

The Role of Grammatical Constraints in Lexical Component in Functional Generative Description

Jarmila Panevová, Magda Ševčíková

Charles University in Prague, Faculty of Mathematics and Physics
Malostranské nám. 25, 118 00 Prague
{panevova,sevcikova}@ufal.mff.cuni.cz

Abstract

The present paper deals with selected morphological and syntactic features of Czech verbs. Working within the framework of Functional Generative Description (FGD), we demonstrate which features of lexical entries are required by the syntactic component of the description. In addition to the passive voice, traditionally described as diathesis, we briefly describe other kinds of proposed diatheses (resultative-1, resultative-2, and recipient). The constraints for their application will be present as features in the corresponding lexical entry; they will be a part of verbal paradigm in formal morphology. Regular operations within hierarchy of valency participants and their surface-syntactic positions are introduced into the grammatical component. Reciprocalization is characterized as a kind of shifting of valency complementations into the surface-syntactic position. We also specify the requirements of the verbs governing the infinitive and content clauses and point out to the interplay between the governing verb and modality of the content clause.

Keywords

Deep-syntactic structure, passive voice, diatheses, resultative, recipient diathesis, reciprocalization, modality of the dependent content clauses.

1 Introduction

Recent linguistic models are based on the division of labor between the lexical and the grammatical component. Though both components have been considered indispensable, individual linguistic approaches usually declare one of them as more central and important; the respective component is then elaborated more extensively as for the scope and depth of the issues involved (cf., for instance, the prevailing concern with grammatical issues in Chomsky's generative approach).

Since its original proposal (Sgall, 1967), the Praguian Functional Generative Description (FGD) adopts both the lexical and grammatical module; nevertheless, the main focus has been laid on the grammatical, esp. syntactic issues (Sgall et al., 1986). During the elaboration of the theory and most importantly during the application of the theory to the building up of the Prague Dependency Treebank (Hajič et al., 2001; Hajič et al., 2006), the lexicon has turned out to be of crucial importance. FGD is a multi-level description of language, where synonymous sentences are represented by the same representation on its uppermost level (tectogrammatrics). On the contrary, ambiguous sentences have different representations on the tectogrammatical level, while they differ on some of the lower levels of representation.

In the present paper we want to demonstrate several issues where the grammatical component strongly requires an introduction of particular features and data in the lexicon. The aim of these constraints is to block the generation of ill-formed sentences and to contribute to the theoretical description of the syntactic and morphological properties of the verbs. The structure of the lexical entry has been studied in connection with the treatment of valency, coreference between valency complementations of the governing and the embedded verb (Section 2). In Section 3 several examples of lexical entries and the syntactic rules cooperating with them are given.

2 Reflections of the Grammatical Constraints in the Lexical Component of FGD

2.1 Valency Frames of Verbs and Diatheses

In FGD, a lexical entry of a verb contains a valency frame, consisting of inner participants (Actor, Patient, Addressee, Origin, Effect) and those free modifications that are determined as semantically obligatory for the respective verb by the so-called dialogue test (e.g. the modification of manner for the verb *chovat se* ‘to behave’; Panevová, 1974/75). The inner participants are classified as (semantically) obligatory or optional with respect to the verb, for each of the inner participants its (morphological) form is further specified. The following features related to the valency and to the properties of the inner participants and obligatory modifications as well as their flexibility to express particular grammatical diatheses are to be specified in the lexical entry of the verb.

- A. Surface **deletability** of a valency complementation which does not lead to ungrammaticality and which is not a textual ellipsis, see ex. (1).
- B. Differences in valency behavior between **aspectual counterparts** (ex. (2a) vs. (2b)).
- C. **Lexicalization** of some meanings of the verb that influence the valency, see ex. (3), where the Patient *sluchátko* ‘receiver’ is implied.
- D. Possibility of **generalization (Gen)** of an inner participant, see ex. (4), – which means that everything written by the author is witty.
- E. Though the **passive diathesis** is productive enough in Czech, there are some verbs (intransitive stative verbs, reflexives and some other), which cannot be passivized (e.g.

běžet ‘to run’, *spát* ‘to sleep’, *stát* ‘to stand’, *bát se* ‘to be afraid’, *plakat* ‘to cry’, *lhát* ‘to lie’). There are also some transitives without passivization (e.g. *mít* ‘to have’; *pít* ‘to drink’ in imperfective aspect). The feature “-pass” will be assigned to them in the lexical entry. Moreover, the verbs participating in the passive diathesis differ in which participant is shifted to the subject position of the passive construction: Patient is involved with the verb *přeložit* ‘to translate’ (ex. (5)), Addressee is converted into the subject with *informovat* ‘to inform’ (ex. (6)), the shift of an Effect into the subject position can be seen with *psát/napsat* ‘to write’ (ex. (7)).

- F. The **resultative diathesis** is a less productive, though still grammaticalized category of Czech verbs (Mathesius, 1925). It has two variants: the objective resultative (*res1*) consisting of the auxiliary *být* ‘to be’ and a passive participle, and the possessive resultative (*res2*) with the auxiliary *mít* ‘to have’. The differences between these two types (syntactic and semantic) are described in Panevová (2011). Some of them could be understood from Sect. 3 (Ex. 3.1 and 3.2). The possible participation of the verb *otevřít* ‘to open’ in the possessive resultative (*obchod má otevřeno od 8 hodin* ‘the shop is opened since 8 o’clock’) will be included in its lexical entry as the feature “+res2”. The resultative is prototypically used with perfective aspect of transitive verbs, while imperfective verbs with the stative meaning usually do not participate in this category (**je/bylo spáno* ‘it-is/was sleeping’, **je/má leženo* ‘it-is/he-has laid’, **je/má chlubeno* ‘it-is/he-has boasted’). However, there are exceptions of the resultative combined with imperfective verbs (*je/má chráněno* ‘it-is/he-has saved’) and with intransitive verbs (*je/má namířeno* ‘it-is/he-has aimed’, *je/má našlápnuto* ‘it-is/he-has trodden on’). Thus the possibility of the lexical item to form this category is to be marked in the lexicon by the feature “+res1” (for the objective resultative), see (7), or by “+res2” (for the possessive resultative), see (8), (9), (10).
- G. The number of verbs participating in **the recipient diathesis** is more limited than the number of verbs with the resultative diatheses; however, the recipient forms are constituted paradigmatically with the verbs which enter the semantic groups listed in Daneš (1985) and in Panevová et al. (ms.), see examples (11), (12).

- (1) *Moji přátelé* *právě* *přijeli.*
(my friend-NOM-PL just arrive-PST-PFV)
‘My friends have just arrived.’
- (2a) *Včera* *Jan* *četl* *až do půlnoci.*
(yesterday John-NOM read-PST-IPFV till midnight)
‘Yesterday, John read till midnight.’
- (2b) **Včera* *Jan* *přečetl* *až do půlnoci.*
(yesterday John read-PST-PVF till midnight)
- (3) *Jan* *rychle* *zavěsil.*
(John quickly hang-PST-PVF)
‘John has hung up quickly.’

- (4) *Autor M.K. píše vtipně.*
 (author M.K. write-PRS-IPFV witty)
 ‘Author M.K. writes with wit.’
- (5) *Tento román byl překladatelem přeložen*
 (this novel_Patient-NOM be-AUX-PST translator-INS translate-PTCP-PASS
neadekvátně.
 inadequately)
 ‘This novel was translated by the translator inadequately.’
- (6) *Turisté byli informováni průvodcem*
 (tourist_Addressee-NOM be-AUX-PST inform-PTCP-PASS guide-INS
o historii zámku.
 about history castle)
 ‘The tourists were informed by the guide about the history of the castle.’
- (7) *O tom hrozném neštěstí byla novináři*
 (about this terrible accident_Effect be-AUX-PST journalist-PL-INS
napsána jen stručná zmínka.
 write-PTCP-PASS –F only short remark-NOM-F)
 ‘Only a short remark was written by journalists about this terrible accident.’
- (8) *Na neděli už je uvařeno.*
 (for Saturday already be-AUX-PRS cook-PTCP-PASS-SG-N)
 ‘It is already cooked for Saturday.’
- (9) *Matka už má na neděli oběd*
 (Mother already have-AUX-PRS for Saturday lunch-ACC-M
uvařen.
 cook-PTCP-PASS-M)
 ‘Mother already has cooked a lunch for Saturday.’
- (10) *Jan má posun zkoušky schválen*
 (John have-AUX-PRS shift-ACC-M exam-GEN confirm-PTCP-PASS-M
děkanem.
 dean-INS)
 ‘John has the shift of his exam confirmed by the dean.’
- (11) *Očividně dostal dávno odpuštěno. (SYN2005)*
 (Obviously get-AUX-PST-M long time ago excuse-PTCP-PASS-N)
 ‘He obviously has got to be excused long time ago.’
- (12) *V domově budou mít obyvatelé*
 (in hostel be-AUX-FUT have-AUX-INF inhabitant-PL
zajištěno nejen ubytování, ale i stravu. (SYN2006PUB)
 arrange-PTCP-PASS-N not only accomodation but also food)

‘In the hostel the inhabitants will have not only accommodation, but also food arranged.’

2.2 Valency Frames and Coreference

Verbs that take an infinitive construction as a complementation in a special valency position require the coreference between the participant triggered in the valency frame and the implied subject of the infinitive. Such requirements must be reflected in its valency frame. The member of valency frame controlling the (unexpressed) subject of its infinitive complement (as its antecedent) is marked in the lexical entry by the upper index “-er” (controller): *bát se* ‘to be afraid’ *Actor^{-er} (NOM), Patient (GEN/INF/Clause)*; *nařídít* ‘to order’ *Actor (NOM), Addressee^{-er} (DAT), Patient (Clause/INF)*. We present here only examples of two types of coreference between the controller and its controlee as prototypes of the requirements for the infinitive constructions in valency positions which are to be reflected in the given lexical entry; other types of Czech infinitive constructions with different types of coreference (control) are described in Panevová (1998). In (13), (14), the verb *bát se* ‘to be afraid’ in one of its meanings requires the identity (coreference) between its Actor and the subject of the embedded infinitive, while in (15), (16) for *nařídít* ‘to order’ the coreference between its Addressee and the subject of the infinitive is required. The differences between (13) and (14) and between (15) and (16) documents the fact that the role of the controlee is filled by the surface (unexpressed) subject:

- (13) *Jan_i se bojí [Sb_i] jít do lesa sám.*
 (John se-REFL is-afraid go-INF to forest alone)
 ‘John_i is afraid [Sb_i] to go alone to the forest.’
- (14) *Jan_i se nebojí [Sb_i] být zařazen do družstva pokročilých.*
 (John se-REFL is-not-afraid be-AUX include-PTCP-PASS
 in team advanced]
 ‘John is not afraid to be included in the advanced team.’
- (15) *Učitel_i nařídil studentům_j [Sb_j] zorganizovat soutěž v matematice.*
 (teacher order-PST student-PL-DAT organize-INF competition in
 mathematics]
 ‘The teacher_i ordered the students_j [Sb_j] to organize a competition in mathematics.’
- (16) *Rodiče nařídili synovi být rychle připraven k odjezdu domů.*
 (parents order-PST son-SG-DAT be-AUX-INF quickly
 prepare-PTCP-PASS to leaving home)
 ‘Parents ordered to their son to be quickly prepared for leaving for home.’

2.3 Valency Frames and Reciprocity

Another item which has been included into the lexical entry of Czech verbs in the lexical component of FGD is the information on the ability of valency complementations of the given

verb to enter a reciprocal relation. This ability is marked within the valency frame of the verbs and nouns by the “*Rcp*” index attached to the respective complementations.

The verb *líbat/políbit* ‘to kiss’, which has the valency frame *Actor^{Rcp} (NOM) Patient^{Rcp} (ACC)*, occurs in its basic (non-reciprocal) usage in the sentence *Pavel políbil Evu* ‘Paul kissed Eva’. The reciprocalization (according to the *Rcp* indices) results in the sentence *Pavel a Eva se políbili* ‘Paul and Eva kissed each other’ (which is to be interpreted that Pavel kissed Eva and at the same time Eva kissed Paul). For the discussion about the boundary between “inherent reciprocals” and reciprocity diathesis see Panevová (1999) and Panevová & Mikulová (2007). The attachment of the *Rcp* index has the following syntactic consequences:

- one of the involved valency slots is omitted,
- the lexeme from the omitted slot becomes a part of a coordinated subject or the subject is in plural,
- the reflexive form of the verb is to be used (if the verb itself is not a reflexive tantum or a derived reflexive, see Panevová, 2008),
- optionally, the lexeme *vzájemně/navzájem* [mutually/one another] etc. can be added into the sentence.

The classification of reciprocalization within the FGD approach is in accordance with Meřčuk’s (2006a: 215) arguments why reciprocals should not be classified as a voice.

2.4 Valency Frames and Modality of Dependent Content Clauses

Inner participants of some verbs can be expressed by a dependent (so-called content) clause. At the tectogrammatical level, dependent content clauses are classified as a Patient or an Effect with most verbs, less often as an Actor, and rather rarely as an Addressee or an Origin. The dependency of the content clause on the governing verb is expressed by a subordinating conjunction or by a pronoun (a pronominal adverb/numeral). The choice of the conjunction or pronominal is connected with the semantic properties of the governing verbs and with the modality of the dependent content clause.

A detailed analysis of the PDT 2.0 data has demonstrated that most of the verbs are compatible with a dependent content clause of one modality only (mostly with declarative modality, substantially less frequently an imperative or an interrogative dependent content clauses occur; cf. ex. (17) to (19), respectively). Only with a restricted number of verbs dependent content clauses of other modality types were used, most of them belong to verbs of communication; for instance, *diskutovat* ‘to discuss’ or *upozornit* ‘to point out’ (ex. (20a) with a declarative clause and (20b) with an imperative clause). Information on which modality type the verb is compatible with is proposed to be involved in the lexical entry of the respective verb. Three values for the description of the modality of dependent content clauses have been introduced: *declarative*, *imperative* and *interrogative*.

Dependent content clauses that express declarative modality are prototypically introduced by the conjunction *že* ‘that’, imperative dependent clauses by the conjunctions *aby* and *ať* ‘so that’, interrogative content clauses by the conjunctions *zda*, *zdali*, *jestli*, *-li* ‘whether/if’. The

conjunction listed for imperative clauses as well as the conjunctions of interrogative clauses are considered synonymous.

Description of the relatively transparent relations between the governing verb, the modality of their dependent content clauses and the conjunction used in these clauses is complicated by the fact that there are verbs in Czech with which modality of the dependent content clause (and thus the conjunction) changes depending on the change of grammatical categories of the governing verb (ex. (21a) with the indicative governing verb and (21b) with the conditional).

With verbs like *upozornit* (ex. (20a,b)), several modality values are to be listed in the lexicon since the compatibility of these verbs with dependent content clauses of different modalities is involved in the lexical meaning of the verbs. On the contrary, only the basic modality is to be marked in the lexical entry of the verb *uvítat* in (21a,b) (i.e. *declarative*).

(17) *Na závěr schůzky ředitel dodal, že smlouva bude podepsána do týdne.* <declarative>
 ‘In the end of the meeting, the director added **that** the contract will be signed in a week.’

(18) *Učitel nařídil žákům, ať zůstanou ve třídě.* <imperative>
 ‘The teacher ordered the pupils **that** they **should** stay in the classroom.’

(19) *Studenti se ptají, zda se zítra koná přednáška.* <interrogative>
 ‘The students are asking **whether** the talk is given tomorrow.’

(20a) *Upozornil je, že večerní představení začíná o hodinu později.* <declarative>
 ‘He pointed out to them **that** the evening performance begins an hour later.’

(20b) *Upozornil je, aby o této skutečnosti nehovořili.* <imperative>
 ‘He pointed out to them **that** they **should** not speak about this fact.’

(21a) *Opozice uvítala, že prezident zákon podepsal.* <declarative>
 ‘Opposition welcomed **that** the president had signed the law.’

(21b) *Opozice by uvítala, aby prezident zákon podepsal.* <imperative>
 ‘Opposition **would** welcome **that** the president would sign the law.’

3 Examples of Lexical Entries and Grammatical Rules Operating on Them

The Czech verb *připravit* / *připravovat* ‘to prepare’ has the valency frame (for one of its meanings) and other features analyzed in Sect. 2:

připravovat-IPFV / *připravit*-PFV ^{+pas, res1, res2} Actor (NOM), Patient ^{Sb}(ACC), Effect ^{opt}(k + DAT, na + ACC)

For the generation of the sentence in Ex. 3.1 and Ex. 3.2 with two types of the *res2* diathesis, the syntactic rules in (Rule I) are applied:

Ex. 3.1

Referent už má připraveny slidy na prezentaci.
(speaker already have-AUX prepare-PTCP/PASS slide-PL-M for presentation)
'The speaker already has his slides for presentation prepared.'

Ex. 3.2

Pavel má od matky připravenu večeři.
(Paul have-AUX from mother-GEN prepare-PTCP-PASS-F dinner-ACC)
'Paul has the dinner prepared by his mother'

Rule I

- (i) Predicate → AUX-mít_i + -n / -t participle V_j (*připraveny* 'prepared')
- (ii) Actor → Sb_i (*referent* [speaker]) / ADV (od + GEN) (*od matky* 'from mother')
- (iii) Addressee → Ø (for the sentence in Ex. 3.1 without an Addressee)
Addressee → Sb_i (*Pavel* 'Paul') (for the sentence in Ex. 3.2)
- (iv) Patient (N_j - ACC) → Obj_j -ACC (*slidy* 'slides' / *večeři* 'dinner')

The valency frame for one of the meanings of the Czech verb *slibit* [to promise] represents a lexical item with possible reciprocalisation and with several possible diatheses:

slibit-PFV / *slibovat*-IPFV^{+pas,+res1,+res2-+recipient} 'to promise' Actor^{Rcp}(NOM), Patient^{Sb}
(ACC/Clause/INF), Addressee^{Rcp}

The verb *slibit* is compatible with the passive, both resultative and recipient diathesis. The rules for reciprocalization are described in an informal way in Section 2.3. The sentence in Ex. 3.3 with the reciprocity relation between the Actor and Addressee and Ex. 3.4 in recipient diathesis could be generated by them, the Rule II is applied for the generation of Ex. 3.4.

Ex. 3.3

Pavel a Táňa si slibili věrnost.
(Paul and Tanya si-REFL promise-PST-PL faithfulness)
'Paul and Tanya promised to be faithful each other.'

Ex. 3.4

Pavel dostal za aktivní účast slibenu
(Paul get-AUX for active participation promise-PTCP-PASS-F-SG-ACC
od trenéra odměnu.
from coach payment)
'Paul was promised to receive a payment for his active participation from the coach'

Rule II

- (i) Predicate → AUX-dostat_i + -n / -t participle V_j -SG-F-ACC (*slibenu* 'promised')
- (ii) Actor → ADV (od + GEN) (*od trenéra* 'from the coach')
- (iii) Addressee → Sb_i (*Pavel* 'Paul')
- (iv) Patient (N_j - ACC) → Obj_j - ACC (*odměnu* 'payment')

4 Conclusions

In the present paper, we focused on the role of the interplay between the lexical and grammatical component in the language description. Starting from an analysis based both on available theoretical descriptions and real corpus data, we have tried to explain that the analyzed grammatical categories of verb need to be treated adequately in the lexicon entries of the respective verb. In the lexical component of FGD an explicit mark of passivization with verbs has been introduced. The same treatment has been proposed for the resultative and recipient diathesis. In addition to the diathesis information, there are many features to be stored in the lexicon that are interconnected with individual valency complementations of the verb; for instance, surface deletability, possibility of generalization or possibility to be expressed by a dependent content clause with a certain modality must be specified for the respective complementations of individual verbs.

The proposed treatment of the issues discussed should allow for an economic and an effective interconnection of the grammatical and lexical module within the Functional Generative Description and should block ill-formed structures, for instance, at the output of the English-to-Czech machine translation procedure.

We agree with the comparison of FGD and MTT given by Žabokrtský (2005), where the similarities between these two models are described in detail. Multi-level and dependency approach are shared as well as Mel'čuk's idea that the deep-syntactic level unifies the synonymous sentence because synonymy is "one of the underlying intuitive notions for the whole of linguistics" (Mel'čuk, 2012: 48).

Acknowledgements

This paper was supported by the projects GA ČR P406/10/0875 and GA ČR P406/12/P175. This work has been using language resources developed and/or stored and/or distributed by the LINDAT-Clarin project of the Ministry of Education of the Czech Republic (project LM2010013).

Bibliography

- Daneš, F. 1985. *Věta a text*. Academia: Praha.
- Hajič, J. 2001. *Prague Dependency Treebank 1.0*. Philadelphia: Linguistic Data Consortium.
- Hajič, J. 2006. *Prague Dependency Treebank 2.0*. Philadelphia: Linguistic Data Consortium.
- Mathesius, V. 1925. Slovesné časy typu perfektivního v hovorové češtině. *Naše řeč*, 9:200–202.
- Mel'čuk, I. 2006a. *Aspects of the Theory of Morphology*. Berlin – New York: de Gruyter.
- Mel'čuk, I. 2012. *Semantics. From meaning to text*. Amsterdam – Philadelphia: John Benjamins Publishing House.

The Role of Grammatical Constraints in Lexical Component in Functional Generative Description

Panevová, J. 1974/75. On Verbal Frames in Functional Generative Description. Part I. *Prague Bulletin of Mathematical Linguistics*, 22:3–40, Part II, *Prague Bulletin of Mathematical Linguistics*, 23:17–52.

Panevová, J. 1999. Česká reciproční zájmena a slovesná valence. *Slovo a slovesnost*, 60:269–275.

Panevová, J. 2008. Problémy se slovanským reflexivem. In *Česká slavistika. Příspěvky k XIV. mezinárodnímu sjezdu slavistů*, *Slavia*, 77 :153–163.

Panevová, J. 2011. O rezultativnosti (zejména) v češtině. In: *Gramatika i leksika u slovenskim jezicima*. Novi Sad, Beograd: Matica Srbska, Institut za srpski jezik. 165 – 176

Panevová, J. & M. Mikulová. 2007. On Reciprocity. *The Prague Bulletin of Mathematical Linguistics*, 87:27–40.

Panevová, J. et al. (ms.) *Mluvnice současné češtiny II. Syntax na základě anotovaného korpusu*. Praha: Karolinum.

Sgall, P. 1967. *Generativní popis jazyka a česká deklinace*. Praha: Academia.

Sgall, P. & E. Hajičová & J. Panevová. 1986. *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Dordrecht: Reidel Publ. Company.

Žabokrtský, Z. 2005. Resemblances between Meaning-Text Theory and Functional Generative Description. In *Proceedings of the 2nd International Conference of Meaning-Text Theory*. 549–557. Moskva: Slavic Culture Languages Publishers House.