

Linguistic Well-formedness of Semantic Structures

Lidija Iordanskaja

OLST – Université de Montréal
C.P. 6128 Centre-ville
Montréal, H3C 3J7 Canada

lidija.iordanskaja@umontreal.ca

Igor Mel'čuk

OLST – Université de Montréal
C.P. 6128 Centre-ville
Montréal, H3C 3J7 Canada

igor.melcuk@umontreal.ca

Abstract

The paper formulates the problem of language-specific constraints on linguistically well-formed Semantic Structures: how can we draw a borderline between semantically well-formed and ill-formed sentences, and how can we tell lexical incongruence from semantic ill-formedness? The constraints that determine semantic ill-formedness in a given language are based on a simple criterion: If the meaning of a linguistically anomalous sentence can be expressed by a linguistically correct paraphrase, then the anomaly is not semantic; otherwise, it is semantic. Several relevant examples (in English, French and Russian) are analyzed; a sketch of a typology of semantic constraints within the Meaning-Text framework is proposed. We also define the distinction between extralinguistic and linguistic well-formedness of SemSs and propose a calculus of possible cases.

1 The Problem Stated

This paper deals, within the Meaning-Text framework, with the transition

$$\{ \text{Semantic Structures [= SemSs]} \} \Leftrightarrow \{ \text{Deep-Syntactic Structures [= DSyntSs]} \},$$

examined from one particular angle:

How to specify the linguistic well-formedness of the input set, that is, the set of SemSs for a particular language **L**?

Logically, there are two major types of well-formedness of the representational structures adopted in this framework: purely formal well-formedness (satisfying the general, language-independent conditions imposed on Meaning-Text semantic networks and trees) and linguistic well-formedness (satisfying language-specific conditions). This paper considers only **linguistic well-formedness of SemSs**.

Specifying the output set—that is, DSyntSs—has been tackled before. In this domain, Apresjan's work (1978, 1989) advanced some deep ideas, introducing two types of syntagmatic constraints on well-formedness of DSyntSs—constraints based on semantic restrictions. Apresjan showed that the absurdity of the starting SemS leads to linguistic anomaly only if the expression of a contradictory combination of meanings includes 1) contradicting grammemes/grammaticized lexemes or 2) contradicting lexemic pre-suppositions/modal frames. Here we will try to develop his ideas as applied to the level of SemSs—namely, language-specific syntagmatic constraints on combinations of semantemes, that is, on Sem-configurations. (A **semanteme** is, roughly speaking, the signified of a full lexical unit [= LU] of given language **L**.)

Speakers judge SemSs mostly proceeding from actual sentences. If a sentence is anomalous—odd or outright incorrect, we want to see at what level of representation this anomaly can be detected. According to Apresjan, the distinction must be drawn between extralinguistic “distortions” (he considered contradictions and tautologies) and strictly linguistic anomalies. We concentrate on a further distinction, namely

the distinction between two types of linguistic anomaly: semantic *vs.* lexical-syntactic. This distinction rests on the following simple criterion (slightly reformulated from Apresjan, 1978 [1995: 609]):

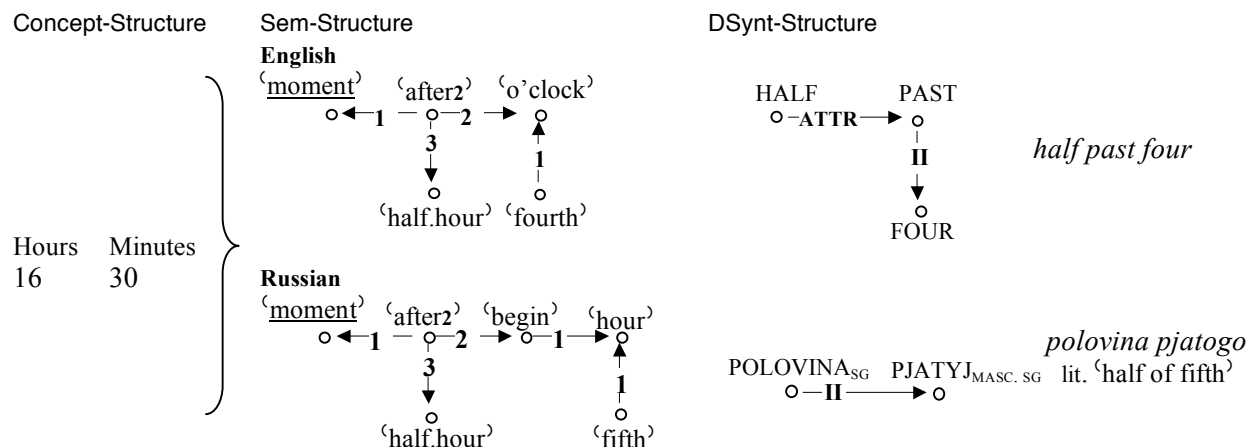
If the meaning ' σ ' of a linguistically anomalous sentence can be expressed by a linguistically correct paraphrase without modifying ' σ ', then the anomaly is not semantic; otherwise, that is, if, in order to obtain a correct paraphrase, ' σ ' has to be replaced with another, although (quasi-)equivalent meaning ' σ' ', the anomaly is semantic.

Two different meanings, and, consequently, two different SemSs are (quasi-)equivalent if they have the same denotation.

Let us emphasize, lest a misunderstanding arises, that by “a linguistically correct expression” we mean an expression that is normally acceptable in everyday non-technical speech.

To make our basic idea clear, we will begin with an example. Namely, we will present the main structures at the three levels of representation for two semantically equivalent expressions of English and Russian: Concep(tual)S(tructure), SemS and DSyntS. The expressions chosen for illustration are equivalent phrases *half past four* and *polovina pjatogo*, both naming a particular moment during the day: “16: 30.”

In order to save space, we give semantemes of Russian in English (even if this contradicts the nature of semantemes, which are language-specific; we do the same with French semantemes later on); the distinguishing lexicographic numbers are taken from *Longman Dictionary of Contemporary English Online*.



The semanteme 'o'clock' means 'one of 24 moments that divide a day in equal parts', and the semanteme 'hour' is 'duration between two consecutive o'clocks'; the ordinal numeral corresponds—in this case—to the order of 'o'clock'/'hour' after 'noon', i.e., after the moment when the Sun is at the highest point (rather than to the quantity of 'o'clocks' and 'hours'); finally, the underscoring in SemSs specifies the communicatively dominant node.

Figure 1. SemSs and DSyntSs for an English and a Russian expression naming a moment.

As seen from Fig. 1, the same informational content, in our case, the identification of a particular moment, can be expressed in two different languages by two non-isomorphic SemSs—that is, by two different meanings (which are of course equivalent). The moment in question is named in English by giving the time interval that has passed from the 4-th o'clock till this moment ('half.hour after the 4-th o'clock'), while in Russian this is done by indicating a moment in the middle of the full hour following the 4-th o'clock ('half of the fifth hour'). Thus, English and Russian use different meanings in order to name the same moment: in English, the point of reference is 'the moment "the fourth o'clock"', while in Russian, it is 'the moment of the beginning of the fifth hour'. (Cf. B. Russell's famous example of the two different meanings corresponding to the same referent: 'Sir Walter Scott' and 'the author of *Waverley*'.)

Different SemSs corresponding to the same extralinguistic content are quite a common phenomenon in natural language; such correspondences are found between different languages as well as inside the same language. What makes such SemSs interesting is the fact that some of them can be good in one language/ in one context, but bad in a different language/in a different context. Thus, the English SemS of *half past*

four is not allowed in Russian (**polovina posle četyrëx*), and vice versa: **half of [the] fifth* is impossible in English. One can speak of well- or ill-formed SemSs of a **given language**, aiming at their **linguistic** well-/ill-formedness—namely, at Sem-configurations that are correct/incorrect in a given language. As a result, the following problem arises:

What are the constraints that ensure, in a given language, the specification of linguistically well-formed SemSs and only such SemSs?

Section 2 proposes several such constraints, in particular—for the case presented in Fig. 1.

2 Constraints on Configurations of Semantemes

We decided to explore language-specific constraints on well-formed SemSs for the three following reasons:

1) Formally, it is necessary to specify as accurately as possible the set of objects with which our model is supposed to deal as its input.

2) Practically, it seems more economical, under synthesis, to filter out bad sentences as soon as possible—at the semantic source. This is processing economy.

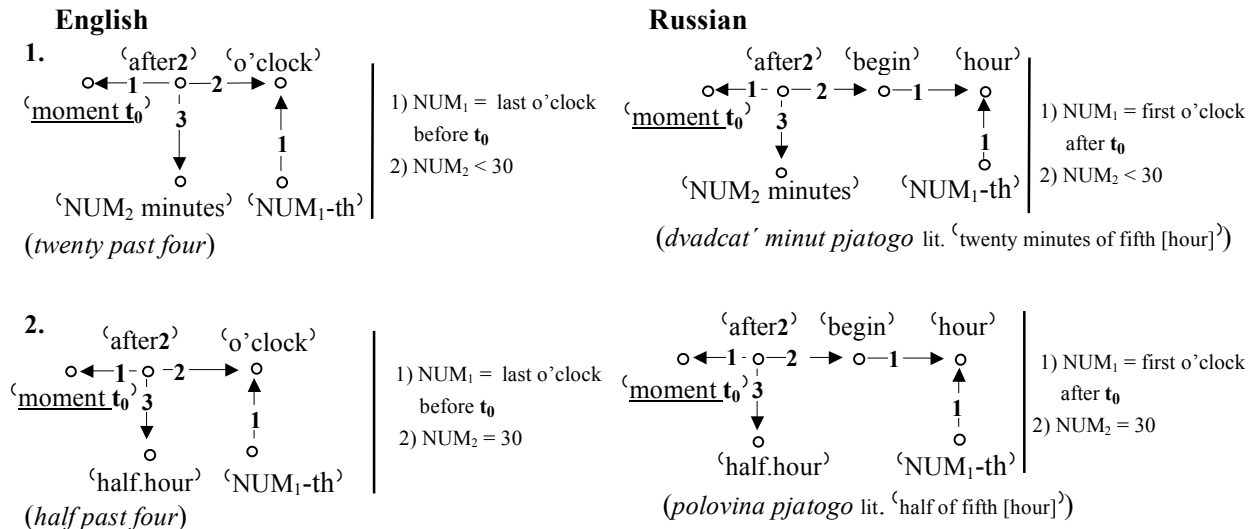
3) Linguistically, it allows for significant economy in the lexicon. Restrictions on cooccurrence attached to a semanteme are valid, as a rule, for all LUs in whose meaning this semanteme is the dominant component; and the number of such LUs can reach tens of thousands. Such is the case, for instance, with the semanteme ‘cause1’ and all causative verbs (see 2.5). Otherwise, we would have to repeat all these restrictions with each individual LU.

A bad SemS in a language **L** is bad, of course, only because **L** lacks necessary lexical-syntactic means for its implementation. Therefore, it is logically possible to do without constraints on Sem-configurations—by means of constraints on Lexicalization. From the linguistic point of view, however, this is an unacceptable solution, because it misses an essential generalization: “The meaning of such a structure is **never** lexicalizable in **L**.”

Now we will examine several cases illustrating linguistic well-/ill-formedness of Sem-configurations.

2.1 Naming a Moment

The meaning corresponding to “16: 30” cannot be expressed in English as **half of [the] fifth*, which is a literal rendering of the Russian pattern. It is impossible to save this expression by superficial corrections: **half of the fifth hour* or **thirty (minutes) of the fifth (hour)* are equally unacceptable, although syntactically these expressions are impeccable and 100% understandable. What is at stake here is not a bad realization of a well-formed SemS, but an ill-formed SemS. To specify some well-formed English and Russian SemSs in the case where the moment named does not correspond to a round hour (= to an ‘o’clock’), we need three semantemic patterns:



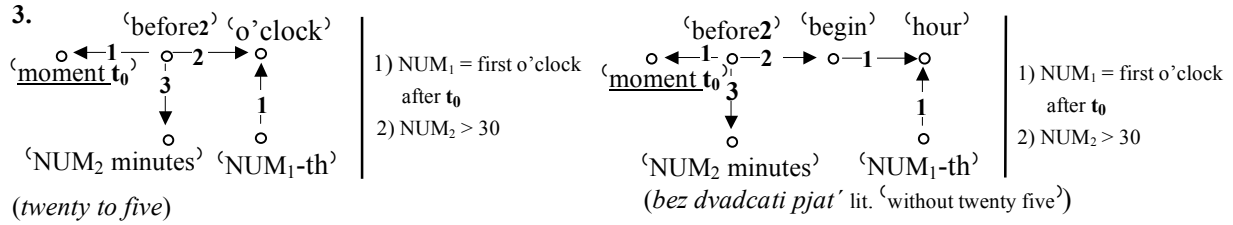


Figure 2. English and Russian Semantemic Patterns for Denoting a Time Moment.

The corresponding constraint must be attached to the semanteme 'moment' in both languages:

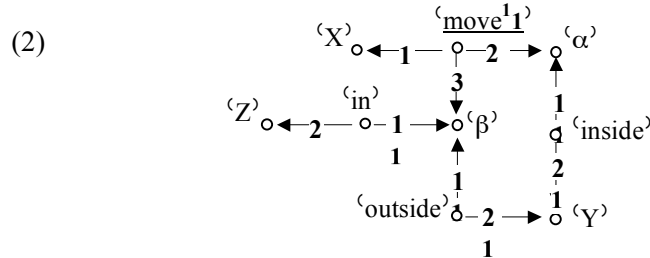
A SemS containing a Sem-configuration of the form 'moment←1→after2/before2-3→minutes' is allowed to be used in order to name a moment only if this configuration is part of one of the three above patterns for a given language.

2.2 Moving Out from Somewhere to Somewhere

Consider the following familiar example (see, e.g., Apresjan, 2000: xii-xiii), the French sentence (1):

(1) *Le serpent est sorti de son trou* 'The snake came out of its hole'.

Our object of interest here is the verb SORTIR ≈ 'X comes out of Y to Z' = 'X moves¹ from point α, which is inside Y, to point β, which is outside Y—in Z'; this definition is represented by the SemS in (2):



As is well known, the SemS (2) cannot be implemented in non-technical, everyday Russian; the sentence (3) is unacceptable:

(3) a. **Zmeja peremestilas' iz nory naružu* lit. '[The] snake moved¹ out.from [its] hole outside'.

The linguistic reason for this is also known: Russian requires that, for any act of moving¹, the way of locomotion be specified—that is, 'by walking', 'by flying', 'by crawling', 'by jumping', etc.¹ Moreover, if the movement is characterized by its beginning point or its end point, this additional obligatory Sem-component must be normally expressed inside a single lexeme—usually, a derived verb of the type VY+JTI 'come out by walking', VY+LETET' 'come out by flying', VY+POLZTI 'come out by crawling'. Thus, (3b) is also unacceptable as a description of the fact described in (1), the correct way being (3c):

- b. **Zmeja peremestilas' iz nory naružu polzja* lit. 'The snake moved¹ out of its hole outside by crawling'.
c. *Zmeja vypolzla iz nory* 'The snake crawled out of its hole'.

(The expression of the correct Sem-configuration 'move¹ by crawling' by a single lexeme in Russian is ensured under the lexicalization of the SemS.)

Therefore, Russian needs the following semantic constraint:

¹ The verb PEREMESTIT'SJA 'change place' as in *Sekretariat peremestilsja v sosednee zdanie* 'Secretaries moved to the next building' is a different lexeme. PEREMESTIT'SJA in (3) can be normally used for generic statements in scientific-type texts: for instance, *Ulitka peremeščaetsja so skorost'ju 0,08 m/min* 'A snail moves with the speed of 0.08 m/min'.

A SemS containing the semanteme ‘move¹1’ is allowed in non-technical everyday speech only if ‘move¹1’ is part of the Sem-pattern ‘move¹1←1–manner–2→Ψ’.

In contrast, French avoids expressing the manner of locomotion together with the indication of movement’s orientation in general—if there is no special communicative need to insist on it. One can say

(4) *Le serpent est sorti de son trou en rampant* lit. ‘The snake came out of its hole by crawling’,

but this puts a communicative emphasis on ‘crawling’. The corresponding constraint for French is:

A SemS containing a Sem-configuration of the form ‘move¹1–2→α’ or ‘move¹1–3→β’ is allowed in non-technical everyday speech if 1) ‘move¹1’ is not part of the pattern ‘move¹1←1–manner–2→Ψ’ or 2) if the manner of locomotion is informationally especially valuable for the Speaker.

Both constraints are to be associated to the corresponding semantemes: Fr. ‘se déplacer = move¹1’ and Rus. ‘peremeščat’sja = move¹1’. In case we are translating from French into Russian, we have to select one of more specific synonyms of ‘peremeščat’sja = move¹1’, whose meanings include the semanteme of manner of locomotion: for instance, ‘vypolzti = move¹1 out by crawling’, ‘vybežat’ = move¹1 out by running’, etc. As a result, the French sentence (1) is translated into Russian as (3c).²

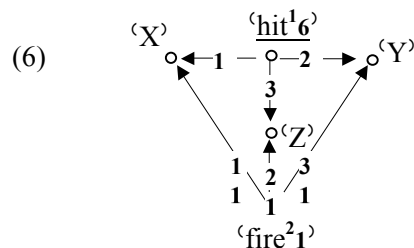
The phenomenon at stake here is what is called **conflation** (Talmy, 2003: 25 and *passim*): different Sem-configurations are covered by different LUs (within one language or across languages).

2.3 Shooting Somebody

The English sentence (5) cannot be easily translated into Russian:

(5) *Mary shot John.*

This happens because Russian has no verb exactly corresponding to SHOOT¹1 ‘X shoots Y with Z’, which has the lexicographic definition ‘X hits¹6 Y with projectile Z which X fires²1 at Y’; the decomposition of the semanteme ‘shoot¹1’ can be represented by the SemS in (6):



HIT¹6 ≈ ‘reach object Y with projectile Z’; to simplify the presentation, we omitted a SemA of HIT¹6 (and, correspondingly, of SHOOT¹1): hit where? (*John was shot/hit in the leg*).
FIRE²1 ≈ ‘send projectile Z at Y’.

The closest Russian equivalent of this SHOOT, the verb STRELJAT’, means ‘X fires at Y’ (= shoots¹2 at Y); it does not imply hitting the target, which constitutes the semantically dominant component of ‘shoot¹1’, cf. (6). SemS (6) can be literally expressed in Russian as (7):

(7) *X popadaet v Y-a Z-om, vystreliv v Y-a Z-om* ‘X hits¹6 Y with Z having fired²1 Z at Y’.

Nevertheless, none of sentences in (8) is acceptable as a translation of (5):

- (8) a. [#]*Mèri popala v Džona, vystreliv (streljaja) v nego* ‘Mary hit¹6 John firing²1 (having fired²1) at him’.
b. [#]*Mèri vystrelila (streljala) v Džona i popala (v nego)* ‘Mary fired²1 at John and hit¹6 him’.

All these sentences are syntactically and morphologically correct; no lexical cooccurrence is violated, either. The only conclusion is, then, that the SemS of (5) is ill-formed in Russian.

² A similar difference between motion and position verbs in French and German is well illustrated in (Malblanc, 1961: 66ff).

However, it is impossible to ban the SemS of the form (6) from Russian altogether, because in particular contexts it can be implemented:

- (9) a. *Mèri popala v zajca, streljaja s kolena* ‘Mary hit¹₆ the hare from the kneeling position’.
 b. *Mèri vystrelila v zajca i, kak ni stranno, popala*
 ‘Mary fired²₁ on the hare and, strange though it may seem, hit¹₆ him’.

SemS (6) can be allowed in Russian only if the speaker is especially interested in the “intermediate” result of firing—namely, hitting the target. In a neutral situation of firing at Y, the goal is to harm Y; reaching this goal implies hitting, but not the other way around (for instance, Y can be protected by good body armor). Therefore, sentences in (8) are informationally deficient with respect to the meaning of (5): they tell us more about Mary’s actions and “successes,” while (5) is about what has happened to John. This can be captured by the following constraint, associated with the Russian semanteme ‘streljat’ ≈ fire²₁:

A SemS containing a Sem-configuration of the form (6) is allowed only if the meaning ‘hit¹₆’ is informationally especially valuable for the Speaker.

English avoids this problem because in (5) the semanteme ‘hit¹₆’ is not expressed by a separate LU.

Since the SemS in (6) cannot be used in Russian to describe the situation present in (5), semantic paraphrasing is triggered—namely, looking for quasi-synonyms of the dominant node of (6), the semanteme ‘hit¹₆’. Its Russian equivalent, ‘popadat’¹, has more specific quasi-synonyms ‘ranit’ ≈ injure¹ and ‘ubit’ ≈ kill¹, one of which must replace Sem-configuration (6)—if Y is a living being. The result is (10a) or (10b):

- (10) a. *Mèri ranila Džona* ‘Mary injured John’.
 b. *Mèri ubila Džona* ‘Mary killed John’.

These are, of course, very approximate translations of (5): they lose information about firing, but, at the same time, they add information about the final result of the firing. This means that the translator needs knowledge about the real extralinguistic situation (he has to know what actually happened to John).

2.4 Be Localized Somewhere

Sentence (11a) cannot be rendered in Russian literally—as (11b); the correct form is (11c):

- (11) a. *Mary stiffened up beside John* ⟨in her chair⟩.
 b. **Mèri vsja naprjaglas’ rjedom s Džonom* ⟨v svoem kresle⟩.
 c. *Mèri, kotoraja byla rjedom s Džonom* ⟨sidevšaja v kresle⟩, *vsja naprjaglas’*
 ‘Mary, who was beside John ⟨sitting in her chair⟩, stiffened up’.

The problem with (11b) is caused by the fact that the Russian semanteme ‘be.localized’ (the dominant node in the decomposition of any locative preposition)—in contrast to the English one—cannot accept as its SemA 1 the designation of a particular class of states such as some uncontrollable human physiological reactions (stiffen up, go limp, shake). The corresponding restriction must be attached to the Russian semanteme ‘be.localized’. Because of this constraint, the Sem-configuration ‘stiffen.up←1–beside [Y]’ is ill-formed in Russian; the rules of Sem-paraphrasing will modify it to produce a different, but equivalent SemS, in which ‘beside [Y]’ will bear on ‘Mary’, so that the SemS for (11c) is as follows:

Rus. ‘John←2–beside–1→Mary←1–stiffened.up’.

2.5 Causing Something

The English semanteme ‘cause1’ (= ‘be the cause of’) takes as its SemA 1 the designation of an object impacting something; this is true of any transitive verb whose meaning’s dominant node is ‘cause1’:

- (12) a. *The hammer falling from the balcony broke the window.*
 b. *John's bullet killed the wolf instantly.*

But sentences in (13) are bad:³

- (13) a. *John struck the window with a hammer, and [#]the hammer broke it.*
 b. *[#]John's rifle killed the wolf instantly [for the situation in which John fired at the wolf].*

‘Cause1’ cannot take the name of an object as its SemA 1, if, in the situation described, this object is an instrument being used by somebody (Kahane & Mel’čuk, 2006); this restriction must accompany the semanteme ‘cause1’ in English (and also in French and Russian).

In the same vein, (14a) cannot be literally rendered in Russian as (14b); the correct way is (14c):

- (14) a. *An ambush killed four soldiers.*
 b. Rus. **Zasada ubila četyrěx soldat.*
 c. Rus. *V zasade pogiblo četyre soldata* ‘Four soldiers died in an ambush’.

The Russian semanteme ‘cause1’ has an additional restriction on its SemA 1: at least, in the situation of killing, the cause X (of the death) must be in physical contact with Y that undergoes the causation; ‘cause1’ is the dominant node of the corresponding semanteme ‘kill2’. Therefore, the SemS

‘ambush←1-kill2’

is ill-formed in Russian, and must be replaced by an equivalent one:

‘ambush←2-use-1→α←1-kill1-2→soldiers’ (≈ ‘α kills1 soldiers by using an ambush’),

which eventually gives (14c).

3 Semantic vs. Lexical-Syntactic Linguistic Ill-formedness

The borderline between ill-formed semantemic combinations (in a SemS) and ill-formed lexical combinations (in a DSyntS) is not always obvious. Let us illustrate the respective cases.

A sentence synthesized from a well-formed SemS may be ill-formed at the DSynt-level for one of two reasons: either an incorrect choice of LU for a correctly used semanteme, but without breaking the rules of mutual cooccurrence of LUs; or an incorrect choice of LU that leads to impossible or undesirable lexical cooccurrence.

- **Bad lexical choice:** Rus. *IZ-ZA* ≈ ‘because of’

Russian sentences (15a) and (15b) clearly contrast, (15b) being pragmatically bizarre:

- (15) a. *Ivan pogib iz-za svoej rassejannosti* ≈ ‘Ivan died because of his absentmindedness’.
 b. *[#]Ivan spassja iz-za svoej rassejannosti* ‘Ivan was saved because of his absentmindedness’.

The explanation is clear: (15b) is bad because *spassja* ‘was saved’ is, generally speaking, a desirable event, while the preposition *IZ-ZA* (≈ ‘because of’) implies the absence of desirability of X for X’s corresponding SemA or for the Speaker (Iordanskaja & Mel’čuk, 1996: 169–172):

‘X iz-za Y-a’ = ‘X because of Y, it being not true that X is desirable for SemA₁(‘X’) or for the Speaker’.

³ They are bad in a non-contrastive context only; cf. *John failed to break the window until he struck it with a hammer; the hammer broke it instantly* or *My pistol had no effect on the wolf, but John's rifle killed the animal*.

The meaning of (15b) is then as follows: ‘Ivan was saved because of his absentmindedness, it being not true that being saved is desirable for Ivan or for the Speaker’. In spite of its pragmatic strangeness, this meaning is well-formed: the maxim “to be saved is good” is not a semantic fact of a particular language, but rather a piece of encyclopedic information. One can even imagine a situation in which being saved is not considered as something desirable (for instance, Ivan wanted to die, but failed). The problem with sentence (15b) lies in a lexical choice: if the Speaker does not want to negate the desirability of Ivan’s being saved, he has to express the semanteme ‘because of’ by the preposition *BLAGODARJA* ‘thanks to’:

‘X blagodarja Y-u’ = ‘X because of Y, X being desirable for $SemA_1(X)$ or for the Speaker’.

The problem of the correct selection of LUs in accordance with the starting SemS and the knowledge and intentions of the Speaker is orthogonal to the problem of correct combination of semantemes in accordance with their properties. The anomaly in question cannot be blocked by a syntagmatic filter of the DSynt-level: no lexical cooccurrence restrictions are violated in (15b). The result of a bad lexical choice of this type is always a relatively incorrect sentence (incorrect with respect to a given meaning). Its pragmatic unacceptability is due to the fact that the wrongly selected LU may add some unwarranted information, which renders the sentence odd.

• **Bad lexical combination: Rus. NEMNOGO ≈ ‘a little’ and Rus. RUSYJ ‘light-brown [hair]’**

We will present two examples of bad lexical combinations banned by syntagmatic constraints (more precisely, semantic conditions) on lexical cooccurrence.

— As shown in (Iordanskaja & Mel’čuk, 2004), the Russian adverb *NEMNOGO* ≈ ‘a little’ does not combine with a “threshold” lexeme *L*—a lexeme whose meaning includes the component ‘sufficiently bigger/smaller than the norm’. The semanteme ‘sufficiently’ determines the threshold after which *L* can apply. Thus, one can be *nemnogo grusten* ‘a little sad’, but not **nemnogo vesel* ‘a little cheerful’: one is called *grusten* as soon as he is even slightly below the norm of neutral emotional state, but to be called *vesel*, one has to be **sufficiently** over the norm—one has to cross a particular threshold. Yet the combination of the semanteme ‘nemnogo = a little’ with a threshold semanteme is quite possible, although in such a case, the lexeme *NEMNOGO* itself cannot be used:

for RANEN ‘injured’	: ‘nemnogo ranen’	⇔	<i>legko ranen</i>	‘lightly injured’	⟨ <i>*nemnogo ranen</i> ⟩
for SOGLASEN ‘agree’	: ‘nemnogo soglasen’	⇔	<i>otčasti soglasen</i>	‘agree partially’	⟨ <i>*nemnogo soglasen</i> ⟩
for ŽELTYJ ‘yellow’	: ‘nemnogo želtyj’	⇔	<i>želt+ovat-yj</i>	‘yellowish’	⟨ <i>*nemnogo želtyj</i> ⟩

The observed ungrammaticalities are due to the violation of lexical restricted cooccurrence, not ill-formed SemSs; thus, ‘nemnogo vesel’ can be correctly expressed as *vesel, no ne očen’* ‘cheerful, but not much’ or *ne sliškom vesel* ‘not too cheerful’. Of course, the phrase **nemnogo vesel* also illustrates bad lexical selection; however, this anomaly, in contrast to the preceding case, is blocked syntagmatically. The semanteme ‘a little’ corresponds to the lexical function [= LF] *AntiMagn*, and the choice of the proper element of an LF is done at the surface-syntactic level.

— Many languages use special adjectives to denote the color of human hair and human eyes. Thus, Russian has *rusye volosy* ‘light-brown hair’ and *karie glaza* ‘brown eyes’, and these adjectives cannot be applied to anything else; their definitions are as follows: ‘rusye X’ = ‘light-brown hair X’ and ‘karie X’ = ‘brown eyes X’. As a consequence, the expressions **karie volosy* ‘brown hair’ and **rusye glaza* ‘light-brown eyes’ are ungrammatical, but this ungrammaticality corresponds to violation of lexical cooccurrence. SemSs ‘light-brown hair’ and ‘brown eyes’ are well-formed; however, for hair and eyes the brown color cannot be expressed by the adjective (SVETLO-)KORIČNEVYJ ‘(light-)brown’: **svetlo-koričnevye volosy* ‘light-brown hair’ and **koričnevye glaza* ‘brown eyes’ are bad phrases in Russian because of the restricted combinability of particular LUs.

Bad lexical combinations produce absolutely incorrect sentences.

Thus, semantic constraints on Sem-configurations must be consistently distinguished from semantic conditions on lexical combinations (Iordanskaja & Mel’čuk, 2004: 122ff). Two LUs units can be non-combinable only because of their meanings; but this does not entail that their meanings are non-com-

binable. Very often two semantemes are perfectly combinable, but one of them must be lexically expressed as a function of the lexical implementation of the other; any violation results in lexical-syntactic ill-formedness. This is the case of lexical functions.

4 Typology of Semantic Well-formedness Constraints

The data analyzed allows us to sketch a preliminary typology of linguistic constraints that come into play at the semantic level and specify linguistically well-formed SemSs.

A linguistic constraint attached to a particular semanteme ' σ ' of **L** can be of two types—as a function of the constraint's target: this is either the Sem-Governor of ' σ '/of one of SemAs of ' σ ' or the Sem-Dependent of ' σ '; in other words, either the semanteme ' σ_G ' of which ' σ ' (or a SemA of ' σ ') is a SemA, or the semanteme ' σ_D ' which is a SemA of ' σ '. These two types can be dubbed “up-oriented” and “down-oriented.”

1. Up-oriented semantemic constraints. Two subtypes are distinguished: semantic clichés and lexical confluences.
 - 1.1. Sem-clichés: Sem-configurations to be expressed by lexical-syntactic schemata (see 2.1: “naming a moment”).
 - 1.2. Lexical confluences: Sem-configurations to be expressed by single LUs or to be avoided, since there is no good lexical expression, that is, “positive” or “negative” confluences—some semantemes in a language are inseparable, while in another language they can be incompatible (see 2.2 and 2.3: “moving out from somewhere” and “shooting somebody”).
2. Down-oriented semantemic constraints. The data examined so far show also two subtypes: the constraint bears either only on one SemA of ' σ ' (it seems that this is SemA 1) or it concerns also other semantemes.
 - 2.1. Sem-A 1 of ' σ ' must belong/not belong to a specific class of entities or facts (see 2.4: “be localized somewhere”).
 - 2.2. Sem-A 1 of ' σ ' must satisfy a complex set of conditions that concern other semantemes as well (see 2.5: “causing something”).

This is no more than a beginning of a fuller typology; at the time being, this is the first step to be taken. Note that proposed constraints on linguistically well-formed SemS are formally what are known as filter rules.

5 Concluding Remarks

Three questions will be touched upon in order to round up our exposition.

5.1 Minimality of SemS

The notion of well-formedness of SemSs is intimately related to the definition of SemS. It is impossible to discuss this definition here, but we will nevertheless indicate two properties of SemS essential from our perspective.

- The vocabulary of SemSs must be minimal (Apresjan, 1994 [1995: 468]): not all signifieds of full LUs of **L** are admitted as semantemes in a SemS. Thus, if the signifieds of two LUs L_1 and L_2 share a Sem-configuration ' σ ', while the distinguishing parts (if any) contain only restrictions on the Sem-actants of ' σ ', then this ' σ ' is taken as a semanteme (rather than ' L_1 ' or ' L_2 '). For instance, the adjective HAZEL [X] 'green-brown eyes [X]' differs from GREEN-BROWN [X] only by a restriction on the SemA X; therefore, only the shared configuration 'green-brown' must be taken as a legitimate semanteme. The signifieds of the Russian causal prepositions IZ-ZA 'because of' and BLAGODARJA 'thank to' share the component '[Y] caused₁ by X', and their differences are only restrictions imposed on their actants: (un)desirability of Y for the corresponding SemA of Y or for the Speaker; as a result, none of these signifieds is allowed to the SemS, the corresponding semanteme being 'caused₁ by X'. In a similar vein, the Russian prepositions IZ, OT and PO share the Sem-component '[Y] directly caused₁ by X' and are differentiated only by various

restrictions on their SemAs (Iordanskaja & Mel'čuk, 1996); all three are represented in the SemS by the Sem-configuration 'directly caused1'.

- The composition of a specific SemS must also be minimal: not all elements of the extralinguistic situation represented by a SemS must (or even can) be reflected in it. Under lexicalization, the LUs selected can add semantic information and thus render the starting SemS richer. For instance, to express the semanteme '[Y] caused1 by X' in Russian, one may have to choose between IZ-ZA and BLAGODARJA, and for this, one must know whether Y is desirable. It is impossible to require that any Russian starting SemS contain such an indication for everything that is caused1. Rather, the choice between IZ-ZA and BLAGODARJA triggers the search for this piece of information and in this way develops the “underspecified” starting SemS.

5.2 Extralinguistic and Linguistic Well-formedness of a SemS

The substantive well-formedness of a SemS can be characterized with respect either to extralinguistic reality or to the language under consideration.

- **Extralinguistic** well-formedness of a SemS corresponds to its **interpretability** in terms of extralinguistic reality (in the largest sense possible). This interpretability hinges on encyclopedic and pragmatic knowledge. For instance, the famous sentence *Colorless green ideas sleep furiously* has an encyclopedically ill-formed SemS: because of logical contradiction between semantemes, currently leading to absurdity (cf. 'colorless green [ideas]'). From the linguistic viewpoint, however, its SemS is well-formed.

- **Linguistic** well-formedness of a SemS is its conformity to **language-specific constraints** on Sem-configurations. In other words, logical contradiction, tautology or any other type of absurdity do not automatically entail the linguistic ill-formedness of a given SemS.

A native speaker reacts to an utterance whose SemS is extralinguistically ill-formed by “What do you mean? This is nonsense.” An utterance with a linguistically ill-formed SemS triggers a different type of reaction: “You don't say this like this; that's what you have to say.”

Extralinguistic and linguistic well-formedness are logically independent properties of SemSs, so that four types of SemSs are possible.

1. Extralinguistically and linguistically well-formed SemSs underlie meaningful correct expressions: *Mary shot John*.

2. Extralinguistically well-formed, but linguistically ill-formed SemSs give rise to meaningful, but incorrect expressions: Rus. **Zmeja peremestilas' iz nory* 'The snake moved out of the hole' (⇒ *Zmeja vypolzla iz nory*).

3. Extralinguistically ill-formed, but linguistically well-formed SemSs produce meaningless expressions that can be linguistically correct or incorrect—at the level closer to the surface—as a function of the explicitness in expressing the contradiction. (By **explicitness** we mean using a separate full lexeme for a given meaning.)

└ If the expression of a logical contradiction is not explicit, this contradiction gives rise to a linguistic anomaly (based on Apresjan, 1989 [1995: 624]).

Thus, in Russian:

- The synonymous phrases *kroxotnyj ogromnyj dom* 'tiny huge house' and **kroxotnyj domišče* 'tiny huge.house' have both extralinguistically ill-formed, but linguistically well-formed SemS; however, the first phrase is correct at the DSynt-level, while the second one is not, since the meaning 'huge' is expressed in it by a derivational suffix **-išče**, i.e., not explicitly.

- The synonymous phrases *rovno desjat' knig, xotja ja i ne znaju, skol'ko v točnosti* and **rovno knig desjat'* lit. 'exactly books ten', both meaning 'exactly ten books, although I don't know precisely how many' are both extralinguistically ill-formed, but linguistically well-formed at the semantic level; however, the first phrase is correct at the DSynt-level, while the second one is not, since it expresses the meaning 'although I don't know precisely how many' by a syntactic construction with inversion of the numeral and the quantified noun, i.e., not explicitly.

The sentence *Colorless green ideas sleep furiously* belongs here. In this connection, the following observation seems important. An extralinguistically ill-formed SemS normally features violations of some

constraints that the semantemes in this SemS impose on their SemAs: thus, 'green' requires for its SemA 1 to denote a visible physical entity. However, when such a constraint, of a very general taxonomic character, is transgressed, the effect is an obvious absurdity expressible by a linguistically correct sentence. In sharp contrast, semantic constraints that control linguistic well-formedness of SemSs are much more specific and therefore much less obvious. Violating them produces the effect of a linguistic mistake.

4. Extralinguistically and linguistically ill-formed SemSs lead to meaningless incorrect expressions: Rus. **Zelěnaja ideja peremestilas' iz nory* 'The green idea moved out of the hole' (no less absurd, but linguistically correct expression would be, for instance, *Zelěnaja ideja vyletela/vypolzla/vyprygnula iz nory* 'The green idea flew out/crawled out/jumped out of the hole').

In this paper we have examined exclusively extralinguistically well-formed, but linguistically defective SemSs (Type 2 above).

5.3 Constraints on Sem-Configurations vs. Semantic Components in Lexicographic Definitions

Language-specific constraints on Sem-configurations cannot in principle be always reduced to particular semantic components in the decomposition of the semantemes under description. Such a constraint is of a different nature: it is itself not a *bona fide* Sem-configuration, but a requirement that a Sem-configuration must satisfy in order to be an actant or a Sem-governor of a given semanteme. Moreover, these constraints may be organized as complex Boolean formulas that cannot be easily included into a semantic decomposition; see, for instance, the restrictions on the use of the French semanteme 'causer1' in (Kahane & Mel'čuk, 2006: 260). It follows that semantemes, although they are not signs, may have their own syntactics, which controls their combinability.

Acknowledgments

The draft of this paper has been read and commented upon by Ju. Apresjan, D. Beck, I. Boguslavskij, S. Kahane, J. Milićević, and A. Polguère; we thank all these colleagues and friends for their remarks and suggestions that have allowed us to significantly improve the presentation.

References

- Apresjan Jurij. 1978. Jazykovaja anomalija i logičeskoe protivorečie. [1995: 598-621]
- Apresjan Jurij. 1989. Tavtologičeskie i kontradiktornye anomalii. [1995: 622-628]
- Apresjan Jurij. 1994. O jazyke tolkovanij i semantičeskix primitivax. [1995: 466-484]
- Apresjan Jurij. 1995. *Izbrannye trudy. Tom II. Integral'noe opisanie jazyka i sistemnaja leksikografija*. Škola «Jazyki russkoj kul'tury», Moskva.
- Apresjan Jurij. 2000. *Systematic Lexicography*. Oxford University Press, Oxford.
- Iordanskaja Lidija & Igor Mel'čuk. 1996. K semantike russkix pričinnyx predlogov (*IZ-ZA ljubvi ~ OT ljubvi ~ IZ ljubvi ~ *S ljubvi ~ PO ljubvi*). *Moskovskij Lingvističeskij Žurnal*, 2: 162-211.
- Iordanskaja Lidija & Igor Mel'čuk. 2004. The Meaning and Cooccurrence of Russian NEMNOGO 'a little.' In Ju. Apresjan (ed.), *Sokrovennye smysly: slovo, tekst, kul'tura. Sbornik statej v čest' N. D. Arutjunovoj*, 112-127. Jazyki slavjanskoj kul'tury, Moskva.
- Kahane Sylvain & Igor Mel'čuk. 2006. Les sémantèmes de causation en français. *LINX* [Revue de linguistique de l'Université Paris X Nanterre], n° 54, 247-292.
- Malblanc, Alfred. 1961. *La stylistique comparée du français et de l'allemand*. Didier, Paris.
- Talmy, Leonard. 2003. *Toward a Cognitive Semantics. Volume I. Typology and Process in Concept Structuring*. MIT Press, Cambridge, MA—London.