral degree in Physics and Mathematics at the St.-Petersburg State University in 1992. He held research positions at the academic Institute for Pure Mathematics in Novosibirsk and at the Keldysh Institute for Applied Mathematics in Moscow. He taught at Novosibirsk and Tver Universities, the Russian State University for the Humanities, Moscow State University, Bilkent University (Turkey), Paris-12 University (France). Since 1998 he is a Professor of Computer Science at Nantes University (France).

Current scientific interests: dependency grammars, natural language semantics and acquisition, and modelling of cognitive processes.

Александр Диковский родился 9 августа 1945 в Ленинграде (ныне Санкт-Петербург). Автор более 100 работ в области теории формальных грамматик, логического программирования, немонотонных выводов, поведения распределенных систем, формального синтаксиса, семантики и грамматической индукции.

Проф. Диковский получил степень кандидата наук в Ленинградском отделении Математического института им. В. А. Стеклова и степень доктора физ.-мат. наук в академическом Институте математики в Новосибирске и в Институте прикладной математики им М. В. Келдыша в Москве. Занимался исследовательской работой в Институте математики в Новосибирске и в Институте прикладной математики им М. В. Келдыша. Преподавал в Новосибирском и Тверском университетах, Российском государственном гуманитарном университете, университете Билкент (Турция), университете Париж-12 (Франция). С 1998 года — профессор информатики в Нантском университете (Франция).

Его нынешние научные интересы включают грамматики зависимостей, семантику, а также проблемы овладения естественным языком и моделирование когнитивных процессов.

Dmitrij Dobrovol'skij is Senior Researcher at the Russian Language Institute of the Russian Academy of Sciences (Moscow) and professor at the Moscow State University (German department). His main field of research is lexical semantics in

Russian and German with the focus on figurative language and idioms in particular. His publications include monographs such as *Kognitive Aspekte der Idiom-Semantik* (1995), *Idiome im mentalen Lexikon* (1997), *Figurative language: cross-cultural and cross-linguistic perspectives* (2005 — with Elisabeth Piirainen), numerous papers published in Russian, German, English, Spanish and French, as well as some dictionaries mostly compiled in co-authorship with other colleagues. At present he is a participant in the Austrian Academy Corpus project at the Austrian Academy of Sciences in Vienna.

Добровольский Д. О. — ведущий научный сотрудник Института русского языка РАН и профессор факультета иностранных языков МГУ. Основные сферы научных интересов — лексическая семантика и фразеология русского и немецкого языков. Среди основных публикаций — монографии *Kognitive Aspekte der Idiom-Semantik* (1995), *Idiome im mentalen Lexikon* (1997), *Figurative language: cross-cultural and cross-linguistic perspectives* (2005 — в соавторстве с Э. Пириайнен), статьи на русском, немецком, английском, испанском и французском языках, а также ряд словарей, написанных в соавторстве. В настоящее время Д. О. Добровольский участвует в проекте «Корпус Австрийской академии» в Австрийской академии наук в Вене.

Ronald F. Feldstein, Professor and Chair, Department of Slavic Languages and Literatures, Indiana University, Bloomington, IN Ph.D., Slavic Languages and Literatures, Princeton University, 1973.

«On the Structure of Syncretism in Romanian Conjugation», in *Contemporary Approaches to Romance Linguistics*, Benjamins, 2004. P. 177—195; «The Unified Monophthongization Rule of Common Slavic», *Journal of Slavic Linguistics* 11: 2, 2003. P. 251—285; *Polish* (with Steven Franks). Lincom, Languages of the World/Materials Series, 2002; «Czyta vs. Czytaj and the Determination of Polish Conjugational Desinences», *International Journal of Slavic Linguistics and Poetics* 35—36 (1987). P. 65—78; «The Inverse Proportionality of Segmental and Suprasegmental in the Russian Verb», *Slavic and East European Journal*, 31: 4, 1987. P. 582—594.

Рональд Ф. Фелдстин, профессор, заведующий отделением славянских языков и литератур университета Индианы, г. Блумингтон. Доктор филологии (Славянские языки и литературы, Принстонский университет, 1973). Некоторые публикации: «О структуре синкретизма в румынском спряжении», в кн. «Современные подходы к романскому языкознанию», Benjamins, 2004. С. 177—195; «Унифицированное правило монофтонгизации в общеславянском языке», *Journal of Slavic Linguistics* 11:2, 2003. С. 251—285; «Польский язык» (в соавторстве с Стивеном Франком), Lincom, Languages of the World/Materials Series, 2002; «Czyta vs. Czytaj и определение польских финалей спряжения», *International Journal of Slavic Linguistics and Poetics* 35—36 (1987). С. 65—78; «Обратная пропорциональность сегментного и супraseгментного в русском глаголе», *Slavic and East European Journal*, 31: 4, 1987. С. 582—594.

Gavrilova V. I have graduated from MSU Philological Faculty Department of applied and structural linguistics. I have had my post-graduate course at the same speciality at VINITI, where I have been working at different departments for a long time. For about ten years I worked at the research centre in the Russian Ministry of Foreign Affairs. At present time I work at The Moscow City Pedagogical University. The theme of my post-graduate work was "Quasipassive constructions in Russian language". All the time I have been mostly interested in the Russian verb and its functioning in texts describing different physical and mental areas and interaction of such universals categories as transitivity, reflexivity, and passivity.

Гаврилова В. Окончила Отделение структурной и прикладной лингвистике филологического факультета МГУ им М. В. Ломоносова. Окончила аспирантуру по специальности структурная и прикладная лингвистика при ВИНТИ АН СССР. В течение длительного времени работала в различных отделах этого института. Около десяти лет проработала в НИЦИ (Научно-исследовательском центре по информатике) при МИД России. В настоящее время является доцентом романо-германского факультета МГПУ (Московского городского педагогического университета). Защитила кандидатскую диссертацию на тему «Квазипассивные конструкции в русском языке». Основной сферой научных интересов всегда был русский глагол, его функционирование в текстах, описывающих разные физические и ментальные пространства, и взаимодействие, в которое при этом вступают такие универсальные присущие русскому глаголу категории как переходность, возвратность и страдательность.

Kim Gerdes. Maître de Conférence (associate professor) in linguistics at the University Michel Montaigne (Bordeaux 3). Current works in the field of syntax and computational linguistics.

Studies of Mathematics, Philosophy, Japanese, and Linguistics in Berlin, Paris, and Utrecht.

1996 — Masters in Pure Mathematics, University Paris XI. 1998 — Masters in Linguistics, University Paris 7. 2002 — Ph. D. in Linguistics, University Paris 7.

Ким Гердес, доцент, преподает лингвистику в Университете им. Мишеля Монтеня (Бордо, Франция). Изучал математику, философию, японский язык и лингвистику в университетах Берлина, Парижа и Утрехта. Степень магистра математики получил в 1996 году в университете Париж 11, магистра лингвистики — в 1998 году в университете Париж 7, доктора филологии — в 2002 году в университете Париж 7.

A. V. Gladky. Born in 1928. Originally was a «pure» mathematician by education and scientific interests. Taught at pedagogical institutes and universities, and at high school. Got interested in mathematical linguistics in the late 1950s and took up linguistic self education under consulting support from active linguists (V. V. Ivanov recommended the first list of references, then I. A. Melchuk was giving essential advice). Was engaged mostly in the mathematical aspects of

the theory of formal grammars till the late 1960s, and the book «Formal grammars and languages» published in 1973 was the principal outcome. Later his scientific interests gradually moved towards purely linguistic problems, including the theory of syntax structures, the theory of case, the semantics of «logical» words in Russian. The current position is the leading scientist at the Moscow Institute of Open Education.

Гладкий А. В. Родился в 1928 г. По образованию и первоначальным научным интересам «чистый» математик. Преподавал в педагогических институтах и университетах, а также в средней школе. В конце 50-х гг. заинтересовался математической лингвистикой и занялся лингвистическим самообразованием, пользуясь консультациями лингвистов (первый список литературы для самообразования получил от В. В. Иванова, в дальнейшем особенно важны были консультации И. А. Мельчука). До конца 60-х гг. занимался преимущественно математическими аспектами теории формальных грамматик (итогом исследований в этой области была вышедшая в 1973 г. книга «Формальные грамматики и языки»), но постепенно научные интересы сместились в сторону собственно лингвистики (теория синтаксических структур, теория падежа, семантика «логических» слов русского языка). В настоящее время ведущий научный сотрудник Московского института открытого образования.

Hi-Yon Yoo. Maître de Conférence (associate professor) in linguistics and phonetics at the University Paris 7 — Denis Diderot 2003 Ph.D. in Linguistics, University Paris 7. Current works in the field of phonetics, syntax-phonology interface, communicative structure of Modern Greek, Korean, and French.

Хийон Ю. Доцент, преподаватель лингвистики и фонетики в университете Париж 7 (Дени Дидро). Степень доктора филологии получила в 2003 г. в университете Париж 7. Работы в области фонетики, интерфейса между синтаксисом и фонологией, коммуникативной структуры новогреческого, корейского и французского языков.

Leonid Iomdin was born on September 1, 1947. Graduated in 1969 from Novosibirsk State University, Department of Mathematical Linguistics, diploma cum laude. Ph. D. in structural and applied linguistics & Russian language from Moscow State University (1982). Since 1985 works in the Laboratory of Computational Linguistics, Institute for Information Transmission Problems, Russian Academy (presently as acting head of laboratory). Main topics include development of machine translation systems (English-to-Russian & vice versa) and multipurpose NLP processors, development of deeply annotated text corpora. Worked in universities of Novosibirsk, Moscow, Munich, Saarbrücken, Essex. Since 2004 gives regular lecture courses in linguistic foundations of natural language processing at the Russian State University of the Humanities. Research interests lie in the field of theoretical linguistics (esp. syntax and semantics). Has more than 80 publications, including three books and a dictionary.

Иомдин Леонид Лейбович родился 1 сентября 1947 г. Окончил с отличием Отделение математической лингвистики Новосибирского государственного университета (НГУ). В 1982 г. в Московском государственном университете им. М. В. Ломоносова (МГУ) защитил кандидатскую диссертацию. С 1985 года работает в Лаборатории компьютерной лингвистики Института проблем передачи информации Российской академии наук (в настоящее время — исполняющий обязанности заведующего лабораторией). Занимается разработкой систем машинного перевода (англо-русский, русско-английский перевод) и многоцелевых лингвистических процессоров, созданием глубоко аннотированных корпусов текстов. Работал в НГУ, МГУ, университетах Мюнхена, Саарбрюккена, Эссекса. С 2004 г. читает курс лекций по лингвистическим основам автоматической обработки текстов в Российском государственном гуманитарном университете (РГГУ). Научные интересы лежат в области теоретической лингвистики (в первую очередь, синтаксиса и семантики). Автор более 80 научных работ, в том числе трех монографий и словаря.

Iordanskaja Lidija, member of Observatoire de Linguistique Sens-Texte, Department of Linguistics and Translation, University of Montreal. Born in Chirchik, UzSSR (now Uzbekistan) in 1936. Graduated in 1958 from the Moscow University; obtained the PhD degree in 1970 from the Institute of Linguistics, Academy of Sciences of the USSR (*Automatic Syntactic Text Processing*). Junior Research Fellow at the Institute of Linguistics, Academy of Sciences of the USSR, Moscow, till 1973, Senior Research Fellow at Informelektro Institute till 1976. In 1977 emigrated to Canada. Associate Research Fellow at the Department of Linguistics and Translation, University of Montreal (1979—1998), Research Fellow in R&D companies Odyssey Research Associates (1987—1989) and CoGenTex (1990—1994). Research areas: Syntax, Semantics, Lexicography, Machine Translation, Computational Linguistics. Publications: 3 dictionaries (with co-authors), 2 books and about 40 articles.

Лидия Иорданская, член Обсерватории лингвистики «Смысл — Текст» Отделения лингвистики и перевода Монреальского университета. Родилась в Чирчике УзССР (ныне Узбекистан) в 1936 году. Окончила Московский университет в 1958 году. В 1970 г. получила степень кандидата наук в Институте языкознания АН СССР за диссертацию по автоматическому синтаксическому анализу и синтезу текстов. Была младшим научным сотрудником Института языкознания до 1973 года и старшим научным сотрудником института «Информэлектро» до 1976 г. В 1977 году эмигрировала в Канаду. Научный сотрудник Отделения лингвистики и перевода Монреальского университета (1979—1998), научный сотрудник исследовательских компаний Odyssey Research Associates (1987—1989) и CoGenTex (1990—1994). Круг научных интересов — синтаксис, семантика, лексикография, машинный перевод, вычислительная лингвистика. Опубликовала 3 словаря (в соавторстве), 2 книги и около 40 научных статей.

Maarten Janssen, born 28.01.1971 in Nijmegen, The Netherlands. Ph.D. in Linguistics /Artificial Intelligence, 1997—2002: «SIMuLLDA: a Multilingual Lexical Database using a Structured Interlingua». Worked at the ERSS in Toulouse, France in 2003 on automatic knowledge extraction using Lexicon-Semantic Pattern Extraction, responsible for the YakwaSI corpus query and extraction tool. Currently working at ILTEC in Lisboa, Portugal. Responsible for the MorDebe lexical database and the NeoTrack system for semiautomatic neologism detection.

Мартен Янссен родился 28 января 1971 года в Ниймегене (Нидерланды). Имеет степень доктора по лингвистике и искусственному интеллекту. Диссертация посвящена системе SIMuLLDA (многоязыковая лексическая база данных, использующая структурированную интерлингву), над которой автор работал в 1997—2002 гг. В 2003 году работал в группе ERSS (Тулуза, Франция), где занимался проблемами автоматического извлечения знаний с помощью механизма извлечения знаний по словарно-семантическим образцам (этот механизм лежит в основе системы обращения к корпусам текстов YakwaSI). В настоящее время работает в Институте теоретической и компьютерной лингвистики в Лиссабоне (Португалия). Занимается лексической базой данных MorDebe и системой для полуавтоматического обнаружения неологизмов NeoTrack.

Sylvain Kahane is Professor of Linguistics at Université Paris 10 — Nanterre. He holds a PhD in Mathematics from Université Paris 6 (on Descriptive Set Theory and its applications to Harmonic Analysis) and was Associate Professor of Mathematics at Université Paris 10 for 10 years. He joined the Talana-Lattice team of Université Paris 7 in 1996 and is now at Modyco at Université Paris 10. He began to publish in linguistics in 1996 and he intensively collaborated with Igor Mel'cuk in 1997 on the mathematical foundations of the Meaning-Text Theory and the modeling of extraction. Since 2001 Sylvain Kahane is developing a mathematical model of natural language, the Meaning-Text Unification Grammar, which is based on the Meaning-Text Theory but also inspired from TAG, LFG and HPSG. His main contribution to linguistics concerns a modelling of extraction and coordination using bubble trees instead of standard dependency trees, a formalization of the concepts of equative and transductive correspondence grammars (opposed to generative ones) and a modelling of word order, with Kim Gerdes, based on the traditional topological model for German and applied to German, French and Arabic.

Сильвен Каан — профессор лингвистики в Университете Париж 10 — Нантер. Получил степень доктора наук по математике в Университете Париж 6 (за работу по дескриптивной теории множеств и ее применении к гармоническому анализу) и в течение 10 лет занимал должность доцента математики в Университете Париж 10. В 1996 году вошел в состав группы Talana-Lattice Университета Париж 10, а в настоящее время работает в лаборатории Modyco Университета Париж 10. Он начал публиковать работы

по лингвистике в 1996 году, а в следующем, 1997 году интенсивно сотрудничал с И. Мельчуком, работая над математическими основами теории «Смысл — Текст» и моделями извлечения знаний. С 2001 г. Сильвен Каан разрабатывает математическую модель естественного языка, в основе которой лежит теория «Смысл — Текст», однако используются и достижения теории TAG, LFG и HPSG. Его основной вклад в лингвистику касается моделирования извлечения знаний и сочинения с использованием пузырьковых деревьев вместо стандартных деревьев зависимостей, формализации понятий эквативных и трансдуктивных грамматик соответствия (в противоположность порождающим грамматикам), а также, в соавторстве с Кимом Гердесом, моделирования порядка слов на основе традиционной топологической модели для немецкого языка применительно к немецкому, французскому и арабскому языкам.

Andrej A. Kibrik. Graduated from Moscow State University in 1984. Got the degree of Cand. Sc. in the Institute of Linguistics RAN in 1988. Got the degree of Doctor of Science in 2003. Now heads the working group «Languages of the World» in the Institute of Linguistics RAN. Teaches linguistics at MSU. Areas of research: cognitive linguistics, discourse analysis, structure of spoken language, linguistic diversity, native American languages, Athabaskan languages.

Кибрик А. А. Окончил МГУ в 1984 г. Окончил аспирантуру в Институте языкознания РАН и защитил кандидатскую диссертацию в Институте языкознания РАН в 1988 г. Степень доктора филологических наук — в 2003 г. Руководитель группы «Языки мира» в Институте языкознания РАН. Преполагает лингвистические курсы в МГУ. Области научных интересов: когнитивная лингвистика, анализ дискурса, структура устной речи, языковое разнообразие, языки индейцев Сев. Америки, атабаскские языки.

Alexander Evgenyevich Kibrik was born in 1939 in Leningrad. In 1961, he graduated from the Department of Classical Philology of the Philological Faculty of the Moscow State University. Since 1961, he has been working at the Chair of Theoretical and Applied Linguistics of this faculty (earlier, it was called the Chair of Structural and Applied Linguistics). He became head of this Chair in 1992.

Alexander Kibrik is a recognized leader in the domain of Russian field linguistics and the author of a unique technique of team field work. Since 1967, he has headed numerous linguistic expeditions organized by the Moscow University and supervised studies of over forty languages of the former Soviet Union. He participated in the creation of grammars of seven minority languages and in the two-volume presentation of comparative studies of the lexicon and the grammar of 23 languages of Daghestan. Alexander Kibrik is also working in the field of language theory and typology, advocating the cognitive approach to the language. He is the author of over 250 research works, including 40 books. The following publications deserve special mention: *Methods of Field Studies (on Problem Statement)*. Moscow, Moscow University Press, 1972. (In Russian.); *Essays on Gen-*

eral and Applied Issues of Linguistics (The Universal, the Typical and the Specific in Language). Moscow, Moscow University Press, 1992. (In Russian.); Constants and Variables of the Language. St. Petersburg, Aleteya, 2003. (In Russian.)

Кибрик Александр Евгеньевич. Родился в 1939 г. в Ленинграде. Окончил отделение классической филологии филологического факультета МГУ в 1961 году. С 1961 г. по настоящее время работает на кафедре теоретической и прикладной (ранее — структурной и прикладной) лингвистики филологического факультета МГУ. С 1992 года заведует кафедрой.

А. Е. Кибрик является признанным лидером в области отечественной полевой лингвистики, автор уникального метода коллективной полевой работы. Он руководит лингвистическими экспедициями МГУ с 1967 года, под его руководством исследовалось более сорока языков бывшего СССР. При его участии созданы грамматики семи малочисленных языков, а также двухтомное сопоставительное исследование лексики и грамматики 23-х языков Дагестана. А. Е. Кибрик работает также в области теории языка и типологии, разрабатывает когнитивно-ориентированный подход к языку. Он автор более 250 научных работ, в том числе сорока монографий. Особо можно отметить следующие издания: Методика полевых исследований (к постановке проблемы). М.: Изд-во МГУ, 1972; Очерки по общим и прикладным вопросам языкознания (универсальное, типовое и специфичное в языке). М.: Изд-во МГУ, 1992; Константы и переменные языка. СПб.: Алетейя, 2003.

Richard Kittredge is Professor of Linguistics in the Département de Linguistique et Traduction at the Université de Montréal. He received a Ph.D. in Formal Linguistics from the University of Pennsylvania in 1969 and directed Montreal's TAUM project in machine translation, including development of TAUM-METEO and AVIATION systems during 1971—1976. During the 1980s and 1990s his research and publications were mainly on sublanguage analysis, and on text generation applied to reports in English and French sublanguages. He was Visiting Scientist at the Australian CSIRO research organization in 2000. Currently, Professor Kittredge serves on the editorial boards of the journal *Machine Translation*, and John Benjamins' book series on Natural Language Processing. He is also affiliated with CoGenTex, Inc. (www.cogentex.com), a company he co-founded in 1990 for research and development in text generation and machine translation.

Ричард Киттредж — профессор лингвистики Отделения лингвистики и перевода Монреальского университета. Он получил степень доктора наук по формальной лингвистике в Пенсильванском университете в 1969 г. и руководил проектом машинного перевода TAUM в Монреале (включая разработку систем TAUM-METEO и AVIATION в 1971—1976 гг.) В 1980—1990 гг. его исследования и публикации в основном касались анализа подязыков, а также порождения текстов применительно к отчетам на английском и французском подязыках. В 2000 г. он был приглашенным исследователем

в научно-исследовательской организации CSIRO (Австралия). В настоящее время профессор Киттредж входит в состав редакционных коллегий журнала "Machine Translation" и книжной серии по автоматической обработке естественного языка издательства John Benjamins. Он также работает в компании CoGenTex, Inc. (www.cogentex.com), сооснователем которой является. Компания была основана в 1991 г. и занимается исследованиями и разработками в области порождения текстов и машинного перевода.

Mikhail V. Kopotev.

Date of birth: September 27, 1973.

Address: P.O. Box 24 (Unioninkatu 40 B), 00014 University of Helsinki, Finland. Tel.: (+3589) 191 22 028.

Position: Research fellow, Department of Slavonic and Baltic Languages and Literatures, University of Helsinki.

Academic preparation: postgraduate program in Russian and General linguistics (Petrozavodsk State University, Russia), postgraduate program in Russian language (in progress, University of Helsinki, Finland).

Research interests: (Old) Russian syntax, idiomacity, computer linguistics.

Михаил Вячеславович Копотев.

Дата рождения: 27 сентября 1973 г.

Адрес: P.O. Box 24 (Unioninkatu 40 B), 00014 University of Helsinki, Finland. Телефон: (+3589) 191 22 028.

Место работы и должность: исследователь, Отделение славянских и балтийских языков и литератур, Хельсинкский университет.

Академическая подготовка: аспирантура по общему и русскому языкознанию (Петрозаводский государственный университет, Россия), докторантура по русскому языку (по настоящее время, Хельсинкский университет, Финляндия).

Исследовательские интересы: (древне)русский синтаксис, фразеология, компьютерная лингвистика.

Sergey Alexandrovich Krylov (born on July, 13, 1958, in Moscow) is a Russian linguist living in Moscow. Senior researcher of the Institute of Eastern Studies of the Russian Academy of Sciences (Department of languages of Asia and Africa) since 1984. Since 2004, he is an assistant professor at the University of Nature, Society and Man (Dubna) and at the same time head of the Chair of general and Slavic linguistics of the Russian St. John the Evangelist Orthodox University. Ph.D. degree in 2002. Sergey Krylov graduated from the Department of Structural and Applied Linguistics of the Moscow State University and took post-graduate courses in the Institute of Linguistics (1983). Major publication: Theoretical Grammar of Mongolian and Related Issues of General Linguistics. Part 1, Morphemics, morphonology, elements of phonological transformatomics (in the light of the general theory of morphological and morphonological models). 269 papers on general linguistics, semantics, phonology, history of linguistics, semiotics, computational linguistics, Russian and Mongolian.

Крылов Сергей Александрович (р. 13 июля 1958, Москва) — русский лингвист, живет в Москве. С.н.с. Ин-та востоковедения РАН (Отдел языков Азии и Африки), доц. (с 2004) Университета Природы, общества и человека (г. Дубна), зав. кафедрой общего и славянского языкознания Российского Православного университета Иоанна Богослова (с 2004). К. ф. н. (с 2002). Окончил МГУ (фил. ф-т) в 1980 по спец. «структурная и прикладная лингвистика» и аспирантуру Ин-та языкознания АН СССР (1983) по спец. «общее языкознание». С 1984 — в ИВ РАН. Кн. «Теоретическая грамматика монгольского языка и смежные вопросы общей лингвистики. Ч. 1. Морфемика, морфонология, элементы фонологической трансформаторики (в свете общей теории морфологических и морфонологических моделей)» и 269 публикаций по общей лингвистике, грамматике, семантике, фонологии, истории языкознания, семиотике, компьютерной лингвистике, русскому и монгольскому языкам.

François Lareau is about to finish his Ph. D. thesis at the University of Montreal and the University Paris 7, co-directed by professors Igor Mel'čuk and Sylvain Kahane. The subject of his thesis is French verbal inflection in the framework of Meaning — Text Unification Grammar. His main research interests lie in theoretical and formal linguistics, text generation and lexicology. After working a few years for the University of Montreal as a research assistant on the DiCo, a formal lexical database, he is now turning to the development of Meaning-Text grammars for a text generation system at the University Pompeu Fabra in Barcelona.

Франсуа Ларо заканчивает докторскую диссертацию в Монреальском университете и Университете Париж 7 под руководством профессоров Игоря Мельчука и Сильвена Каана. Тема диссертации — французское глагольное словоизменение в рамках унификационной грамматики типа «Смысл — Текст». Основные научные интересы лежат в области теоретической и формальной лингвистики, порождения текста и лексикологии. В течение нескольких лет он работал научным сотрудником Монреальского университета, разрабатывая формальную лексическую базу данных DiCo, а в настоящее время начинает разрабатывать грамматики типа «Смысл — Текст» для системы порождения текста в Университете Помпеу Фабра в Барселоне.

Lennart Lönngren. I am a Swede working in Norway. My «cradle» is the University of Uppsala, where I got my Ph.D. in 1970 and taught Russian for many years. Since 1993, I am a professor of Russian linguistics at the University of Tromsø. My research interests include contemporary Russian morphology, syntax and semantics.

Леннарт Лённгрен. Я швед, работаю в Норвегии. Моей колыбелью был Уппсальский университет, в котором я получил степень доктора наук (1970) и много лет преподавал русский язык. В настоящее время я профессор университета Тромсе, преподаю русскую лингвистику. В область моих научных интересов входят современная русская морфология, синтаксис и семантика.

Leonardo Lesmo is Full Professor of Man-Machine Interaction at the Dipartimento di Informatica of the University of Torino. His research activity is concerned with Natural Language Processing, ontologies and agent-based models of communication. He is Vice-President of the Italian Association for Artificial Intelligence, member of the Steering Committee of the Italian Association of Cognitive Sciences and is responsible of the Computer Science Unit within the Interuniversity Center for Theoretical and Applied Ontology. Currently, he is the local coordinator of a National Project aimed at the development of legal ontologies and is technical responsible for the Turin unit within the HOPS European project, having as goal the development of natural language interfaces (written and spoken) towards the public administration services.

Леонардо Лесмо — профессор, занимается человеко-машинным взаимодействием в Отделении информатики Туринского университета. Его исследовательские интересы касаются автоматической обработки текстов, онтологий и агентных моделей коммуникации. Он является вице-президентом Итальянской ассоциации искусственного интеллекта, членом Управляющего комитета Итальянской ассоциации когнитивных наук и возглавляет отдел информатики Межуниверситетского центра теоретической и прикладной онтологии. В настоящее время является локальным координатором Национального проекта разработки юридических онтологий и техническим руководителем Туринского отделения Европейского проекта HOPS, целью которого является разработка естественно-языковых интерфейсов (письменных и голосовых) для общественно-административных органов.

Letuchiy A. B. was born May 4, 1981. Graduated from the Department of Linguistics of Russian State University for Humanities in 1999 (M. A. «Labile (ambitransitive) verbs in accusative languages: semantics and argument structure»). His personal research focuses on linguistic typology, syntax in the framework of formal theories of language, semantics, and nonverbal semiotics. He has attended the international conference «Case, valency and transitivity» (Nijmegen, the Netherlands, 2003), and the Third Winter Typological School (Moscow, 2002). Since 2001 he has taken part in the fieldtrips organized by RSUH to the Hakas and Adyghe spoken areas. Currently he is admitted to the graduate program of the RSUH Linguistic Department.

Летучий А. Б. родился 4 мая 1981 г. Выпускник Института лингвистики РГГУ 1999 г. (дипломная работа на тему «Лабильные глаголы в аккумулятивных языках: семантика и актантная структура», научный руководитель — Н. Р. Сумбатова). Научные интересы: лингвистическая типология, синтаксис и способы его представления в различных моделях, семантика, невербальная семиотика, полевая лингвистика. Участвовал в международной конференции «Case, valency and transitivity» (2003 г., Неймеген, Нидерланды), а также в Третьей зимней типологической школе (2002 г., Москва). С удовольствием принимал участие в экспедициях РГГУ в Республику Хакасия и Республику Адыгея (2001—2004 гг.). Сейчас аспирант

Центра лингвистической типологии ИЛ РГГУ, научный руководитель — В. А. Плунгян.

Levontina Irina Borisovna (1963), PhD, Russian Language Institute of the Russian Academy of sciences (Vinogradov institute), one of the authors of the New Explanatory Dictionary of Russian Synonyms, under the guidance of Ju. D. Apresjan; specialist in semantics and lexicography, ethnolinguistics, linguopoetics, forensic linguistic examination, the author of popular articles and radio show on the Russian language.

Левонтина Ирина Борисовна (1963), кандидат филологических наук, старший научный сотрудник Института русского языка им. В. В. Виноградова РАН, один из авторов «Нового объяснительного словаря синонимов русского языка» (под общим руководством акад. Ю. Д. Апресяна), автор работ по лексической семантике, лексикографии, этнолингвистике и языку поэзии, специалист по судебной лингвистической экспертизе, автор популярных статей о русском языке, один из авторов программы «Грамотей» (радио «Маяк»).

Olga Lyashevskaja graduated from the Russian State University for the Humanities, Department of Theoretical and Applied Linguistics in 1995. In 1999 she defended Ph.D. thesis «Non-standard Semantics of Russian Nominal Number». Currently she is senior research fellow in the Department of Linguistic Research, All-Russian Institute for Scientific and Technical Information (VINITI RAN), Moscow.

Ляшевская Ольга Николаевна в 1995 г. окончила факультет теоретической и прикладной лингвистики РГГУ. В 1999 г. защитила кандидатскую диссертацию на тему «Нестандартное числовое поведение русских существительных». В настоящее время является старшим научным сотрудником отдела лингвистических исследований Всероссийского института научной и технической информации (ВИНИТИ РАН), Москва.

Mel'čuk Igor, Professor of Linguistics at the University of Montreal (since 1977), Member of the Royal Society of Canada, Corresponding Member of the Austrian Academy of Sciences, Doctor Honoris Causa of the Univ. of Besançon (France), born 19.10.1932 in Odessa, now Ukraine. Graduated in 1956 from the Moscow University (Spanish linguistics); obtained the PhD degree in 1962 from the Russian language Institute of the Academy of Sciences of the USSR (*Theoretical Problems of Automatic Syntactic Analysis of Texts*). Till 1976, Junior, then Senior Research Fellow at the Institute of Linguistics, Academy of Sciences of the USSR, Moscow (was expelled for political reasons and had to emigrate). Recipient of several scientific awards (1988—1990, Killam Scholarship, Canada; 1990, Guggenheim Fellowship, USA; 1991, Alexander von Humboldt Research Award, Germany; 1997, Award of the French Canadian Association for Advancement of Sciences). Nominated at the International Chairs at Collège de France

(1996—1997) and École Normale Supérieure (2002—2004, Paris). Publications: 34 books, 206 articles. Hobbies : reading (history and memoirs), skiing, hiking.

Игорь Мельчук, профессор лингвистики Монреальского университета с 1977 г., член Королевского общества Канады, член-корреспондент Австрийской академии наук, почетный доктор Безансонского университета (Франция). Родился 19 октября 1932 г. в Одессе (ныне Украина). В 1956 году окончил Московский университет по специальности «испанская филология». Получил степень кандидата наук в Институте русского языка АН СССР в 1962 г. (диссертация о проблемах автоматического синтаксического анализа текстов). До 1976 г. — младший, затем старший научный сотрудник Института языкознания АН СССР (был изгнан по политическим причинам и вынужденно эмигрировал). Лауреат нескольких научных премий (1988—1990 — Килламская стипендия, Канада; 1990 — стипендия Гуттенхайма, США; 1991 — Исследовательская стипендия фонда Александра фон Гумбольдта, Германия; 1997 — награда Франко-канадской ассоциации за достижения в науке). Отмечен наградами в Коллеж де Франс (1996—1997) и Эколь Нормаль (2002—2004, Париж). Опубликовал 34 книги и 206 статей. Увлечения: чтение (исторические произведения, мемуары), лыжи, туристические походы.

Irina Mikaelian, Ph.D., University of Provence, France, 2002; graduated from Moscow State University in 1982. Initially specialized in Romance (French and Italian) Languages, she taught French at The Russian State University of Humanities (RGGU), Moscow. Subsequently, she spent ten years in France (1994—2004) teaching the Russian language and related subjects such as History of Russian Languages, Advanced Russian Grammar and Syntax, Translation from and into Russian at the Universities of Provence, of Lyon II, and of Grenoble III. Since 2004, she lives in the USA and teaches Russian at the Penn State University (University Park). Her main research interests are Russian and general linguistics with a particular emphasis on syntax and semantics within cognitive and comparative approach.

Ирина Микаэлян закончила романо-германское отделение МГУ в 1982 г. С 1991 по 1994 гг. преподавала французский язык на Факультете теоретической и прикладной лингвистики РГГУ. С 1994 по 2004 год провела во Франции. Преподавала русский язык, а также курсы истории русского языка, теоретической грамматики, синтаксиса русского языка и перевода в университетах Экс-ан-Прованса, Лиона и Гренобля. В 2002 г. защитила диссертацию по русской и общей лингвистике в университете Экс-ан-Прованса под руководством проф. Маргерит Гиро-Вебер. В настоящее время живет в США и преподает русский язык в университете штата Пенсильвания (Penn State University, University Park). Ее интересуют когнитивный и сравнительный аспект в изучении синтаксиса и семантики русского языка.

Jasmina Milićević got her Ph. D in linguistics from the Université de Montréal. She is a member of the research group *Observatoire de linguistique Sens-*

Texte at that same university. Currently, she works as an Assistant Professor of Linguistics at the French Department of the Dalhousie University in Halifax, Canada. Her research interests are in semantics, lexicology, morphology and second-language teaching.

Ясмينا Миличевич получила степень доктора филологии в Монреальском университете. Она является членом исследовательской группы «Обсерватория лингвистики «Смысл — Текст»» этого университета. В настоящее время — доцент Отделения французского языка Университета Дальхаузи (Галифакс, Канада), преподает лингвистику. Ее исследовательские интересы включают семантику, лексикологию, морфологию и преподавание иностранного языка.

Alexis Nasr is assistant professor in Computational Linguistics at University Paris 7. He obtained his PhD from University Paris 7 on the design of a computational model for the Meaning Text Theory and its application to automatic text reformulation. Presently he is engaged in research on probabilistic parsing.

Алексис Наср — доцент, преподает компьютерную лингвистику в Университете Париж 7. Степень доктора филологии получил в Университете Париж 7 за создание компьютерной модели для теории Смысл — Текст и ее применение к задаче автоматического перифразирования текста. В настоящее время занимается исследованиями в области вероятностного синтаксического анализа.

Professor **Elena Viktorovna Paducheva** is a Doctor of Science, a Foreign member of the American Academy of Arts and Sciences, member of the European Linguistic society and member of the Grammatical Commission of the International Committee of Slavists.

She graduated from the Moscow State University. Since 1957, she has been working in the All-Russian Institute of Scientific and Technical Information of the Russian Academy of Sciences. Her research activities included mathematical methods in language description, machine translation, transformational syntax of Russian, logical analysis of language, semantics of grammatical tenses and aspects, prosody, and communicative structure of the sentence. In 1974, she published the book “On the Semantics of Syntax” devoted to the problem of syntactic structure contribution to the meaning of the sentence. In 1985, her second book “The Utterance and its Relation with the Reality” was published: it was focused on the theory of reference and pragmatic aspects of the language. Her book “Semantic Studies” (1996) was devoted to the spoken and narrative modes of discourse interpretation and the semantics of the narrative. Her recent book “Dynamic Models in the Semantics of the Lexicon” reflects the author’s experience in working with databases of Russian verbal semantics.

She taught and lectured at universities of Finland, Germany, Italy, Sweden, the USA, and Switzerland. She gives regular courses of lectures on theoretical semantics and computer dictionaries at the Philological Faculty of the Moscow

State University and the Institute of Linguistics of the Russian State University for the Humanities.

Падучева Елена Викторовна — доктор филологических наук, профессор, иностранный член Американской академии наук и искусств, член Европейского лингвистического общества, член Грамматической комиссии Международного комитета славистов.

Окончила Московский университет. С 1957 г. работает во Всероссийском институте научной и технической информации РАН. Занималась применением математических методов в описании языка, машинным переводом, трансформационным синтаксисом русского языка, логическим анализом языка, семантикой грамматических форм времени и вида, просодией и коммуникативной структурой предложения. В 1974 г. опубликовала книгу «О семантике синтаксиса», посвященную проблеме описания вклада синтаксической структуры в смысл предложения; в 1985 г. — книгу «Высказывание и его соотносительность с действительностью» о теории референции и прагматических аспектах языка. Книга «Семантические исследования» (1996) посвящена речевому и нарративному режиму интерпретации дискурса и семантике нарратива; книга «Динамические модели в семантике лексики» отражает опыт работы автора с базами данных по семантике русского глагола.

Преподавала и выступала с лекциями в университетах Финляндии, Германии, Италии, Швеции, США, Швейцарии. Периодически читает курсы лекций по теоретической семантике и компьютерным словарям на филологическом факультете МГУ и в Институте лингвистики РГГУ.

Anna Sergeyevna Panina graduated from the Faculty of Theoretical and Applied Linguistics of the Russian State University for the Humanities. Since 1998, she has been working in the group developing YARAP, a system of Japanese-Russian machine translation. She got her Ph.D. degree in 2003 for the thesis "Problems of Lexical Ambiguity in Translating from Japanese into Russian". The main research activities include theory of translation and lexical semantics of Japanese with regard to natural language processing problems.

Панина Анна Сергеевна окончила факультет теоретической и прикладной лингвистики РГГУ (Москва); с 1998 г. работает в Институте востоковедения РАН в составе группы, разрабатывающей систему японско-русского автоматического перевода ЯРАП. В 2003 г. защитила кандидатскую диссертацию на тему: «Проблемы лексической неоднозначности при переводе с японского языка на русский». Основное направление научной деятельности — исследования в области теории перевода и лексической семантики японского языка в связи с задачами автоматической обработки текста.

Барбара Х. Парти изучала синтаксис под руководством Н. Хомского в аспирантуре Массачусетского Технологического Института (MIT, Ph. D. 1965). С 1965 по 1972 она преподавала в Калифорнийском университете в

Лос-Анджелесе (UCLA) и начала заниматься семантикой после появления работ Р. Монтегю (R. Montague). Ее основные результаты в это время — синтез хомскианского синтаксиса и семантики Монтегю, это привело к появлению формальной семантики в лингвистике. Ее основные научные интересы — формальная семантика и ее связи с синтаксисом, прагматикой, логикой, философией и когнитивной наукой. С 1972 г. она преподает в Массачусетском университете, Амхерст, (с 2004 г. — Professor Emerita). Она преподавала как приглашенный профессор в Мексико-Сити, Праге, Лейпциге и Москве.

Барбара Парти является почетным доктором четырех университетов (включая РГГУ). Она член Национальной Академии Наук США и Американской Академии Искусств и Наук и иностранный член Королевской Нидерландской Академии Искусств и Наук. В 1986 г. она была Президентом Американского Лингвистического Общества.

Barbara H. Partee studied syntax with Chomsky at MIT (Ph.D. 1965), then moved into semantics after exposure to Montague's work at UCLA, where she taught from 1965 to 1972. Her major early contributions involved the synthesis of Chomskyan syntax with Montague's semantics, bringing formal semantics into linguistics. Her research interests center on formal semantics and its connections with syntax, pragmatics, logic, philosophy, and cognitive science. Partee has taught at the University of Massachusetts since 1972, and has been Emerita since 2004. She has been a Visiting Professor in Mexico, Prague, Leipzig, and Moscow.

Partee has received four honorary doctorates, including one from RGGU. She is a fellow of the National Academy of Sciences and the American Academy of Arts and Sciences, and a Foreign Member of the Royal Netherlands Academy of Arts and Sciences. She was President of the Linguistic Society of America in 1986.

Alain Polguère is, since 1995, professor at the Department of Linguistics and Translation of the Université de Montréal. Previously, he taught for four years lexicology, computational linguistics and general linguistics at the Department of English Language and Literature of the National University of Singapore. His main areas of research are in lexicology, lexicography, formal semantics, natural language processing and applied linguistics. He is the head of the research group Observatoire de linguistique Sens-Texte (OLST).

Ален Польгер с 1995 г. является профессором Отделения лингвистики и перевода Монреальского университета. До этого он в течение четырех лет преподавал лексикологию, компьютерную лингвистику и общее языкознание в Отделении английского языка и литературы Национального университета Сингапура. Его основные научные интересы — лексикология, лексикография, формальная семантика, автоматическая обработка текстов на естественном языке и прикладная лингвистика. Он возглавляет исследовательскую группу «Обсерватория лингвистики “Смысл — Текст”».

Amanda Pounder is currently Associate Professor of German and Adjunct Associate Professor of Linguistics at the University of Calgary, Canada. Originally a student of Romance Philology at the Université de Montréal, she owes her conversion to morphology to her former professor, I. A. Mel'čuk. She has an M. A. from the Universität Trier and a Dr. Phil. from the Universität Wien. Her research focuses on theoretical morphology and Germanic historical linguistics.

Аманда Паундер — доцент немецкого языка и адъюнкт-доцент лингвистики в Университете Калгари, Канада. Во время обучения в Монреальском университете на отделении романской филологии она заинтересовалась проблемами морфологии, чем обязана своему бывшему профессору И. А. Мельчуку. Получила степень магистра искусств в Трирском университете и степень доктора наук в Венском университете. Ее исследовательские интересы включают теоретическую морфологию и историческую германистику.

Vladimir Alexandrovich Plungian was born in 1960 in Moscow. After graduating from the Department of structural and applied linguistics of the Moscow State University he joined the Department of African languages of the Institute of Linguistics, Russian Academy of Sciences. He is currently heading the Typology Section of the same institute and teaches in the Moscow State University as well as in the Moscow State University of the Humanities. Dr. Sci. (Philology) (1998). He is the author of over 160 publications, including 6 books. Plungians' major research interests include general morphology and typology of grammar categories.

Плунгян Владимир Александрович родился в 1960 г. в Москве. После окончания Отделения структурной и прикладной лингвистики МГУ им. М. В. Ломоносова поступил на работу в Отдел африканских языков Института языкознания РАН; в настоящее время заведует сектором типологии того же института, а также преподает в МГУ и РГГУ. Доктор филологических наук (1998). Автор более 160 научных работ, в том числе 6 монографий. Основная область исследований — общая морфология и типология грамматических категорий.

Ekaterina Vladimirovna Rakhilina was born in 1958 in Moscow. After graduating from the Department of structural and applied linguistics of the Moscow State University he joined the VINITI Research Institute, Russian Academy of Sciences. She is currently heading the Department of linguistic studies there and teaches at the Institute of Linguistics of the Moscow State University of the Humanities. Dr. Sci. (Philology) (2000). She is the author of over 120 publications, including 3 books. Her major research interests include general and Russian semantics and lexicology.

Екатерина Владимировна Рахилина родилась в 1958 г. в Москве. После окончания Отделения структурной и прикладной лингвистики МГУ им. М. В. Ломоносова поступила на работу в ВИНТИ РАН, где в настоящее

время заведует Отделом лингвистических исследований; также преподает в Институте лингвистики РГГУ. Доктор филологических наук (2000). Автор более 120 научных работ, в том числе 3 монографий. Основная область исследований — общая и русская семантика и лексикология.

Tilmann Reuther, *1953, Klosterneuburg, Austria. Attended school and university in Vienna, M.A. in Russian language and literature and mathematics, PhD in Russian linguistics. Associate professor at the University of Klagenfurt, Austria. Lecturer at the University of Vienna and the Vienna Business School. Visiting scholar at the Université de Montréal and Tbilisi State University. Editor of the journal *Wiener Slawistischer Almanach* and its linguistic monograph series. Former president of the Austrian Association of University Teachers. Main linguistic interests: semantics, lexicography, grammar (Russian and other Slavic languages). Personal interests: music, art, nature.

Тильманн Ройтер, род. в 1953 в г. Клостернойбург, Австрия. Окончил гимназию и университет в Вене, диплом по специальностям «русский язык и литература», и «математика», кандидат и доктор наук по специальности «русистика». Доцент Клагенфуртского университета, Австрия. Преподает также в Венском университете и в Венском экономическом институте. Научные стажировки в Монреальском университете и в Тбилисском государственном университете. Издатель научного журнала *Wiener Slawistischer Almanach* и серии лингвистических монографий. Долгие годы — председатель Австрийской ассоциации вузовских преподавателей. Основные лингвистические интересы: семантика, лексикография, грамматика (русский и другие славянские языки). Личные интересы: музыка, искусство, природа.

Livio Robaldo is a Ph. D Student in Computer Science at the Dipartimento di Informatica of the University of Torino. His research activity is concerned with Natural Language Processing and Ontologies. In particular, his work is focused on Semantic Underspecification and on Logical Representations which take into account cognitive aspects of Natural Language. He is member of the Italian Association for Artificial Intelligence. Currently, he is a part of the team of the HOPS European project.

Ливьо Робальдо — аспирант Отделения информатики Туринского университета. В область его исследовательских интересов входит автоматическая обработка текстов и онтологии. Его работа, в частности, касается семантической недоопределенности, а также логических представлений, учитывающих когнитивные аспекты естественного языка. Он является членом Итальянской ассоциации искусственного интеллекта. В настоящее время входит в состав рабочей группы Европейского проекта HOPS.

Elena Rudnitskaya (born in 1968) graduated from the MGU, Department of Structural and Applied Linguistics (OSIPL/OTIPL), in 1990. She finished her

Candidate dissertation on sentence and manner adverbs in Russian (in the Moscow Semantics School framework) in 1993, the MGU, OSIPL. From 1994, Elena has been working at the Department of Languages at the Institute of Oriental Studies, Moscow. She has around 15 publications on Korean grammar and around 10 conference presentations on this topic. In 2000, Elena finished her PhD dissertation on long-distance reflexive pronouns in Russian (in the Generative Grammar framework), the City University of New-York. She has around 15 publications on Russian lexical semantics and generative syntax and around 10 conference presentations on this topic.

Елена Рудницкая родилась в 1968 году. В 1990 году закончила Отделение структурной и прикладной лингвистики (ОСИПЛ/ОТИПЛ) Московского государственного университета. В 1993 году в аспирантуре этого же отделения завершила работу над диссертацией о сентенциальных наречиях и наречиях образа действия в русском языке (в русле Московской семантической школы). С 1994 года работает в Отделе языков Института восточных исследований в Москве. Елена — автор около 15 статей по корейской грамматике и около 10 выступлений на конференциях на эту же тему. Получила степень доктора филологии в Городском университете Нью-Йорка за диссертацию о дистантных возвратных местоимениях в русском языке (в русле порождающей грамматики). Ей принадлежит около 15 публикаций по русской лексической семантике и порождающему синтаксису и около 10 докладов на конференциях, посвященных данной проблематике.

Zygmunt Saloni, born 1938, Professor of Polish (grammar, lexicography) at the University of Varmia and Masuria, Olsztyn, Poland. Main publications: *Cechy składniowe polskiego czasownika* (Syntactic Features of Polish Verbs), 1976; (with M. Świdziński) *Składnia współczesnego języka polskiego* (Syntax of Contemporary Polish), 1981, 5th ed. 2001; *Czasownik polski* (Polish Verb), 2001.

Зигмунт Салёни родился в 1938 г. Профессор-полонист, преподает в Варминско-Мазурском университете (Ольштын, Польша). Основные труды: *Cechy składniowe polskiego czasownika* (Синтаксические свойства польского глагола), 1976; *Składnia współczesnego języka polskiego* (Синтаксис современного польского языка), 1981, 5 изд. 2001, в соавторстве с М. Свидиным; *Czasownik polski* (Польский глагол), 2001.

Begoña Sanromán graduated in Hispanic Filology at UNED (Spain). She worked in the University of La Coruña (Spain) as a member of the research project “Base de datos léxico-semánticos: unidades léxicas descriptivas y no descriptivas” led by Dr. Margarita Alonso Ramos. She doctorated at the University of Helsinki with the dissertation *Semántica, sintaxis y combinatoria léxica de los nombres de emoción en español*. At the moment she is working as university lecturer at the University of Helsinki.

Бегонья Санроман закончила отделение испанской филологии в Национальном университете дистанционного образования (UNED), Испания.

Она работала в университете Ла Корунья (Испания) в составе исследовательской группы, разрабатывающей под руководством д-ра Маргариты Алонсы Рамос проект «база лексико-семантических данных: дескриптивные и недескриптивные лексические единицы». В Хельсинкском университете она защитила докторскую диссертацию на тему «Семантика, синтаксис и лексическая сочетаемость существительных со значением эмоций в испанском языке». В настоящее время преподает в Хельсинкском университете.

Serdobolskaya Natalya is a typologist, focusing mainly on non-finite constructions and raising in the languages of the world. For a number of years, she has been taking part in the fieldtrips to the minority languages of Russia, mostly of the Uralic family (Mari, Komi-Zyrian, and Udmurt). Serdobolskaya graduated from the Department of Linguistics, Moscow State University, and has been in the post-graduate three-year program at the Departement. She has recently presented her Ph.D. thesis "Syntactic status of the arguments of non-finite predicates".

Сердобольская Н. В. родилась 25 августа 1979 г. Закончила отделение Теоретической и Прикладной Лингвистики филологического факультета МГУ в 2001 г. Начиная с 2000 г. принимает активное участие в лингвистических экспедициях МГУ и РГГУ в респ. Марий Эл, Коми, Удмуртия и Адыгея.

Сердобольская Наталья Вадимовна — типолог, в сферу ее интересов входят проблемы приписывания падежного оформления, синтаксиса и семантики нефинитной предикации в языках мира. В настоящее время Н. В. Сердобольская закончила аспирантуру, недавно защитила диссертацию на тему «Синтаксический статус актантов зависимой нефинитной предикации».

Zoya Mikhailovna Shalyapina graduated from the Department of Applied Linguistics, Maurice Thorez Moscow State Pedagogical Institute of Foreign Languages in Moscow. In 1968—1977 she worked in the Laboratory of Machine Translation of this Institute, where she headed the group developing the linguistics of the ARAP English-Russian MT system. Ph.D. in 1975 (Thesis on the principles of organization and recording semantic data in an automatic dictionary. Since 1977, Zoya Shalyapina has been heading the group developing YARAP, a system of Japanese-Russian machine translation. The main research activities include theoretical, methodological and experimental studies in the formalization of linguistic competence and performance in machine translation (on the material of Japanese, Russian and English).

Зоя Михайловна Шаляпина. Окончила Отделение прикладной лингвистики МГПИИЯ им. М. Тореза (Москва); в 1968—1977 гг. работала в Лаборатории машинного перевода МГПИИЯ, с 1970 г. — как руководитель группы по разработке лингвистического обеспечения системы англо-русского автоматического перевода АРАП. В 1975 г. защитила кандидатскую диссертацию на тему: «Принципы организации и записи семантической информации в автоматическом словаре». С 1977 г. работает в Инсти-

туте востоковедения РАН, возглавляет группу, разрабатывающую систему японско-русского автоматического перевода ЯРАП.

Основное направление научной деятельности — теоретические, методологические и экспериментальные исследования по формализации языковой компетенции человека и языковой деятельности при переводе (на материале японского, русского и английского языков).

Svetlana Sheremetyeva is a Doctor of Sciences (Habilitation) in Computational Linguistics (from St. Petersburg State University.) Has a considerable teaching and research experience accumulated as Professor and Chair of the Foreign Languages Department of the Southern Ural State University (Russia), as a key researcher at the Computing Research Lab of New Mexico State University (USA), as a lecturer in computational linguistics at Uppsala University (Sweden) and Copenhagen Business School (Denmark), as a managing director of LanA Consulting A/S, Denmark (www.lanaconsult.com). Worked on a number of large-scale projects devoted to the development of Mikrokosmos ontology, machine translation, multilingual text generation, information retrieval, and text summarization. Developed a system for automatic generation of patent claims implemented for English, a novel method for morphological analysis without a stem dictionary implemented in broad-coverage analyzers for Russian and Serbo-Croatian.

Светлана Шереметьева — доктор филологических наук, получила степень в Санкт-Петербургском университете по специальности «компьютерная лингвистика». Имеет большой опыт преподавания и научной работы — в качестве профессора, заведующего кафедрой иностранных языков Южноуральского государственного университета, ведущего научного сотрудника в лаборатории компьютерных исследований Университета штата Нью-Мексико (США), лектора по компьютерной лингвистике в Уппсальском университете (Швеция) и Копенгагенской школы бизнеса (Дания), управляющего директора компании LanA Consulting A/S, (www.lanaconsult.com, Дания). Работала в ряде широкомасштабных проектов, посвященных онтологии (Микрокосмос), машинному переводу, порождению многоязычных текстов, поиску информации и аннотированию текста. Разработала систему автоматического порождения патентных заявок, реализованную для английского языка, а также новый метод морфологического анализа без словаря основ, реализованный в анализаторах широкого покрытия для русского и сербохорватского языков.

Alexei Shmelev, Doctor of Letters in Russian Linguistics, Professor of Russian Linguistics at the Philological Department of Moscow Pedagogical University, Supervisor of a research department at the Institute of the Russian Language (Russian Academy of Sciences). Main research interests: linguistic and semantic theory and their applications, manipulation in Russian political journalism, key-concepts of Russian culture, cross-cultural communication. Authored and co-

authored books include: *Matematika pomogaet lingvistike* (co-authored with G. Krejdlin), 1994; *Referencial'nye mexanizmy russkogo jazyka*, 1996; *Jazykovaja konceptualizacija mira: na materiale russkoj grammatiki* (co-authored with T. Bulygina), 1997; *Russkaja jazykovaja model' mira*, 2002; *Russkij jazyk i vnejazikovaja dejstvitel'nost'*, 2002; *Russkij anekdot: tekst i rechevoj zhanr* (co-authored with Elena Shmeleva), 2002 as well as two books on Russian aspectology co-authored with Anna Zalizniak.

Алексей Шмелев, доктор филологических наук, профессор русистики филологического факультета Московского педагогического университета, руководитель исследовательского сектора в Институте русского языка РАН. Основные научные интересы: лингвистическая и семантическая теория и ее приложения, манипуляции в российской политической журналистике, ключевые концепты русской культуры, межкультурная коммуникация. В числе книг, написанных единолично и в соавторстве — «Математика помогает лингвистике» (в соавторстве с Г. Крейдлиным, 1994), «Референциальные механизмы русского языка» (1996), «Языковая концептуализация мира на материале русской грамматики» (в соавторстве с Т. Булыгиной, 1997), «Русская языковая модель мира» (2002), «Русский язык и внеязыковая действительность» (2002), «Русский анекдот: текст и речевой жанр» (в соавторстве с Еленой Шмелевой, 2002), а также две книги по русской аспектологии в соавторстве с Анной Зализняк.

Boris Smilga, born 05.07.1978 in Moscow. Studied at the Russian State University for the Humanities and at the University of Nantes, where he received his M.S. in Computer Science in 2004 and currently continues his studies as a post-graduate student. Between 2000 and 2003 he worked in the USA, including one year at the Shannon Laboratory (AT&T Research).

Борис Смилга родился 5 июля 1978 г. в Москве. Учился в Российском государственном гуманитарном университете и в Нантском университете, в котором получил степень магистра информатики (2004 г). В настоящее время учится в аспирантуре Нантского университета. В 2000—2003 гг. работал в США (в том числе в течение года — в лаборатории Шеннона (AT&T Research)).

Vitali Specht was born on June 6, 1976 in Merke (Kasakhstan). In October 1997 — April 2003 he studied computer sciences at the University of Stuttgart, Germany. Since Aug. 2003 he is a research assistant at the University of Stuttgart. His research interests focus on machine translation and parsing.

Виталий Шпехт родился 6 июня 1976 в Мерке (Казахстан). С октября 1997 по апрель 2003 г. изучал информатику в Штутгартском университете (Германия). В настоящее время — научный сотрудник Штутгартского университета. Его научные интересы включают машинный перевод и автоматический анализ текстов.

Jacques Steinlin is a student at the University of Paris 7. For the last two years, his work has been mainly centered around computational tasks with the DiCo lexical database these last two years. At the same time, he is working on his thesis in the field of Natural Language Generation focusing on the issues of generating collocations using the Lexical Functions.

Жак Стенлен — студент университета Париж 7. В последние два года его работа касалась компьютерных задач в рамках лексической базы данных DiCo. Одновременно Жак работает над диссертацией о порождении естественного языка с упором на порождение словосочетаний с помощью лексических функций.

Vladimir Trub was born in 1944. He graduated from the Department of the structural and applied linguistics, Faculty of Philology of the Kyiv State University in 1966. For more than 20 years he worked in the Department of information systems of the Institute of Cybernetics of the Ukrainian Academy of Sciences. His Ph.D thesis «The peculiarities of semantic interpretation of utterance (communicative, presuppositional, referential and truthfulness aspects)» was completed in 1994.

Since 1995 Vladimir Trub has been working as a senior scientific fellow of the Department of social linguistics in the Institute of Ukrainian language of the National Academy of Sciences of Ukraine.

His scientific interests include the problems of linguistic semantics and social linguistics. His works appeared in the «Logical analysis of the language» series and in the Proceedings of the Dialogue International conferences. Vladimir Trub is one of the authors of the Russian-Ukrainian dictionary for business documentation (Київ, «Основи», 1995) in which the apparatus of lexical functions is used to reflect idiomatic combinatorics of words and their translational equivalents.

Труб Владимир Михайлович родился в 1944 г. В 1966 г. закончил Отделение структурной и прикладной лингвистики филологического факультета Киевского госуниверситета. Более 20 лет работал в Отделе документальных информационных систем Института кибернетики АН УССР. В 1994 г. защитил кандидатскую диссертацию на тему «Особенности семантической интерпретации высказывания (коммуникативные, презумптивные, референционные и истинностные аспекты)». С 1995 г. работает старшим научным сотрудником Отдела социолингвистики Института украинского языка Национальной АН Украины. В круг его научных интересов входят проблемы лингвистической семантики и социолингвистики. Неоднократно публиковался в сборниках серии «Логический анализ языка» и Трудах международных конференций Диалог. Является одним из авторов русско-украинского словаря по деловой документации (Київ, «Основи», 1995), в котором для отражения идиоматической сочетаемости слов и их переводных эквивалентов активно используется аппарат лексических функций.

Agnès Tutin. I am professor in linguistics at the Université de Grenoble (France). My research issues deal with lexicon, semantics and NLP. I am especially interested in the treatment of collocations from a semantic and NLP viewpoint.

Агнес Тютэн, профессор лингвистики в университете Гренобля (Франция). Исследовательские интересы касаются словаря, семантики и автоматической обработки текстов. Особенный интерес вызывает обработка словосочетаний с точки зрения семантики и автоматической обработки текстов.

Elena Uryson, doctor of philological sciences, leading scientific researcher, V. V. Vinogradov's Institute of Russian Language RAS. Main works: description of more than 60 synonymic sets in: «New Explanatory Dictionary of Russian Synonyms» / General supervision by acad. Ju. D. Апресян. 2 ed., revised and updated. M. 2004. (In Russian.); «Problems of Language Model of the World Description: Analogy in semantics». M., 2004. (In Russian.)

Урысон Елена Владимировна, доктор филологических наук, ведущий научный сотрудник Института русского языка им. В. В. Виноградова РАН. Основные работы: словарные статьи более чем шестидесяти синонимических рядов в «Новом объяснительном словаре синонимов русского языка» (под общим руководством акад. Ю. Д. Апресяна. 2-е изд., испр. и доп. М., 2004); монография «Проблемы исследования языковой картины мира: Аналогия в семантике» (М., 2003).

Leo Wanner is a Research Professor at the Institució Catalana de Recerca i Estudis Avançats (ICREA) and Professor at the Technology Department of the Pompeu Fabra University, Barcelona. He holds a PhD degree in linguistics from the University of the Saarland, Germany. Before joining ICREA and UPF, he was Assistant Professor at the Intelligent Systems Institute, University of Stuttgart and Ramon y Cajal Fellow at the Applied Linguistics Institute of the Pompeu Fabra University. Dr. Wanner has extensive research experience in the field of automatic text generation, graph grammar formalisms and automatic acquisition of lexical resources from text corpora. Currently, he is leading a large scale project on the development of a multilingual multimodal document generator that is based on the Meaning-Text Theory. He has also been working on the automatic compilation and classification of lexical function material using machine learning techniques.

Лео Ваннер — профессор-исследователь в Каталонском институте углубленных исследований (ICREA) и профессор Отделения технологии Университета Помпеу Фабра в Барселоне. Степень доктора филологии получил в Университете земли Саар, Германия. До перехода на работу в ICREA и Университет Помпеу Фабра был доцентом Института интеллектуальных систем в Штутгарте и стипендиатом исследовательской программы Ramon y Cajal при Институте прикладной лингвистики Университета Помпеу Фабра. Д-р Ваннер обладает большим опытом научной работы в области автоматического порождения текста, графовых формализмов для описания

грамматики и автоматического создания лексических ресурсов из корпусов текстов. В настоящее время возглавляет масштабный проект, целью которого является разработка многоязыкового и многомодульного генератора документов на основе теории Смысл — Текст. Лео Ваннер работает также над проблемами автоматического порождения и классификации лексических функций с использованием методов машинного обучения.

Victor Samuilovich Khrakovsky was born in Leningrad on November 20, 1932, in Leningrad. He graduated from the Eastern Studies faculty of the Leningrad University and took post-graduate courses at the Chair of General Linguistics of the Philological Faculty of the same university. He is the author of over 200 publications. His research interests include theory of linguistics, typology, Russian language studies, and Arabic studies. In 1961, he joined the Group of structural and typological studies of language organized by Professor Kholodovich at the Leningrad department of the Institute of Linguistics of the Soviet Academy of Sciences. At present, this is the Laboratory of typological studies of language at the Institute of Linguistic Studies of the Russian Academy of Sciences, and Victor Khrakovsky is heading the Laboratory.

Храковский Виктор Самуилович родился 20 ноября 1932 г. в Ленинграде. Окончил Восточный факультет и аспирантуру при кафедре общего языкознания филологического факультета Ленинградского университета. Область научных интересов: теория языкознания, типология, русистика, арабистика. Автор более 200 научных работ. С 1961 г. поступил на работу в группу структурно-типологического изучения языков, организованную проф. А. А. Холодовичем в Ленинградском отделении Института языкознания Академии наук СССР. В настоящее время — это Лаборатория типологического изучения языков Института лингвистических исследований РАН, а В. С. Храковский является ее заведующим.

Tatiana Evgenjevna Yanko is a research fellow of the Department of theoretical linguistics of the Institute for linguistics in Moscow. She is the author of the book «Kommunikativnyje strategii russkoj rechi. [The communicative strategies of the Russian speech. Moscow. 2001]». Her fields of interest in linguistics are: Topic-focus structure, intonation, aspectology.

Татьяна Евгеньевна Янко, доктор филологических наук, работает в Институте языкознания в Москве. Автор монографии «Коммуникативные стратегии русской речи. Москва. 2001». Работает над проблемами коммуникативной структуры, интонации, аспектологии, лексической семантики.

Anna A. Zalizniak, graduated from the Department of Structural and Applied Linguistics, Moscow State University, in 1981. Doctor of Letters in General Linguistics (2003); leading researcher in the Department of Linguistic Typology, Institute of Linguistics, Russian Academy of Sciences. In 1994—1997 was Assistant Professor of Russian Linguistics, Institute of Slavic Philology, University

of Munich. Research interests: semantics and semantic typology; logical analysis of language, Russian aspectology; semantic mechanisms of prefixation; Russian linguistic conceptualization of the world in cross-linguistic perspective; cross-cultural communication, cognitive linguistics. Main publications: «Investigations in the Semantics of Inner State Predicates», 1992; «Vvedenie v russkuju aspektologiju» (co-authored with A. D. Shmelev), 2000; «Ključevye slova ruskoj jazykovoj kartiny mira», (co-authored with I. B. Levontina and A. D. Shmelev), 2005; «Mnogoznachnost' v jazyke i sposoby ee predstavlenija», 2005 (in print).

Анна А. Зализняк закончила отделение структурной и прикладной лингвистики МГУ в 1981 г. Доктор филологических наук по специальности «Общее языкознание» (2003), ведущий научный сотрудник Сектора лингвистической типологии Института языкознания РАН. В 1994—97 гг. преподавала в Институте славянской филологии Мюнхенского университета. Сферу научных интересов составляют: семантика и семантическая типология, логический анализ языка, русская аспектология, словообразовательная семантика (механизм префиксации), русская языковая картина мира в межкультурной перспективе, межкультурная коммуникация, когнитивная лингвистика. Основные публикации: «Введение в русскую аспектологию» (в соавторстве с А. Д. Шмелевым), 2000; «Ключевые слова русской языковой картины мира» (в соавторстве с И. Б. Левонтиной и А. Д. Шмелевым), 2005; «Многозначность в языке и способы ее представления», 2005 (в печати).

In 2001 **Zdeněk Žabokrtský** (1977) graduated from the Faculty of Electrical Engineering, Czech Technical University, Prague. Currently he is a PhD student at the Institute of Formal and Applied Linguistics, Charles University, Prague. His research interests are dependency syntax, computer lexicography, treebanking and parsing. He participates in the development of the Prague Dependency Treebank and valency lexicon of Czech verbs VALLEX.

Зденек Жабокритски (1977 г. р.) окончил факультет электротехники Чешского технического университета в Праге в 2001 г. В настоящее время — аспирант Института формальной и прикладной лингвистики Карлова университета в Праге. Его научные интересы — синтаксис зависимостей, компьютерная лексикография, синтаксически размеченные корпуса текстов и автоматический синтаксический анализ. Он участвует в разработке Пражского синтаксического размеченного корпуса текстов, использующего дерева зависимости, а также в разработке валентностного словаря чешских глаголов VALLEX.

EAST — WEST ENCOUNTER:
Second International Conference on Meaning ⇔ Text Theory.

ВОСТОК — ЗАПАД:
Вторая международная конференция по модели «Смысл ⇔ Текст»

Издатель А. Кошелев

Оригинал-макет изготовлен В. Гусевым

Художественное оформление обложки Ю. Саевича и О. Макисмовой

Художник-консультант Л. М. Панфилова

Подписано в печать 28.05.2005. Формат 60х90 ¹/₁₆.
Бумага офсетная № 1, печать офсетная. Гарнитура Times.
Усл. печ. л. 37. Тираж 400. Заказ №

Издательство «Языки славянской культуры».

ЛР № 02745 от 04.10.2000.

Phone: **207-86-93** Fax: **246-20-20** (для аб. M153)

E-mail: **Lrc@comtv.ru** Site: **<http://www.lrc-press.ru>**

*

Оптовая и розничная реализация — магазин «Гнозис».
Тел./факс: (095) 247-17-57, тел.: 246-05-48, e-mail: gnosis@pochta.ru
Костюшин Павел Юрьевич (с 10 до 18 ч.).
Адрес: Zubovskiy b-r, 2, str. 1 (Метро «Парк Культуры»)

Foreign customers may order this publication
by E-mail: koshelev.ad@mtu-net.ru
or by fax: (095) 246-20-20 (for ab. M153).