

Delimitation of information between grammatical rules and lexicon

Jarmila Panevová, Magda Ševčíková

Charles University in Prague

Faculty of Mathematics and Physics

Institute of Formal and Applied Linguistics

{panevova|sevcikova}@ufal.mff.cuni.cz

Abstract

The present paper contributes to the long-term linguistic discussion on the boundaries between grammar and lexicon by analyzing four related issues from Czech. The analysis is based on the theoretical framework of Functional Generative Description (FGD), which has been elaborated in Prague since 1960's. First, the approach of FGD to the valency of verbs is summarized. The second topic, concerning dependent content clauses, is closely related to the valency issue. We propose to encode the information on the conjunction of the dependent content clause as a grammatical feature of the verb governing the respective clause. Thirdly, passive, resultative and some other constructions are suggested to be understood as grammatical diatheses of Czech verbs and thus to be a part of the grammatical module of FGD. The fourth topic concerns the study of Czech nouns denoting pair body parts, clothes and accessories related to these body parts and similar nouns. Plural forms of these nouns prototypically refer to a pair or typical group of entities, not just to many of them. Since under specific contextual conditions the pair/group meaning can be expressed by most Czech concrete nouns, it is to be described as a grammaticalized feature.

1 Introduction

Theoretical approaches to natural languages, regardless of which particular theory they subscribe to, usually work with grammar and lexicon as two basic modules. The delimitation between these modules

is not given by the language itself, the “balance” between the modules is “entirely an empirical issue” (Chomsky, 1970). There are core grammatical and lexical topics, such as agreement or lexical meaning, respectively, whose classification as belonging to the grammar on the one hand and to the lexicon on the other is shared across languages and linguistic theories, while classification of borderline cases as either grammatical or lexical ones is strongly theory-dependent.

A brief overview of selected approaches laying more stress either on the lexical or on the grammatical module is given in Sect. 2 of the present paper; the approach of Functional Generative Description, used as the theoretical framework of our analysis, is briefly presented. In Sect. 3 to 6, the delimitation of information between the two modules is exemplified by four topics, which have been studied for Czech.

2 Grammar vs. lexicon in selected theoretical approaches

The interplay between grammar and lexicon has been discussed for decades in linguistics. Although the former or the latter module plays a predominant role in particular frameworks, the preference of one of the modules does not mean to exclude the other one from the description, they are both acknowledged as indispensable.

According to Bloomfield (1933), the lexicon has a subordinated position.¹ The grammatical rules are the main component either within Chomskyan generative (transformational) grammar, though the im-

¹“The lexicon is really an appendix of the grammar, a list of basic irregularities.” (Bloomfield, 1933)

portance of the lexical component was strengthened by the decision to treat certain types of nominalizations within the lexicon rather than within the transformational (grammatical) component (Chomsky, 1970).

On the other side of the scale of grammatically vs. lexically oriented approaches,² there is the lexicalist approach of Meaning-Text Theory by Mel'čuk et al. Within this framework, a richly structured lexicon, so-called Explanatory Combinatorial Dictionary, has been systematically compiled for individual languages; the Dictionary is considered as a central component of description of language, cf. (Mel'čuk, 1988; Mel'čuk, 2006). Lexicon plays a crucial role in categorial grammars (Ajdukiewicz, 1935) as well as in the lexicalized tree adjoining grammar, see Abeillé – Rambow (2000), just to give two further (chronologically distant) examples.

Functional Generative Description (FGD) works with both grammatical and lexical modules since the original proposal of this framework in 1960's (Sgall, 1967); nevertheless, the main focus has been laid on grammatical, in particular syntactic, issues (Sgall et al., 1986). FGD has been proposed as a dependency-based description of natural language (esp. Czech). The meaning-expression relation is articulated here into several steps: the representations of the sentence at two neighboring levels are understood as the relation between form and function. The "highest" level (tectogrammatics) is a disambiguated representation of the sentence meaning, having the counterparts at lower levels. On the FGD framework the multi-layered annotation scenario of Prague Dependency Treebank 2.0 (PDT 2.0) has been built.

PDT 2.0 is a collection of Czech newspaper texts from 1990's to which annotation at the morphological layer and at two syntactic layers was added, namely at the layer of surface syntax (so-called analytical layer) and of deep syntax (layer of linguistic meaning, tectogrammatical layer) (Hajič et al., 2006).³ At the morphological layer each token is assigned a lemma (e.g. nominative singular for nouns) and a positional tag, in which the part of speech and formal-morphological categories are specified (e.g.

case, number etc. for nouns). At the analytical layer, the surface-syntactic structure of the sentence is represented as a dependency tree, each of the tree nodes corresponds exactly to one morphological token and is labeled as a subject or object etc. At the tectogrammatical layer, the linguistic meaning of the sentence is captured as a dependency tree, whose nodes correspond to auto-semantic words only.⁴ The nodes are labeled with a tectogrammatical lemma (which is often different from the morphological one), functor (semantic role, label; e.g. Actor ACT) and a set of grammatemes, which are node attributes capturing the meanings of morphological categories which are indispensable for the meaning of the sentence.⁵ The tectogrammatical representation is further enriched with valency annotation, topic-focus articulation and coreference.

The lexical issues have been becoming more central in the FGD framework for the recent ten years as two valency lexicons of Czech verbs based on the valency theory of FGD (Panevová, 1974/75) have been built; cf. the VALLEX lexicon (Lopatková et al., 2008) and the PDT-VALLEX (Hajič et al., 2003), which is directly interconnected with the tectogrammatical annotation of PDT 2.0.

The approach of FGD to valency is summarized in Sect. 3 of the present paper. The delimitation of information between grammar and lexicon in FGD is further illustrated by the description of dependent content clauses in Czech (Sect. 4), grammatical diatheses of Czech verbs (Sect. 5) and representation of a particular meaning of plural forms of Czech nouns (Sect. 6).

3 Valency

The problem of valency is one of most evident phenomenon illustrating the interplay of lexical and grammatical information in the language description. Lexical units are the bearers of the valency information in any known theoretical framework. The form of this information is of course theory-dependent in many aspects, first of all in (i) to (iii):

²For this opposition, the terms 'transformationalist' vs. 'lexicalist' approaches are used, the former ones are called also 'syntactic' or simply 'non-lexicalist' approaches, depending on the theoretical background.

³See also <http://ufal.mff.cuni.cz/pdt2.0>

⁴There are certain, rather technical exceptions, e.g. coordinating conjunctions used for representation of coordination constructions are present in the tree structure.

⁵Compare the related term 'grammems' in Meaning-Text Theory (Mel'čuk, 1988).

- (i) how the criteria for distinguishing of valency and non-valency complementations are determined,
- (ii) how the taxonomy of valency members looks like,
- (iii) how the relation between the deep valency labels (arguments in some frameworks, inner participants in FGD) are reflected (see also Sect. 4).

3.1 Valency approach of FGD

In FGD we use the valency theory presented in Panevová (1974/75; 1994) and Sgall (1998) and in its application in valency dictionaries (VALLEX, PDT-VALLEX). Valency complementations enter the valency frame as an obligatory part of lexical information. The empirically determined set of inner participants (Actor ACT, Patient PAT, Addressee ADDR, Origin ORIG and Effect EFF) and those free modification which were determined as semantically obligatory with the respective verb (by the criterion of grammaticality or by the so-called dialogue test see (Panevová, 1974/75)) are included in the valency frame.

FGD avoids the concept of a wide number of valency positions known, for instance, from the Meaning-Text Model (recently see Apresjan et al. (2010)), where e.g. the verb *vyzvat* ‘to call’ has six valency slots, cf. ex. (1). In FGD the valency frame of the corresponding verb *povolat* consists of three slots: ACT(Nom) PAT(Acc) DIR3.

- (1) *vyzvat*’ *kogo-libo iz Peterburga v Moskvu po telefonu na soveshchanije*
 ‘to call somebody from Petersburg to Moscow by phone for the meeting’

3.2 Valency in the lexicon and grammar

In FGD functors are defined with respect to their language patterning and to their position in the valency frame (a verb with two valency slots includes ACT and PAT, a verb with three valency slots includes ACT, PAT and its third position is determined according to its semantics). Inner participants are divided into obligatory and optional, this information is a part of the valency frame in the lexicon.

Any grammatical module, whatever its aim is (be it analysis, or generation, or having determinative or declarative character) is based on the combinatorial nature of the verb (as a center of the sentence) with its obligatory valency slots. If the valency requirements are not fulfilled, the sentence is in some sense wrong (either as to its grammaticality or as to its semantic acceptability). The surface deletions are checked by the dialogue test or by the contextual conditions.

4 Dependent content clauses in Czech

The next topic concerns the description of dependent content clauses in Czech. Dependent content clauses are object, subject and attributive clauses that express a semantic complementation of the particular governing verb (or noun, these cases are not addressed in the paper).

4.1 Dependent content clauses in FGD

Within the valency approach of FGD, dependent content clauses are inner participants of a verb, they (more precisely, main verbs of these clauses) are classified as PAT and EFF with most verbs, less often as ACT and rarely as ADDR or ORIG in the tectogrammatical structure of the sentence according to the PDT 2.0 data.

Dependent content clauses are connected with their governing verbs by subordinating conjunctions or by pronouns and pronominals (cf. ex. (2) and (3)).⁶ Pronouns and pronominals are considered as semantically relevant parts of the tectogrammatical sentence structure and thus represented as nodes in the tectogrammatical tree whereas subordinating conjunctions are not. Conjunctions introducing content clauses are listed within the dictionary entry of the particular verb in the PDT-VALLEX, VALLEX as well as in Svozilová et al. (1997).

- (2) *Rozhodl se, že zůstane.*
 Decided.3.sg.pst.anim REFL, that stays.3.sg.fut.
 ‘He decided to stay.’
- (3) *Otec často vyprávěl, jak jezdíval.* EFF autem.
 how drove.3.sg.pst.anim car.instr.sg.neut.
 ‘Father often recounted how he drove a car.’

⁶The Czech examples are translated literally first, followed by a standard English translation.

4.2 Modality in dependent content clauses

Since conjunctions introducing dependent content clauses correspond to the modality expressed by these clauses, it seems to be theoretically more adequate to interconnect the choice of the conjunction with the modality information and to handle it within the grammatical component of the linguistic description. A pilot study based on the valency theory of FGD was carried out by Kettnerová (2008), who suggested a classification of selected Czech verbs of communication into four classes of assertive, interrogative, directive and “neutral” verbs.

Though Kettnerová’s approach concerns just verbs of communication (and so do nearly all theoretical studies dealing with dependent content clauses in Czech, e.g. Bauer (1965)), Daneš (1985), according to our preliminary analysis of dependent content clauses in the tectogrammatical annotation of PDT 2.0 clauses of this type occur with a number of other verbs. Besides verbs of communication, a dependent content clause is governed by verbs classified as verbs of mental actions according to (Lopatková et al., 2008) (e.g. *dočíst se* ‘to learn by reading’, *přehodnotit* ‘to rethink’), further by verbs of “preventing” somebody or oneself from an action (*odradit* ‘to discourage’, *předejít* ‘to prevent’) and many other which do not share a common semantic feature (*usilovat* ‘to aim’, *divit se* ‘to be surprised’).

4.3 Interconnecting lexical and grammatical information

Aiming at an extension of Kettnerová’s approach, all the verbs governing a dependent content clause were analyzed in order to find a correspondence between the conjunctions introducing the respective dependent content clause and the modality expressed by this clause. As the dependent content clause is closely related to the meaning of the governing verb, the repertory of modality types of the dependent content clauses and the conjunctions used is mostly restricted:

Most of the analyzed verbs occurred only with a dependent content clause expressing assertive modality; assertive dependent content clauses are mostly introduced by the conjunction *že* ‘that’ (less often also by *jestli*, *zda*, *zdali* or *-li* ‘whether/if’ – see the next paragraph). Substantially fewer verbs

governed only either an imperative or an interrogative dependent content clause; imperative clauses are introduced by *aby* or *ať* ‘that’, the interrogative ones by *jestli*, *zda*, *zdali* or *-li* ‘whether/if’. Only with a restricted number of verbs dependent content clauses of more modality types (and thus with several introducing conjunctions) occurred, most of them belong to verbs of communication;⁷ the conjunctions corresponded to the modality in the same way as with verbs with dependent content clauses of only a single modality type.

However, since there are semantic nuances in the modality of the dependent content clauses that cannot be described by means of the common inventory of modality types, the inventory has to be extended or revised; cf. ex. (4) and (5) that both are classified as assertive but the difference between them consist in the fact that in the former example the content of the dependent content clause is presented as given and in the latter one as open.

(4) *Ověříme, že robot*
Check.1.pl.fut, **that** robot.nom.sg.anim
místnost uklidil.
room.acc.sg.fem cleaned_up.3.sg.pst.anim.
‘We check **that** the robot had cleaned up the room.’

(5) *Ověříme, zda robot*
Check.1.pl.fut, **whether** robot.nom.sg.anim
místnost uklidil.
room.acc.sg.fem cleaned_up.3.sg.pst.anim.
‘We check **whether** the robot had cleaned up the room.’

After a solution for this question is found, at least the following issues have to be discussed before modality of the dependent content clauses is included into the grammatical module of FGD:

- the differences between conjunctions introducing dependent content clauses of a particular modality type have to be determined in a more detailed way,
- the impact of the morphological characteristics of the governing verb on the modality of the dependent content clause should be clarified.

⁷They are classified as “neutral” verbs of communication by Kettnerová (2008).

5 Grammatical diatheses of Czech verbs

The number of the diathesis proposed for the modified version of FGD as well as for the new version of Prague Dependency Treebank (PDT 3.0) was broadened in comparison with the previous version and with the usual view considering namely passivization; see Panevová – Ševčíková (2010). We exemplify below three types of grammatical diatheses with their requirements on syntax, morphology and lexicon.

5.1 Passivization

Passivization is commonly described as a construction regularly derived from its active counterpart by a transformation or another type of grammatical rule. However, though this operation is productive, there are some constraints blocking it and requirements how the result of this diathesis looks like. It is very well known that, at least for Czech, passivization cannot be applied for intransitive verbs, moreover it is not applicable for some object-oriented verbs and for reflexive verbs in Czech (ex. (6) to (8), respectively).

(6) *spát* – **je* *spáno*
sleep.inf – is.3.sg.pres slept.nom.sg.neut
'to sleep' – 'it is slept'

(7) *přemýšlet* *o něčem* – **je*
think.inf about something.loc.sg.neut – is.3.sg.pres
přemýšleno *o něčem*
thought.nom.sg.neut about something.loc.sg.neut
'to think about something' – 'it is thought about something'

(8) *rozloučit* *se* – **je* *se*
say_good.bye.inf REFL – is.3.sg.pres REFL
rozloučeno
said_good_bye.nom.sg.neut
'to say good bye' – 'it is said good bye'

There are also constraints on passivization which are lexical dependent: Some verbs having direct object in Accusative cannot be passivized, e.g. *mít* 'to have', *znát* 'to know', *umět* 'to know' at all, with some verbs this constraint concerns only their imperfective form (e.g. *jíst* 'to eat', *potkat* 'to meet'). From the other side, the passivization is not restricted only on the verbs with direct object in Accusative (e.g. *věřit komu* 'to believe + Dat', *plýtvat čím* 'to waste + Instr', *zabránit čemu* 'to avoid + Dat').

The operation of passivization is based on the shift of the certain verbal participant to the position of the surface subject. However, reminding the theory of valency in FGD sketched briefly in Sect. 3, which participant is shifted, depends on the type of the verb. Sometimes it is PAT (ex. (9)), with other verbs it is ADDR (ex. (10)), or EFF (ex. (11)). These constraints and requirements concerning passivization have to be marked in the lexicon.⁸

(9) *vykopat jámu*.PAT – *jáma*.PAT
dig.inf hole.acc.sg.fem – hole.nom.sg.fem
je *vykopána*
is.3.sg.pres dug.nom.sg.fem
'to dig a hole' – 'the hole is dug'

(10) *informovat někoho*.ADDR *o něčem*
inform.inf somebody.acc.sg.anim about something.loc.sg.neut – *někdo*.ADDR
je *o něčem*
is.3.sg.pres about something.loc.sg.neut
informován
informed.nom.sg.anim
'to inform somebody about something' – 'somebody is informed about something'

(11) *O zemětřesení* *napsal*
About earthquake.loc.sg.neut wrote.3.sg.pst.anim
reportáž.EFF – *O zemětřesení*
report.acc.sg.fem. – About earthquake.loc.sg.neut
byla *napsána*
was.3.sg.pst.fem written.nom.sg.fem
reportáž.EFF
report.nom.sg.fem.
'He wrote a report about an earthquake.' – 'A report was written about the earthquake.'

5.2 Resultative constructions

The other constructions understood as the grammatical category of diathesis are less productive than passivization, but they are produced by a regular grammatical operation, which fact points out to their systemic (grammatical) nature. Resultative constructions display this character in both of their forms: so-called objective (Giger, 2003) and possessive forms. The auxiliary *být* 'to be' and passive participle are used for the former type (ex. (12) and

⁸The technical means how to mark this information is left aside here. We can only say that the feature reflecting the relation between the type of the participant and the surface subject must be included in the lexicon.

(13)); the auxiliary *mít* ‘to have’ and passive participle constitute the latter type (ex. (14)).⁹

- (12) *Je otevířeno.*
 Is.3.sg.pres **opened**.nom.sg.neut.
 ‘It is **opened**.’
- (13) *Je zajiřřeno, ře*
 Is.3.sg.pres **arranged**.nom.sg.neut, that
dřkan na schůzi
 dean.nom.sg.anim for meeting.acc.sg.fem
přijde.
 come.3.sg.fut.
 ‘It is **arranged** that the dean will come for the meeting.’
- (14) *Dohodu o spolupřáci*
 Agreement.acc.sg.fem about cooperation.loc.sg.fem
uř máme podepsánu.
 already have.1.p.presl **signed**.acc.sg.fem.
 ‘We **have** an agreement about the cooperation **signed**.’

The form for objective resultative is ambiguous with the passive form. However, the slight semantic difference is reflected here by the grammateme values passive vs. resultative1 (see Table 1); cf. the ex. (15) and (16):

- (15) *Bylo navřřeno, aby se*
 Was.3.sg.pst.neut **proposed**.nom.sg.neut, that REFL
o změně zákona
 about change.loc.sg.fem law.gen.sg.inan
hlasovalo ihned.
 voted.3.sg.cond.neut immediately.
 ‘It **was proposed** to vote about the change of the law immediately.’
- (16) *Tento zákon se pořád*
 This.nom.sg.inan law.nom.sg.inan REFL still
pouřívá, ačkoli byl
 uses.3.sg.pres, though was.3.sg.pst.inan
navřřen už dávno.
proposed.nom.sg.inan already long_time_ago.
 ‘This law is still used, though it **has been proposed** long time ago.’

These constructions are rarely used for intransitives and for verbs in imperfective aspect (ex. (17) and (18)) (Načeva Marvanová, 2010). The syntactic rules for the objective resultative are based either on the shift of the PAT into the position of the surface

subject, or on the surface deletion of the PAT. In the possessive form either ACT or ADDR converts into surface subject.¹⁰ The mark about compatibility of the verb with the resultative meanings (grammateme values resultative1, resultative2) must be introduced into the lexicon, see Table 1.

- (17) *má nakročeno*
 has.3.sg.pres **stepped forward**.nom.sg.neut
 ‘he **has stepped forward**’
- (18) *Toto území máme*
 This.acc.sg.neut area.acc.sg.neut **have**.1.pl.pres
chráňeno před povodněmi.
protected.acc.sg.neut against floods.instr.pl.fem.
 ‘We **have** this area **protected** against the flood.’

5.3 Recipient diathesis

The recipient diathesis is a more limited category than the resultativeness, see also Daneř (1968), Panevová – řevčříková (2010) and Panevová (in press); however, it is again a result of the syntactic process constituting a recipient paradigm. An ADDR of the verb is shifted to the position of surface subject, the auxiliary verb *dostat* ‘to get’, marginally *mít* ‘to have’ with passive participle are used in recipient diathesis (ex. (19)). The semantic features of verbs (such as verb of giving, permitting etc.) are responsible for the applicability of this diathesis rather than the presence of ADDR in the verbal frame (ex. (20)). The mark about a possibility to apply the recipient diathesis will be a part of the lexical information within the respective verbs (see Table 1).

- (19) *Pavel dostal za*
 Paul.nom.sg.anim **got**.3.sg.pst.anim for
posudek zapláceno.
 review.acc.sg.inan **payed**.acc.sg.neut.
 ‘Paul **got payed** for the review.’
- (20) *řřkal někomu*
told.3.sg.pst.anim somebody.dat.sg.anim
*něco – *dostal*
 something.acc.sg.neut – **got**.3.sg.pst.anim
něco řřčeno
 something.acc.sg.neut **told**.acc.sg.neut
 ‘he **told** somebody something’ – ‘he **got told** something’

⁹These constructions studied from the contrastive Slavic view are called “new perfect constructions” by Clancy (2010); see, however, the description of these constructions given by Mathesius (1925).

¹⁰The shift of the ADDR into the subject position is a syntactic operation and it is not connected with the participant shifting; however, the subject has a function of the possessor of the resulting event rather than an agentive role.

PLATIT-1 ‘TO PAY-1’	
Formal morphology:	Vf - - - - - A - - - -
Aspect:	processual (→ complex ZAPLATIT)
Grammatemes:	+passive: PAT ^{Sb} +resultative1, +resultative2, +recipient
Reflexivity:	cor3
Reciprocity:	ACT – ADDR
Valency frame:	ACT(Nom) (ADDR(Dat)) PAT(Acc) (EFF(od+Gen/za+Acc))
Semantic class:	exchange
ŘÍCI-1 ‘TO TELL-1’	
Formal morphology:	Vf - - - - - A - - - -
Aspect:	complex (→ processual ŘÍKAT)
Grammatemes:	+passive: EFF ^{Sb} +resultative1, -resultative2, -recipient
Reflexivity:	cor3
Reciprocity:	ACT – ADDR
Valency frame:	ACT(Nom) ADDR(Dat) (PAT(o+Loc)) EFF(4/V-ind(assert/deliber),imper)
Semantic class:	communication

Table 1: Examples of lexical entries in the lexicon – a preliminary proposal

5.4 Grammatical diatheses in the lexicon and grammar

The exemplified diatheses belong to the grammatemes in FGD, representing the morphological meanings of specific analytical verbal forms. They differ from fully grammaticalized paradigmatic verbal categories (such as verbal tense or mood) in this respect that they use for their constitution not only morphemic means, but also syntactic operations. Due to this fact they are not applicable for all verbs and for their application a lexical specification in the lexicon is needed as well as general syntactic rules in the grammatical module.

Examples of lexical entries according to the suggestions in Sect. 3 to 5 are given in Table 1. In the Table the following notation is used:

- the number accompanying the lemma delimits the particular meaning of an ambiguous lexical item,
- formal morphological features of the lemma are described by the positional tag,
- processual and complex are the tectogrammatical values of the grammateme aspect corresponding to the imperfective and perfective aspect, respectively; a link to the aspectual counterpart is included,
- +/- with a grammateme value indicate the non/applicability of this value,
- for passive the participant converting into subject of the passive construction (if present) is marked as the upper index,

- resultative1, resultative2, and recipient (objective resultative, possessive resultative, and recipient, respectively) are proposed as the values of selected grammatical diatheses whose non/applicability is expressed with +/-,
- the reflexivity value cor3 means that reflexive binding ACT – ADDR is possible with the verb,
- the reciprocity ACT – ADDR is to be interpreted as indicating that syntactic operation of reciprocity can be applied if ACT and ADDR play both roles in this event,
- as for the valency frame, obligatory participants are without brackets, optional participants are in brackets; the morphemic realization of noun participants in brackets is attached to the functor; the verbal participant filling the role of PAT or EFF is denoted as V with the possible modalities compatible with the verb (assert – factual indicative, deliber – nonfactual indicative, imper – imperative).

6 Pair/group meaning of Czech nouns

The last issue that we use to exemplify the interconnection of grammar and lexicon in FGD is related to the category of number of Czech nouns. For this category some “irregularities” occur in their paradigms. The pluralia tantum nouns as *nůžky* ‘scissors’, *brýle* ‘glasses’ perform the formal deviation – they use the same form (plural) for singular as well as for the plural. They constitute a closed class and the

feature for their number deviation must be included into the morphological zone of their respective lexical entry. With the nouns denoting collectives such as *listí* ‘leaves’, *mládež* ‘young people’ the plural forms are semantically excluded; however, they represent again the closed class, so that their semantic deviation have to be included into the semantic zone of their respective lexical entry as a feature blocking the plural value of the grammateme number in their prototypical usage.

6.1 Nouns with pair/group meaning

There are many nouns in Czech that refer by their plural forms prototypically to a pair or typical group of entities and not just to many of them, which is acknowledged as the ‘proper’ meaning of plurals in Czech. This is the case for nouns denoting the human body parts occurring in pairs or typical groups (e.g. *ruce* ‘hands’, *prsty* ‘fingers’), nouns denoting clothes and accessories related to these body parts (*ponožky* ‘socks’), further nouns denoting objects used or sold in collections or doses, such as *klíče* ‘keys’ and *sirky* ‘matches’.

In contrast to other languages (Corbett, 2000), in Czech the pairs or groups are expressed by common plural forms of these nouns, these nouns are not formally marked for the pair/group meaning. However, the pair/group meaning manifests in the form of the numeral in Czech. When denoting pair(s) or group(s) of entities, the nouns are compatible with so-called set numerals only (cf. *jedny ruce* ‘a pair of hands’, *patery sirky* ‘five boxes of matches’), while if they refer simply to a number of entities, they are accompanied with cardinal numerals (*dvě rukavice* ‘two gloves’, *pět sirek* ‘five matches’).

The primary meaning of the set numerals is to express different sorts of the entities denoted by the noun (cf. *dvoje sklenice – na bílé a červené víno* ‘two sets of glasses – for the white and red wine’). However, the same set numerals, if combined with pluralia tantum nouns, express either the amount of single entities (i.e. the same meaning which is expressed by cardinal numerals with most nouns), or the number of sorts, cf. *troje nůžky* ‘three types/pieces of scissors’. The set numerals in combination with the nouns which we are interested in in the present paper express the number of pairs or groups; it means that the set numerals are used

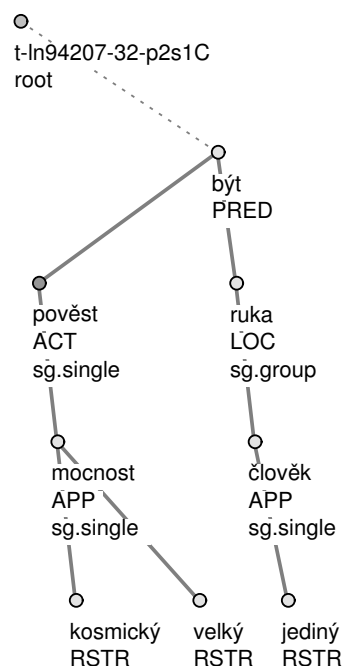


Figure 1: Tectogrammatical tree of the sentence *Pověst velké kosmické moci je v rukách jediného člověka.* ‘Reputation of the big space power is in the hands of a single man.’ For each node the tectogrammatical lemma, the functor and the values of the number and *typgroup* grammatememes are given.

here instead of cardinal numerals while the cardinals combined with these nouns express the number of single entities (cf. *troje boty* ‘three pairs of shoes’ vs. *tři boty* ‘three shoes’).

When considering how to treat the pair/group meaning within the framework of FGD, the fact was of crucial importance that this meaning is still connected with a list of typical pair/group nouns in Czech but not limited to them: If a set numeral co-occurs, the pair/group meaning can be expressed by most Czech concrete nouns, cf. ex. (21).¹¹

- (21) *Najdeme-li dvoje velké*
Find. 1.pl.fut-if two.sets big.acc.pl.fem
stopy a mezi nimi
traces.acc.pl.fem and between them.instr.pl.fem

¹¹Unlike the common co-occurrence of set numerals with nouns for which the pair/group meaning is not frequent, for the above mentioned nouns *ruce* ‘hands’, *klíče* etc. the ‘bare’ plural form is commonly interpreted as pair(s)/group(s) and the set numeral is used only if the concrete number of pairs or groups is important.

jedny menší, řekneme si:
one_set smaller.acc.pl.fem, say.1.pl.fut REFL:
rodina na výletě.
 family.nom.sg.fem on trip.loc.sg.inan.
 ‘If we find **two sets of big traces** and **one set of smaller ones** between them, we say: a family on a trip.’

6.2 Pair/group meaning as a grammaticalized feature

This fact has led us to the decision to treat the pair/group meaning as a grammaticalized category, namely as a special grammatical meaning of the plural forms of nouns (besides the simple plural meaning of several single entities), and to include it into the grammatical component of the description. If we had decided to capture the pair/group meaning as a lexical characteristic, it would have implied to split lexicon entries (at least) of the prototypical pair/group nouns into two entries, an entry with a common singular–plural opposition and an entry for cases in which the plural of the noun denotes pair(s) or group(s); the potential compatibility of the pair/group meaning with other nouns, though, would have remained unsolved. The economy of the lexicon seems to be the main advantage that can be achieved when preferring our solution to the lexicalist one in this particular case.

As a (newly established) grammatical meaning, the pair/group meaning has been introduced as a new grammateme *typgroup* in the grammatical module of FGD. For this grammateme three values were distinguished: *single* for noun occurrences denoting a single entity or a simple amount of them, *group* for cases in which pair(s) or group(s) are denoted, and *nr* for unresolvable cases.

The *typgroup* grammateme is closely related to the number grammateme (values *sg*, *pl*, *nr*). The following six combinations of a value of the grammateme number (given at the first position) with a value of the grammateme *typgroup* (at the second position) are possible according to the pilot manual annotation of the pair/group meaning carried out on the tectogrammatically annotated data of PDT 2.0: *sg-single*, *pl-single*, *sg-group*, *pl-group*, *nr-group*, and *nr-nr*, cf. ex. (22) to (27), respectively, and the tectogrammatical tree in Fig. 1.

- (22) *Na stole leží kniha*.sg-single
 On table.loc.sg.inan lies.3.sg.pres **book**.nom.sg.fem.
 ‘A **book** lies on the table.’
- (23) *Na stole leží knihy*.pl-single
 On table.loc.sg.inan lie.3.pl.pres **books**.nom.pl.fem.
 ‘**The books** lie on the table.’
- (24) *Namaloval to vlastníma rukama*.sg-group
 Draw.3.sg.pst.anim it own.instr.pl.fem
hands.instr.pl.fem.
 ‘He draw it by his **hands**.’
- (25) *Děti, zujte si boty*.pl-group!
 Kids.voc.pl.fem, take_off.2.pl.imp REFL
shoes.acc.pl.fem!
 ‘Kids, take off your **shoes**!’
- (26) *Vyčistil si boty*.nr-group
 Cleaned.3.sg.pst.anim REFL **shoes**.acc.pl.fem.
 ‘He has cleaned his **shoes**.’
- (27) *Odnes boty do opravy!*
 Take.2.sg.imp **shoes**.acc.pl.fem to repair.gen.sg.fem!
 ‘Take **the shoes** to a repair!’

7 Conclusions

In the present contribution we tried to illustrate that the balanced interplay between the grammatical and the lexical module of a language description is needed and document it by several concrete examples based on the data from the Czech language. In Sect. 3 the valency was introduced as an issue that must be necessarily included in the lexical entry of particular words; however, the valency is reflected in the grammatical part of the description as well, where the obligatoriness, optionality, surface deletions etc. must be taken into account.

Content clauses as valency slots of a special kind need a special treatment: The verbs governing content clauses classified as PAT or EFF require certain modality in the dependent clause (assertive, imperative, interrogative expressed on the surface by the subordinated conjunctions), only few verbs are compatible with more than one type of modality, e. g. *říci* ‘to tell’. These requirements (sometimes modified by morphological categories of the respective governor) are a part of valency information in the lexicon, while the rules for their realization by the conjunctions *že*, *zda*, *jestli*, *aby*, *ať* are a part of the grammar (Sect. 4).

Selected grammatical diathesis as a type of meanings of the verbal morphological categories are analyzed as to the constraints on their constitution (as a piece of information to be included in the lexicon) as well as to the regular syntactic operations applied on their participants (as a part of grammar; see Sect. 5). The arguments for an introduction of a new morphological grammateme (*typgroup*) connected with the category of the noun number are presented in Sect. 6. This meaning (with values *single* vs. *set*) is considered to be a grammaticalized category rather than a lexical characteristic of typical pair/group nouns.

Our considerations presented in Sect. 3 to 6 must be reflected within the technical apparatus of FGD both in the realization of lexical entries within the lexicon and in the shape of grammatical rules within the corresponding module.

Acknowledgments

The research reported on in the present paper was supported by the grants GA ČR P406/2010/0875, GA ČR 405/09/0278, and MŠMT ČR LC536.

References

- Anne Abeillé – Owen Rambow (eds.). 2000. *Tree Adjoining Grammars: Formalisms, Linguistic Analysis and Processing*. Center for the Study of Language and Information, Stanford.
- Kazimierz Ajdukiewicz. 1935. Die syntaktische Konnexität. *Studia Philosophica* I, 1–27.
- Jurij D. Apresjan – Igor' M. Boguslavskij – Leonid L. Iomdin – Vladimir Z. Sannikov. 2010. *Teoreticheskie problemy russkogo sintaksisa. Vzaimodejstvie grammatiki i slovarja*. Jazyki slavjanskich kul'tur, Moskva.
- Jaroslav Bauer. 1965. Souvětí s větami obsahovými. *Sborník prací filosofické fakulty brněnské university*, A 13:55–66.
- Leonard Bloomfield. 1933. *Language*. Henry Holt, New York.
- Noam Chomsky. 1970. Remarks on Nominalization. In *Readings in English Transformational Grammar*. Ginn and Company, Waltham, Mass., 184–221.
- Steven J. Clancy. 2010. *The Chain of Being and Having in Slavic*. Benjamins, Amsterdam – Philadelphia.
- Greville G. Corbett. 2000. *Number*. Cambridge University Press, Cambridge.
- František Daneš. 1968. Dostal jsem přidáno a podobné pasivní konstrukce. *Naše řeč*, 51:269–290.
- František Daneš. 1985. *Věta a text*. Academia, Praha.
- Markus Giger. 2003. *Resultativa im modernen Tschechischen*. Peter Lang, Bern – Berlin.
- Jan Hajič – Jarmila Panevová – Eva Hajičová – Petr Sgall – Petr Pajas – Jan Štěpánek – Jiří Havelka – Marie Mikulová – Zdeněk Žabokrtský – Magda Ševčíková Razímová. 2006. *Prague Dependency Treebank 2.0*. CD-ROM. Linguistic Data Consortium, Philadelphia.
- Jan Hajič – Jarmila Panevová – Zdeňka Uřešová – Alevtina Bémová – Veronika Kolářová – Petr Pajas. 2003. PDT-VALLEX: Creating a Large-coverage Valency Lexicon for Treebank Annotation. In *Proceedings of the 2nd Workshop on Treebanks and Linguistic Theories*. Vaxjo University Press, Vaxjo, 57–68.
- Václava Kettnerová. 2008. Czech Verbs of Communication with Respect to the Types of Dependent Content Clauses. *Prague Bulletin of Mathematical Linguistics*. 90:83–108.
- Markéta Lopatková – Zdeněk Žabokrtský – Václava Kettnerová et al. 2008. *Valenční slovník českých sloves*. Karolinum, Praha.
- Vilém Mathesius. 1925. Slovesné časy typu perfektního v hovorové češtině. *Naše řeč*, 9:200–202.
- Igor A. Mel'čuk. 1988. *Dependency Syntax: Theory and Practice*. University of New York Press, New York.
- Igor A. Mel'čuk. 2006. Explanatory Combinatorial Dictionary. In *Open Problems in Linguistics and Lexicography*. Polimetrica, Monza, 225–355.
- Mira Načeva Marvanová. 2010. *Perfektum v současné češtině*. NLN, Praha.
- Jarmila Panevová. 1974/75. On Verbal Frames in Functional Generative Description. *Prague Bulletin of Mathematical Linguistics*, 22:3–40, 23:17–52.
- Jarmila Panevová. 1994. Valency Frames and the Meaning of the Sentence. In *The Prague School of Structural and Functional Linguistics*. Benjamins, Amsterdam – Philadelphia, 223–243.
- Jarmila Panevová. In press. O rezultativnosti (především) v češtině. In *Zbornik Matice srpske*. Matice srpska, Novi Sad.
- Jarmila Panevová – Magda Ševčíková. 2010. Annotation of Morphological Meanings of Verbs Revisited. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*. ELRA, Paris, 1491–1498.
- Petr Sgall. 1967. *Generativní popis jazyka a česká deklinace*. Academia, Praha.
- Petr Sgall. 1998. Teorie valence a její formální zpracování. *Slovo a slovesnost*, 59:15–29.
- Petr Sgall – Eva Hajičová – Jarmila Panevová. 1986. *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Reidel – Academia, Dordrecht – Praha.
- Nad' a Svozilová – Hana Prouzová – Anna Jirsová. 1997. *Slovesa pro praxi*. Academia, Praha.