The verbal valency in the Prague Dependency Treebank from the annotator's point of view

Zdeňka Urešová
uresova@ufal.mff.cuni.cz
ÚFAL & CKL MFF UK, Malostranské nám. 25, 118 00 Praha 1, Česká republika

Abstract

The core ingredient of the Prague Dependency Treebank (PDT; see Hajič, this volume) - „valency" - indicates the capability of lexical units to combine other complementations. The PDT has adopted the concept of the valency theory of the Functional Generative Description (FGD) (see Sgall, 1967, Sgall et al, 1986). The valency theory of the FGD has first been developed for verbs, then also for other parts of speech. We present a description of how we dealt with valency of verbs during the annotation of the PDT and the way how the verbal part of the valency lexicon (PDT-VALLEX) was built. We focus on some specific problems related to verbal valency (as well as some other verbal complementations) from the point of view of the PDT.

1. The concept of valency in PDT

One of the prerequisites of the correct syntactic annotation at the tectogrammatical level (TR) of the Prague Dependency Treebank (see Hajič, this volume) is the knowledge of ***valency frames***. The valency theory (see Panevová, 1974-75, 1980, 1994, 1999) as used in the process of annotation of the Prague Dependency Treebank (PDT) corresponds to the concepts of the Functional Generative Description (FGP) (see Sgall, 1967, Sgall et al, 1986). Within this approach, syntactic as well as semantic criteria are used to identify verbal complementations.

The verb is considered to be the core of the sentence (or clause, as the case may be). Its complementations (***dependents***) are classified either as ***inner participants*** or as ***free modifications***. Both types of verbal complementations can be either ***obligatory*** (semantically always present with a given verb) or ***optional*** (not necessarily present). Only inner participants (obligatory or optional) and obligatory free modifications belong to the verbal valency frame. Optional free modifications are not listed in the valency frame.[1,2,3] The relation between the dependent and its ***governor*** at the TR is labeled by a ***functor***. The functor must be determined and recorded for all complementations in the actual process of data annotation. Annotators choose[4] this value from a set of functors listed in the manual for tectogrammatical

[1] Neither are so called quasi-valency and typical complementations stored in the valency frames of the PDT-VALLEX lexicon (these types of complementations are described by Lopatková et al.(2003), Panevová (2003).

[2] We discuss here the verbal valency frame in a narrow, strict sense, i.e. the verbal valency frame captured in the lexicon. The verbal valency frame in a broader sense consists of all. complementations, which can expand the given verb. The types of all the complementations are captured in the structure of the annotated tree as some of the values of the dependent nodes.

[3] Valency is considered also for many nouns and adjectives, see Řezníčková,V. (2003), Hajič et. (2003).

[4] If the annotators hesitate among the correct value of the functor, they have the choice to mark this uncertainty through multiple selection of several functors.

annotation (see Hajičová et al., in prep.) The intersection of the set of functors used for valency modifications and non-valency modifications is not empty. Had we annotated in the corpus also which verbal complementations are obligatory and which are optional, we could then had simply extracted the valency frames of all verbs from the annotated sentences. However, in order to obtain the highest possible mutual agreement among the annotators and to keep consistency in the course of annotation, such a lexicon is already being built step-by-step *during* the annotation. It is shared by the annotators and it is gradually enlarged. This lexicon is called **PDT-VALLEX**.

2. The concept of the valency frame

Taking the basic principles (see Panevová, 1974-75 and the writings quoted above) as a starting point, we use criteria for **distinguishing inner participants and free modifications**, **the concept of shifting** of "cognitive roles" and **the dialogue test** for determining the obligatoriness of inner participants and free modifications.

2.1. Distinction between inner participants and free modifications

This distinction applies to the set of complementation types (functors) as a whole (i.e., if a functor is classified as an inner participant, it will be called inner participant in any valency frame (of any verb) in which it appears).

If a complementation type modifies the verb only once in any given clause (without regard to possible coordination or apposition) and it occurs just with particular verbs, which, in principle, can be listed, we call it an ***inner participant***. Five inner participants are distinguished at the tectogrammatical level in the PDT. (For a detailed discussion about the position of ADDR, ORIG and EFF see Lopatková, Panevová (this volume); the ideas proposed there are not yet taken into account in PDT-VALLEX):

- ACT (Actor),
- PAT (Patient),
- ADDR (Addressee),
- ORIG (Origin) and
- EFF (Effect).

Inner participants are determined semantically, except for the Actor (ACT) and (to a certain extent) also for the Patient (PAT). The first participant is always the Actor; the second one is always the Patient. Addressee (ADDR) is the semantic counterpart of an indirect object. As a rule, ADDR is animate (*promise something to somebody.*ADDR, *talk to somebody.*ADDR *about something, teach someone.*ADDR *something*). Effect (EFF) is the semantic counterpart of the second object or of the verbal attribute (*break something into something.*EFF, *appoint somebody as somebody.*EFF). Origin (ORIG) also comes from the second (or third or fourth) object, describing origin or something that is being transformed by the verb into something else (*create something from something.*ORIG, *translate something (a book) from Czech.*ORIG *to English*, *expect something from somebody*.ORIG).

On the other hand, if the same type of verbal complementation can be repeated within the same clause and if it can modify any verb (in principle), we call it a *free modification.* There are approximately 50 distinct free modifications used at the TR. The list comprises modifications of different kinds, such as local and directional (LOC, DIR1, DIR2, DIR3), temporal (TWHEN, TSIN, TTILL, TFL, TFHL, THO, TPAR, TFRWH, TOWH), manner

(MANN), intention (INTT) or causal (CAUS), etc. The full list of all functors is given in the annotation manual (see Hajičová et al., in prep.).

A secondary criterion for distinguishing the difference between an inner participant and a free modification is *government (rection)*. If the form of the dependent (morphological case, preposition, particular lexeme to be used etc.) is determined by the governing verb, it is considered to be an inner participant, if the dependent is independent in its form on the governing verb then it is considered to be a free modification.

Inner participants are subject to *shifting* but free modifications are not.

2.2. The concept of shifting of „cognitive roles“

In case the valency slots for Actor and Patient are not occupied (for the verb in question), ***shifting*** of participants takes place. The principle of shifting requires that if a verb has only one inner participant, it is always the Actor and if there are two inner participants of the verb, they are always the Actor and the Patient, regardless of their “semantics”. In case of three or more inner participants of one particular verb the first two are always Actor and Patient; for other than the first two slots, the above more or less semantic criteria are taken into account. For instance:

PAT shifted to the position of ACT: *The book*.PAT *came out*;
ADDR shifted to the position of PAT: *she understood him*.PAT;
ADDR shifted to the position of PAT, EFF stays in its slot: *elect him*.PAT *as a chairman*.EFF,
EFF shifted to the position of PAT: *to build a group*.PAT;
ORIG shifted to the position of PAT: *to act on the basis of a presupposition*.PAT.

The shifting does not apply only in some specific cases (see 4.10.)

2.3. The dialogue test

Inner participants and free modifications can be (at the tectogrammatical level) either obligatory or optional. The semantically obligatory dependent does not have to be present at the syntactic (analytical) level (AR); it can be omitted without the sentence becoming ungrammatical. However, the annotator has to restore this node in the tectogrammatical tree that represents the sentence at the TR. The obligatory complementations are thus always present at the TR despite their omission at the analytical level, which might even be the correct or preferred case: e.g., (pronominal) Actor is always an omissible member in the surface structure of a Czech sentence (Czech is a pro-drop language). (Some obligatory free modifications are also in general omissible in the surface realization, for example, in short answers to questions.) Therefore, the semantic obligatoriness cannot be determined by the surface form; but it can be examined by the *dialogue test* (see Panevová, 1999): the answer „I don’t know“ is not acceptable (it would disturb the smoothness of the dialogue) in case the complementation is semantically obligatory. For instance, the functors DIR3 (directional – where to) with the verb *to come* and DIR2 (directional – from where) with the verb *to leave* are obligatory. As long as the answer “I don’t know” is acceptable without disturbing the smoothness of the dialogue, we speak about an optional complementation (again, we mean optional in the Tectogrammatical Representation). For instance, the functors DIR2 (directional – from where) with the verb *to come* and CAUS (cause) with the verb *to leave* are optional.

It is necessary to point out that the application of the dialogue test was largely very helpful but for some verbs it deserves further discussion. Unfortunately there was no time during the process of annotation to construct special semantically related groups of verbs (see Levin, 1993) in order to assist the application of the dialogue test (under the assumption that such verbs behave in a similar way with regard to obligatoriness vs. optionality). We assume that by subsequently using the valency data for various tasks and applications we can achieve further refinement of the relevant criteria.

3. The process of creating verbal valency frames

3.1. Valency frame and its surface realization in the PDT-VALLEX lexicon

For each verb, the appropriate functor as well as its surface realization (surface-syntactic and morphological form) is recorded in every slot of its valency frame. In general, the mapping of the valency frame to its surface realization can be quite complex (see Hajič et al., 2003, Hajič, Urešová, 2003), but with a grain of salt we can assume that each of the valency members (slot fillers) can be mapped to its surface form independently. The surface realization through the morphemic case, preposition and morphemic case, and subordinate sentence with a conjunction is the most common.

For instance:

snížit

- valency frame: ACT(.1) PAT(.4) ?ORIG(z+2) ?EFF(na+4)
- example: *snížit nájem z 8 na 6 tisíc*
  (lit.: *lower the rent from 8 to 6 thousand)*

The question mark in front of the valency member in the above example denotes optionality, the other valency members are obligatory. The valency frame can be also empty - it means that the valency frame does not contain any valency member. For instance the verb *pršet (lit.: rain)* has an empty valency frame (written as EMPTY).

The surface realization of the valency frames is important information for automatic generation procedures of the surface structures as well as for the automatic "translation" of the analytic sentence representations to the tectogrammatical ones. The knowledge of the surface-syntactic realization is of course useful already during the manual annotation in order to distinguish the individual valency members (by being appropriately careful; in so doing one should not forget that the valency lexicon is simultaneously being created and verified during the process of annotation). For polysemic lexemes, the surface realization can indicate more or less subtle semantic differences and thus help the process of manual annotation by distinguishing individual valency frames (and therefore the individual senses or at least of groups of senses of the lexeme). The surface-syntactic realization is aimed at the analytical level of sentence representation (i.e., at the next lower level, where every surface word is represented by one annotation unit; we consider the morphological annotation to be part of the analytical level). All necessary conditions on part of speech or morphemic realization of individual members of verbal frames (or even specific lemmas, such as prepositions) should be specified. The original notation known from the literature on valency for tectogrammatical tree structures has been extended and an enriched formalized notation of surface realizations of individual valency members has been proposed. It captures not only the simple cases (such as the requirement for a certain morphemic case of the dependent member, regardless of the part of speech and other characteristics), but also the surface structure of idioms, often very complicated.

In order to describe the surface realization of the valency frame, first we have to capture the surface structure of this realization in the way it is represented at the analytical level of annotation (see Bémová et al., 1997). Square brackets are used to denote (analytical-level) dependency and comma is used for separating sister nodes: the governing node is written first, then there follows the opening square bracket ('['), the dependent $node_1$, dependent $node_2$, etc., then the closing square bracket (']'). The requirements on the part-of-speech and morphemic characteristics of individual nodes are written in a shorthand form (by means of a single character for each category) after the dividing symbol ´.´ (full stop) or ´:´ (colon) in the following order: part of speech, gender, number, case, degree of comparison and an agreement. For example, for the requirement of accusative we write .4, for plural locative .P6 etc. If any of these characteristics are missing, then it indicates that the given category can take any value in the annotation (with the exception of the first one, the major part-of-speech category, for which more complicated rules apply in case no concrete indication is present). The lemma (its analytical form, i.e. the form which corresponds to the morphological lexicon) is put, in case it is needed, in front of the separator: requirement for the preposition *s* (lit.: *with*) with instrumental looks like this: s[.7]. Some special symbols are used for capturing the omission of the member at the analytical layer. In order to shorten the realization in the most common case (which is the requirement for a preposition and a certain morphological case) an abbreviation "preposition+case" instead of "preposition[.case]" can be used (this is the way usually used in the literature, such as (Panevová, 1974-75)). The difference between a period and a colon as the separators of the lemma and the morphological part of the realization is as follows: the period determines the node of the corresponding analytical tree on which the nodes corresponding to the verbal complementations at the TR should depend (this difference is important mainly for complex phrasal slot descriptions).

For instance:

volat - frame: ACT(.1) PAT(.4) i.e. the Actor in Nominative, the Patient in Accusative

- example: *volejte telefonní číslo 205338* (vytáčet)
  (lit.: *call the phone number 205338*) (to dial)

- frame: ACT(.1) PAT(po+6) i.e. the Actor in Nominative, the Patient with the preposition "po" + locative

- example: *volat po otevřeném trhu* (vyžadovat, usilovat)
  (lit.: *clamour for the open market*) (to ask for, to cry out for)

Different meanings can have the same morphological realization of the valency frame; this is used just when a clear distinction between the meanings (senses) exists (see Lopatková, Panevová, this volume):

For instance[5]:

zakládat - frame: ACT(.1) PAT(.4)

-example: *zakládat sukni* (zkracovat)
(lit: *to shorten a skirt* (to shorten [by folding])

-frame: ACT(.1) PAT(.4)

-example: *zakládat stránky v knize* (označovat)
(lit: *to mark the pages in a book* (to mark)

[5] Please note that the "*zakládat*" valency frames quoted above are only examples and they do not represent all the existing valency frames of this verb.

For obligatory free modifications only, empty parenthesis may be used to denote any surface realization usual for the free modification in question. The realization of inner participants is always given in full, since there is no "standard" or "default" realization for any of them.[6]

Examples of realizations:

Simple morphological case (.1, .2, .3, .4, .5, .6, .7)[7]

Prepositional case (preposition without vocalization and the number of the required morphological case): na+4, k+3, o+6, …; or secondary preposition and the number of the required morphological case: e.g., prospěch[v,.2],[8] lit.: *to the benefit of*

Infinitive: (.f)

Subordinating conjunction: (že, aby, když, zda, jestli, ať, …; lit.: *that, to, when, whether, if, let,* …)

Subordinate clause without conjunction (.c); (if started for instance with an interrogative pronoun or adverb : který, proč, kde, kdy, …; lit.: *which, why, where, when,* …)

Adjective: (usually with a case, e.g., .a7)

Adverb: (.d)

Interjection: (.i)

Numeral: (.m)

Pronoun: (.p)

Construction with '*to be*' (*to be* and the required morphological case, e.g., být[.7])

Direct speech: (.s)

Any common ("standard" for given functor) realization: ()

State: (=)

Empty frame: (EMPTY)

The annotation of idioms (functor: DPHR) is much more complicated. Almost always, it is necessary to capture a particular lemma with an appropriate morphological case and often also with a number: *jít příkladem*: DPHR(příklad.S7) (lit. *go [by an] example*; *give an example*). A lemma with a required prepositional case occurs also very often: *lapat po dechu*: DPHR(po[dech.S6]) (lit. *catch [s-one's] breath*). The phrase is sometimes realized through even more complex (sets of) dependent subtrees: (někomu) *běhá mráz po zádech* (lit.: *[a] frost runs on [sb'] back*; *a shiver runs down sb's spine*): DPHR(mráz:S1,po[záda.P6]).

3.2. The process of building the PDT-VALLEX lexicon

[6] Frequency-wise, of course, some realizations are more frequent than others – for example, for ACT in an active verbal construction, the nominative case is used very often.

[7] Numbers are used in Czech grammars to denote cases: 1 for nominative, 2 for genitive, 3 for dative, 4 for accusative, 5 for vocative, 6 for locative, and 7 for instrumental.

[8] The conjunction "jako" (*as*) is also included in the list of the prepositions, as it requires a particular morphological case in some valency frames. For instance: *bral to jako problém* (lit: *he considered it as [to be] a problem*).

The annotators work primarily only with those verbs (or their senses) found in the PDT data. On the other hand, every occurrence of a verb in the corpus contains a reference to its valency frame (i.e., to an entry in the valency lexicon). The annotators insert the verbs (senses) found in the course of the annotation and their associated valency frames into the lexicon. They create the particular valency frame and write an example (or more examples) of its usage. If they find it reasonable, they can insert a note that refers to of another verb that has one of its valency frames related to the current one (a synonym/antonym, an aspectual counterpart, etc.).

Notes and comments on problems encountered during the creation and/or usage (annotation) of the valency frames can also be recorded.

4. Problems related to the verbal valency

Many problems and confusions naturally came out by verifying and adopting the valency theory to particular verbs during the annotation. Let us focus shortly on some of them.

4.1. Missing optional valency slots

It is natural that the annotators primarily include and describe valency slots according to their surface realization as it occurred in the data. That is the reason why a valency frame in the PDT-VALLEX sometimes might not contain an optional inner participant because it is difficult to determine such inner participants (and the dialogue test is of no help either, because it is not applicable for determining the optional slots). For instance, with the above mentioned verb *snížit* (*to lower)* only two inner participants (ACT and PAT) were listed first in the lexicon and only as late as the construction *to lower the rent from 8 to 6 thousand* occurred the frame was extended with optional inner participants ORIG and EFF. Similarly, a valency frame may not capture all possible morphemic realizations of the given valency slot; however, the valency frame should contain all morphemic realizations that occur in the annotated data. From this point of view, the complex Valency Lexicon VALLEX (see Straňáková-Lopatková and Žabokrtský, 2002) is more complete in describing valency frames in full, using the much bigger (yet syntactically unannotated) Czech National Corpus as its data base; for each verb, its entries are meant to contain all their meanings and all possible surface realizations (as well as some other additional information).

4.2 The competition between an inner participant and a free modification

Competition between two or more functors is understood to be a situation when a valency member occupies (meaning-wise) just one valency slot, but both (or more) functors apply (based on their "semantic" definitions). The current representation of the valency frame does not allow labeling one valency frame slot with more than one functor.

4.2.1. The competition between an ADDR and a LOC/DIR3/DIR1

The obligatory functors LOC (location– answer to a question "where?"), DIR3 (direction to - answer to "where to?") and DIR1 (direction from – answer to "from where?") compete in the valency frame of some verbs with an ADDR (Addressee).

For instance:

podat

frame: ACT(.1) CPHR({přiznání …}.4) ADDR(.3)
podat přiznání úřadu… **to whom**
(lit:. *to-file a-tax-return [to] the-office(Dat)*)
frame: ACT(.1) CPHR({přiznání, …}.4) DIR3()
podat přiznání na úřad... **to where**
(lit.: *to-file a-tax-return into the-office*)
frame: ACT(.1) CPHR({přiznání, …}.4) LOC()
*podat přiznání na úřadě…***where**
(lit.: *to-file a-tax-return at the office*)

Other examples:
***ukrást** peníze bance*.ADDR / *z banky*.DIR1
(lit.: *to steal money the-bank(Dat)*.ADDR / *from the bank*.DIR1)
***odebrat** děti rodičům*.ADDR / *od rodičů*.DIR1
(lit.: *to-take-away the-kids the-parents(Dat)*.ADDR / *from parents*.DIR1)
***dát** listinu úřadu*.ADDR / *na úřad*.DIR3
(lit.: *to-give the-document the-office(Dat)*.ADDR. / *at-the-office*.DIR3)

It is clear that there is only one valency slot (the valency members cannot occur in the given sense simultaneously in one clause) with different morphemic realizations. Because of this different morphemic realizations *and* due to the current definitions of functors ADDR, LOC, DIR3, DIR1 *and* because of the fact that one valency slot cannot be occupied by more than one functor, it is necessary to create three different valency frames.[9]

If the corresponding surface realization is omitted from the actual sentence, it is difficult for the annotator to make a decision which of the competing functors has to be assigned to the restored (obligatory) node. By convention, Addressee (as an inner participant) has a priority, so a node labeled as the Addressee is added to the annotation in such a case.

4.2.2. The competition between ADDR a BEN

The border between the inner participant ADDR (Addressee) and the free modification BEN (Benefactive) is not often clear. The situation is easy in case the dative or the prepositional case "pro+4" *(for+Accusative case)* is present in the annotated clause. The dative is prototypically considered to be an Addressee, whereas the prepositional case "pro+4" is prototypically a Benefactive.

For instance:
***přinesl** jí*.ADDR *pro tatínka*.BEN *dopis*
(lit.: *[he] brought her(PronPers.Dat)*.ADDR *for [her] dad*.BEN *a-letter*).

The situation is more complicated if only one morphemic realization from the previous two is present in the given clause. We have thus used the following criteria for distinguishing an Addressee and a Benefactive:

The dative is prototypically an Addressee; however, a dependent in the dative is labeled Benefactive if the dative construction can be substituted with a possessive pronoun.

For instance:

[9] One might consider using a special "group functor", in this case for Addressee, Locative, and Directional in order to create just one valency frame. For issues of semantic and syntactic coherence see also (Levin, 2003) and (Kingsbury, Palmer, 2002).

barvit *jí.BEN vlasy*…její vlasy
(lit.: *to-color her(PronPers.Dat) hair…her(PronPoss) hair*)
amputova*t mu.BEN nohu*…jeho nohu
(lit.: *to-amputate him the-leg …his leg*)
líbat *jí.BEN ruku*…..její ruku
(lit.: *to-kiss her(PronPers.Dat) hand…her(PronPoss) hand*)
vidět *mu.BEN do duše*…jeho duše
(lit.: *to-see him into soul…his soul)*

This rule still has exceptions, however. For instance, substitution is possible in the following construction: ***odebral*** *nám tři body* (lit.: *he took [from] us(Dat) three points*), but the dative is still labeled as an Addressee here, namely because it is a valency member for the verb *odebrat* (the valency relation has always precedence.) The possibly occurring Directional could be here only a free modification, e.g., *odebral nám*.ADDR *tři body z tabulky*.DIR1 (lit.: *he took [from] us(Dat).*ADDR *three points from the-chart*.DIR1)

If there is an additional valency slot which is not an Addressee but it is a kind of Directional with the particular verb (see 4.2.1.), then the dative construction is labeled Benefactive.

For instance:
odebral *mu*.BEN *krev ze žíly*.DIR1
(lit.: *[he] took-away him*.BEN *the-blood from the-vein.*DIR1)
- frame: ACT(.1) PAT(.4) DIR1()

Compare:
odebrat tělu.ADDR potřebné látky
(lit.: *to take-away the-body(Dat).*ADDR *the-necessary substances*): an obligatory Addressee.
odebrat z těla.DIR1 potřebné látky
(lit.: *to take-away from the-body*.DIR1 *the-necessary substances away*): an obligatory Directional
odebrat mu.BEN z těla.DIR1 potřebné látky
(lit.: *to take-away him(Dat).*BEN *from the-body*.DIR1 *the-necessary substances*): an obligatory Directional, Benefactive in the dative case.

The prepositional form “pro+4”, while being prototypically a Benefactive, expresses an Addressee in case this form can be substituted by a dative without the change of meaning.

For instance:
přinášet *pro úřednici (=úřednici).*ADDR *dopis*
(lit.: *[to] bring for a-clerk (=[to] a-clerk(Dat)) a-letter*)
přivézt *pro maminku (=mamince).*ADDR *květiny*
(lit.: *[to] bring for mum (=[to] mum(Dat)) flowers*)

The adequacy of this treatment of the prepositional form “pro+4” (i.e., the possibility to label it as an Addressee) is attested by examples of coordination of the two different forms (“pro+4” and the prepositionless dative) which should be annotated by the same functor.

For instance:
poskytoval *mu bydlení a pro Alenu taky*

(lit.: *[he] provided him(Dat) accommodation and for Alena too*)
zajistil *nám pobyt a pro sebe taky*
(lit.: *[he] booked us(Dat) a-stay and for himself too*)
zaručil *nám i pro ně stejné podmínky*
(lit.: *[he] guaranteed us(Dat) and for them the-same conditions*)

The presence or absence of Benefactive and Addressee can also distinguish the meaning of the verb.

For instance:
nosil *mu.*BEN (kamarádovy) *batohy* (přenášet)
(lit.: *he carried him(Dat).*BEN *bags*) … (his bags to, e.g., save him work)
vs.
nosil *mu*.ADDR (kamarádovi) *batohy* (přinášet)
(lit.: *he was-bringing him(Dat).*ADDR *bags*) … (moving bags to his proximity)

4.2.3. The competition between ORIG a DIR1

Similar situation arises between the inner participant Origin (ORIG) and the free modification Directional (DIR1), which expresses the direction "from where". The problem lies in the question "from where?", which can, in many cases, indicate not only the Origin but also the Directional.

If the valency slot has in its surface realization description the form "od+2" (from+Genitive case), which is quite typical for Origin, and as another possibility also the form "z+2" ("from inside"+Genitive), which is typical for Directional, we prefer to label such slot as the (inner participant) Origin. We assume that both forms have the same semantics in such cases. This is displayed e.g. by verbs with the meaning "to gain something from somebody (= from somewhere)".

For instance:
čerpat *od kolegy / z textu informace*
(lit.: *[to] gather from a-colleague / from a-text information*)
obdržet *od úřadu / z úřadu povolení*
(lit.: *[to] receive from an-office / from an-officer a-permit*)
dostat *od banky / z banky finanční podporu*
(lit.: *[to] get from a-bank / from[-inside] a-bank financial support*)
(please note the homonymy of the English preposition "*from*" in Czech, cf. also below)
půjčit si *od banky / z banky peníze*
(lit.: *[to] borrow from a-bank / from[-inside] a-bank money*)

Sometimes both the prepositions (*od, z*) can appear in one clause in a text. One of the constructions will then be labeled as the free modification DIR1. It is up to the annotator to distinguish, based on the context as usual, the semantic difference between them.

For instance:
půjčil si *od tatínka.ORIG z účtu.DIR1 značnou sumu*
(lit.:*[he has] borrowed from [his] father.*ORIG *from the-account*.DIR1 *an-appreciable sum*).

If the verbal complementation can only be realized as "z+2", we assume this is the free modification Directional, not an Origin, and the Directional is here not a part of the valency frame. This rule applies e.g. for verbs with the meaning „to pay to somebody something from somewhere".

For instance:

financovat *stavbu z rozpočtu*
(lit.: *[to] finance the-construction from[-inside] the-budget.DIR1*)
hradit *náklady z fondu oprav*
(lit.: *[to] cover costs from[-inside] the-resources.DIR1 of-repair*)
dotovat *výdaje ze státních rezerv*
(lit.: *[to] supplement expenses from[-inside] the-state reserves.DIR1*)

However, if the complementation can be realized by the form "z+2" but there is also another valency member, namely Effect (mostly expressed by the prepositional forms "do+2" (to/into+Genitive), "v+4" (in/into+Accusative), "na+4" (on/to/onto/into+Accusative), we consider this to be a valency complementation and it is labeled Origin.

For instance:

překládat *z češtiny do němčiny*
(lit.: *[to] translate from Czech into German*)
změni*t účes z kudrn na rovné vlasy*
(lit.: *[to] change haircut from curler into straight hair*)
klesnout *z tisíce na pět set*
(lit.: *[to] sink from [one] thousand to five hundred*)

The situation is, however, more complicated in many other cases. The common meaning of "origin" (most often expressed by the prepositional constructions "z+2", "od+2" as discussed above) can become split into more frames with different functors assigned to the slot with this surface realization.

For instance, the verb ***pocházet*** (lit. *come-from*) ended up with three different frames, with the slot in question labeled PAT, DIR1 and TFRWH (temporal "from when"), respectively:

zboží ***pochází*** *z Prahy.PAT* (shifted from ORIG)
(lit.: *the goods come-from from[-inside] Prague)*
(in the sense "from local (Praguian) producers")
matka ***pocházela*** *z Moravy*.DIR1
(lit.: *[the] mother came-from from Moravia*)
kniha ***pochází*** *ze 12. století*.TFRWH
(lit.: *[the] book comes-from from [the] 12th century*)

By using the Origin or Directional functors in the valency frames we often distinguish an abstract and concrete meaning of a verb, respectively (see also 4.4.).

For instance:

přecházet *(cross [over], change, switch)*
přecházet z desetihodinového.ORIG na osmihodinový provoz.PAT
(lit.: *[to]change from ten-hour-long to eight-hour-long shifts*)
vs.

přecházet z jedné strany.DIR1 na druhou.DIR3
(lit.: *[to] cross [the street] from one side to the-other*)

***vymáčknout** (squeeze, press, get out)*
vymáčknout z obyvatel/od obyvatel.ORIG daně
(lit.: *[to] get-out from/from[-inside] the-dwellers the-taxes(Acc)*)
vs.
vymáčknout z citrónu.DIR1 šťávu
(lit.: *[to] press from[-inside] the-lemon the-juice(Acc)*)

The competition of ADDR and BEN described earlier and the competition of ORIG and DIR1 confirms the assumption that the semantic classification cannot always correspond straightforwardly to formal indicators (i.e., to the surface realization by prepositions, morphemic cases, etc.). It is obvious from the PDT annotation that a solution of the problem of the competition of certain functors is a very difficult task and it is not satisfactorily solved yet, both formally (should we allow for more valency frames for a single meaning, or should we use groups of functors, etc.?) and in the annotation practice. This problem and the current solution is considered open for further discussion and it undoubtedly requires a more detailed examination.[10]

4.3. Overlap of miscellaneous types of free modifications

The specification of the functors denoting free modifications is based on their semantics. It is not easy to define the particular functor fully unambiguously and "sharply". The annotators have to help themselves, for the sake of consistency of the annotation, with various criteria (often based on morphosyntactic rules) in the gray zone of an overlap of two (or sometimes even more) functors of miscellaneous types of free modifications.

4.3.1. Overlap of TWHEN and LOC

Even though this overlap seems improbable, it happens fairly often. To help us to resolve it, we have been using a transformation of the construction in question into predication: if the most natural transformation into a complex (subordinate) sentence opens with the time conjunction „*když*" (lit.: *when*), it indicates that a time functor should be assigned. The fact that the noun in the construction is an event noun can be used as a supportive criterion leading us to assign the temporal interpretation, too.

For instance:

***podlehli** v zápase*.TWHEN
(lit.: *[they have] lost in [the] fight*) "…, when they fought"
*v polemikách***.**TWHEN ***likvidoval** soupeře*
(lit.: *in the-argumentation [he] liquidated the-rivals*) "…, when he argued"
***oznámil** to v rozhovoru*.TWHEN
(lit.: *[he] announced it in a-talk*) "…, when he talked"
*akcie **patřily** v první vlně*.LOC *k nejatraktivnějším*
(lit.: *the-shares belonged in the-first wave to the-most-attractive*) "where did they …?"
v tomto příkladu.LOC ***nejde** o jednoduchou úlohu*
(lit.: *in this case is-not-the-matter about an-easy task*) "where it is not …?"

[10] For another account of the alternation of some types of valency complementations, see also (Benešová, 2002).

4.3.2. Overlap of INTT and LOC/DIR3/DIR1

The semantics of the governing verb (mostly a verb of motion) leads in some cases to an uncertainty whether Intention (INTT, a free modification) rather than an obligatory Direction (DIR1, DIR3) or Location (LOC) from the valency frame is not concerned.

For instance:

přišel *se-koupat*
(lit.: *[he] came [to] swim)*
došel *nakoupit*
(lit.: *[he] went [to] shop*),
vydat-se *na jahody*
(lit.: *[to] set-out for strawberries*),
dorazil *překonat record*
(lit.: *[he] arrived [to] beat the-record*),
odešel *se-rozcvičit*
(lit.: *[he] went [to]warm-up)*
zůstal *na oběd*
(lit.: *[he] stayed for lunch*)

The current version of the PDT-VALLEX lexicon prefers to use the Direction (as an obligatory slot) in the above cases as well as in other similar cases, since Intent was not considered to be an obligatory slot in the valency frame but just a "pure" free modification (cf. Lopatková, Panevová, this volume)[11]. On the other hand, some of its properties make its position on the inner participant↔free modification axis not quite clear. We are leaving it open whether INTT should not become an obligatory valency member in case where the original verbal meaning, i.e. intentional movement to somewhere or from somewhere, fades out to such an extent that the spatial meaning is irrelevant . For instance, in such collocations as *jdu se oženit* (lit.: *[I] am-going myself to-marry*) (I want, I mean), *jdu jí napsat* (lit.: *[I] am-going myself her to-write*) (again: I want, I mean), *jde si zapnout kabelku* (lit: *[she] goes herself to-zip-up [her] bag*) (she wants, she means), it is obvious that the voluntative modality of Intent (intent to do something) has a priority over the Direction. In this case the possibly occurring Direction (in the same clause) should be labelled as a free modification.

To summarize, INTT is currently always treated as optional (and because it is a free modification, it is never a member of a valency frame - cf. sect. 1 for definitions and principles), and due to a disagreement on a usable "semantic" definition it is being assigned rather on the basis of its morphemic realization (i.e., infinitive, "pro+4", "na+4" or "k+3") than that of semantics in almost all cases.[12,13]

---

[11] In the earlier works on verbal valency (esp. the works of Panevová quoted above), some free modifications were considered to have properties that would not make it counterintuitive to designate them as obligatory, such as BEN or INTT.

[12] With the exception of ***jít, chodit, vycházet*** *za prací* (lit. go, be going, go-out for/to work), where the surface realization „za+7" („for+Instrumental case") is used.

[13] Similarly, the decision whether to use INTT or AIM (Aim) depends solely on the form: while INTT has only been assigned in the cases just described, AIM has been used only if it was realized on the surface as a subordinate clause with the conjunction *aby* (lit.: *to*). This made the task easier for the annotators and the annotation is thus consistent, at the expense of hiding the semantic difference if it goes against the form. Lately, the specification of the difference between INTT and AIM was reconsidered to become less formal and more semantic (see Panevová, Lopatková, this volume), but such a treatment might only influence the future versions of the PDT.

4.4. Abstract and concrete meaning of surface directional expressions

Abstract and concrete usages of verbs are often distinguished in the lexicon by using separate valency frames. The original examples of "general directionality" are split into several valency frames in the PDT annotation.

For instance:

přijít

přijít ke stromu.DIR3	– "přistoupit" (to move close to)
(lit.: *[to] come to the-tree*)	
přijít k penězům.PAT	– "získat" (to get)
(lit.: *[to] come to the-money*)	
přijít na řešení.CPHR	– "napadnout", "vyřešit" (to solve)
(lit.: *[to] come onto a-solution*)	

Other examples:

ustoupit

ustoupit od zdi.DIR1	– "vzdálit se" (to move away from)
(lit.: *[to] go-away from a-wall*)	
ustoupit od myšlenky.PAT	– "vzdát se" (to abandon [an idea])
(lit.: *[to] go-away from an-idea*)	

vycházet

vycházet z lesa.DIR1	– "opustit a vzdálit se" (to move out of)
(lit.: *[to] come-out from the-forest*)	
vycházet z předpokladu.PAT	– "začít" (start with, from)
(lit.: *[to] start from an-assumption*)	

The consistency of annotation of this kind of problematic valency frames is low in the annotated data, since not all occurrences contain such clear-cut cases like the examples above. This group of valency frames is also considered open for further and more detailed examination. See also 4.2.3.

4.5. Co-occurrence of time and local complementations

Two local or time complementations such as *zítra k večeru* (lit.: *tomorrow towards evening), hluboko pod povrchem* (lit.: *deep under the-surface*) have their own specific character. It is difficult to treat them in the dependency syntax formalism because no clear dependency direction (and/or structure) can be established using the usual (omission-without-loss-of-grammaticality) criteria. Applying these theoretically-based criteria, during the annotation on a large amount of data, we failed to consistently and uniquely determine the governor and the dependent: it was found empirically that one particular member can be sometimes omitted (without making the sentence ungrammatical, with the usual caution) but such a consideration did not generalize well, because in many instances, neither the former or the latter part of such construction can be omitted.

The "grammatical" omission can take place e.g. when the time complementation is in Accusative (*Oblékla-se půl hodiny před začátkem představení*, lit.: *[she] dressed-up half(Acc.) an-hour before the-start of-the-performance)*, where "*půl hodiny*" can be omitted, but it could also be the other way round (*Strávila tam dva měsíce před porodem*, lit.: *[she] stayed there two months(Acc.) before the-delivery*) – here, only "*před porodem*" can be omitted. An example of

a case where neither part can be omitted is e.g. "*Leží to dva kilometry od řeky,"* (lit.: *[it] lies two kilometers away-from the-river*).

The currently used solution for annotation in PDT 2.0 is as follows: the first part is always considered to be the governor; it means that the first part is always being modified by the other local or time verbal complementation.

For instance:

***leží** to hluboko pod povrchem*
*(*lit.: *lies it deep under the-surface)*
***pojedeme** na západ od Prahy*
*(*lit.: *[we will] go to the-west from Prague)*
***dorazil** pět minut po odjezdu vlaku*
*(*lit.: *[he] arrived five minutes after the-departure of-the-train)*
***vrátí se** brzy po Vánocích*
*(*lit.: *[he will] return himself soon after Christmas)*

4.6. The functor "STATE"

A question arose during the annotation whether a modification that is semantically different, but formally identical to the LOC (cf. 1.) or DIR3 (cf. 2.) functor should be distinguished in the valency frames.

For instance:

ocitnout se
*ocitla se v Praze.*LOC
(lit.: *[she] found herself in Prague*)
vs.
ocitla se pod tlakem.???
(lit.: *[she] found herself under pressure*)

dostat se
dostala se do Brna.LOC
(lit: *[she] got to Brno*)
vs.
dostala se do maléru.???
(lit: *[she]got-involved herself in a-mishap*)

Here we believe it is not appropriate to follow only the morpho-syntactic considerations (both complementations would then get the functor LOC). That was the reason we preliminarily set up a new functor which would label this type of dependency as "State". So far this functor has not been used, but a special node attribute with a special value for State will serve this purpose, using the syntactically closest functor label. The annotator currently adds an alternative of „an undefined functor"(in such case a star appears by the primary functor of a node in the annotated data). A special symbol "=" is used in the valency lexicon so far.

For instance:

*dát věci **do souvislostí*** valency frame: ACT(.1) PAT(.4) DIR3(=)
(lit.: *[to] put things into perspective*)
*držel byt **v pořádku*** valency frame: ACT(.1) PAT(.4) LOC(=)
(lit.: *[he] kept the-flat in order*)

hnát řešení ***do krajností*** valency frame: ACT(.1) PAT(.4) DIR3(=)
(lit.: *[to] push the-solution into the-extremes*)

Other examples:
jít ***do likvidace***
(lit.: *[to] go into liquidation*)
nechat sportovce ***v klidu***
(lit.: *[to] leave the-sportsman at rest*)
odsouvat osobnost ***do zapomnění***
(lit.: *[to] shift a-personality into oblivion*)

It is important to say that the "functor" State requires further investigation from different points of view; various subtle semantic differences occur in such constructions and it is not yet clear how to describe them precisely and in a sufficient detail. However, from the standpoint of further research we consider even a mere separation of such constructions in the valency frames useful.

4.7. Valency of the verbs of foreign origin and their Czech counterparts

Valency frames of verbs of foreign origin are created having their "Czech" synonyms (if they exist) in mind. Thus in most cases the valency frame of the "foreign version" of a verb and its Czech counterpart is thus the same in most cases.

For instance:
vystěhovat se *z venkova do města.* valency frame: DIR1()
(lit.: *[to] move from the-countryside to the-city*)
emigrovat *z východu na západ* valency frame: DIR1()
(lit.: *[to] emigrate from the-East to the-West*)

zacházet *s penězi* valency frame: PAT(s+7)
(lit.: *[to]deal(handle) with money*)
disponovat *se zásobami* valency frame: PAT(s+7)
(lit.: *[to] deal(control,handle) with the-reserves*)
manipulovat *s mříží* valency frame: PAT(s+7)
(lit.: *[to] manipulate with the-grid*)

uvažovat *o životě* valency frame: PAT(o+6)
(lit.: *[to] think about life*)
meditovat *o zvycích* valency frame: PAT(o+6)
(lit.: *[to] meditate about traditions*)

Other examples:
dislokovat/umístit
(lit.: *to dislocate/to lie down*)
deportovat/vyhostit
(lit.: *to deport/ to banish***)**
demontovat/rozebrat
(lit.: *to dismantle/to strip down*)
devalvovat/znehodnotit
(lit.: *to devaluate/to invalidate***)**

absolvovat/zakončit
(lit.: *to go through(pass)/to finish*)

4.8. One or two frames?

Verbs with seemingly optional Patient form another uncertainty class. Here, it is often unclear whether one verb has two meanings (and thus should be split into two different frames).[14] This problem concerns verbs such as:

a) ***podnikat, plavat, běhat*** *(to undertake, to swim, to run)*
b) ***kousat, kouřit, kojit, zavěsit*** *(to bite, to smoke, to nurse, to hang up)*
c) ***tančit, cvičit, trénovat*** (*to dance, to exercise, to practice*)
d) ***mluvit, hovořit, číst, psát, zpívat*** *(to speak, to talk, to read, to write, to go, to sing).*

If we decide to use two different frames (cf. the group (a)), then the first frame does not include a Patient slot and the second one does (an obligatory one, of course). Otherwise, we stick with just one frame with either an obligatory Patient (cf. the group (b)) or an optional Patient (cf. the group (c)) or without a Patient slot altogether (cf. group (d)).

The reason for treating these four groups differently is that they behave differently. The meaning of "doing or running an activity (without a specific object in mind)" has the right to have its own valency frame in case of the group (a). These verbs get thus two frames. The first one is simply a single-slot frame ACT(.1) - e.g., *Kamarád už dlouho podniká* (lit.: *a-friend already for-a-long-time has-a-business*), *Anna plave závodně* (lit.: *Ann swims professionally*). The second one is a two-slot frame ACT(.1) PAT(.4) - e.g., *Plaval dvacet bazénů denně* (lit.: *he-had-swam twenty pools daily*), *Jirka podniká velké cesty* (lit.: *Jirka undertakes big journeys*). Notice that also the translation of *podniká* to English is different in these two cases, a strong indication of two different meanings.

On the contrary, the obligatory valency complementation is necessary for verbs in the (b) group. These verbs correlate very strongly with a specific Patient, therefore we consider a Patient here to be always present. The meaning of "doing or running an activity (without specific object in mind)" is just a submeaning of this valency frame. These verbs have thus the following two-slot frame: ACT(.1) PAT(.4).

In the (c) group, we decided to assign only one frame with an optional Patient: ACT(.1) ?PAT(.4)). E.g., in the clause *Jirka denně cvičí a trénuje* (lit.: *Jirka daily exercises and practices*) the verbs *to exercise* and *to practice* have a special "abstract" semantic characteristic „doing or running an activity", where it is unimportant <u>*what*</u> exactly he is exercising or practicing. On the other hand, we can certainly express some particular activity (which is always going on "behind the scenes") that Jirka can exercise or practice. Then, such activity would then be assigned the Patient functor.[15]

We believe that the last group of verbs, (d), has an additional semantics of the verb *umět* (lit.: *to know*) in one of its meanings, in sentences like *Pavel už mluví, ale ještě nečte, nepíše a nepočítá* (lit.: *Paul already talks, but [he] yet [does] not-read, not-write and not-count*), *Anna mluví hezky německy a už i zpívá* (lit.: *Ann speaks well German and already even sings*). These verbs have been assigned a valency frame without a Patient in this meaning (i.e., ACT(.1) only).

[14] Such a verb can have even more senses, which can be quite clear, e.g. *komín kouří* (lit: *the chimney fumes*).
[15] One might argue that because some people are professional trainers, the verb *to train* should have been moved to the group (a). We have not encountered such an example in the data, thus we assigned it to the group (b), but of course its reassignment cannot be excluded in the future.

4.9. Reciprocity in valency

The notion of reciprocity belongs to events where a two-way "direction" between its two participants can be observed, either happening simultaneously (*they met*) or mutually ("reciprocally": *they were helping each other*). It is well known (Panevová, 2003) that reciprocity changes the surface realization of the verbal valency structure in a way non-reciprocal events do not.

In most cases, it is ACT and PAT that are in the relation of reciprocity. For example, ACT(.1) PAT(.4) is the valency frame for the verb *líbat (to kiss)* but the Patient is not expressed in such reciprocal clauses as *Sourozenci se líbali* (lit.: *Siblings each-other kissed*). We could have used the *siblings* here also as a Patient (by creating another node in the tectogrammatical representation and duplicating the *siblings* there), but that would make it indistinguishable from "*Siblings kissed siblings*" which means something different. This type of "missing" valency member thus gave us a reason for setting up a new value of a tectogrammatical lemma, namely "Rcp": a new node with this value recorded in the tectogrammatical lemma attribute and a functor label that corresponds to the "missing" dependent is added if the usage of the verb is reciprocal and the "second" member is reciprocally "included" in the "first" member which is expressed by plural or as a coordination. Whereas the above clause is an example of the former expression (plural), *Honza a Marie se líbají* (lit.: *John and Mary [each-other] are-kissing*) is an example of the latter (coordination).

The "other" reciprocal participant is often seemingly expressed on the surface by forms of the morpheme (particle) "*se*" or "*si*" (Lit.: *him/herself, Acc. or Dat*, or *each other*.), sometimes in conjunction with the preposition "*s*" (*with*) used with the Patient; e.g., *Honza se líbá s Marií* (lit.: *John himself is-kissing with Mary*), *Tom s Pavlem si vyměňují známky* (lit.: *Tom with Paul themselves exchange(Pl.) stamps*). Since in these cases the Patient is expressed on the surface, the valency frames for these verbs have to account for it. The realization of this frame must thus contain also the preposition "*s*" with instrumental: ACT(.1) PAT(.4;s+7). Naturally, the particle "*se*" or "*si*" is discarded in the annotation in all cases, since the reciprocal element can be accounted for by either the expressed participant or by the Rcp node. Other verbs with this kind of frame are e.g. *potkat* (*to meet*), *vítat* and *přivítat* (*to welcome*), etc. Other details about reciprocity (such as the surface realization possibilities of expression) are not marked in the verbal valency frame nor in the annotated data because they can be handled by global "grammatical" rules. It is also true that almost every verb can be used reciprocally, at least in theory. Possible restrictions (if any) must still be studied in the future.

4.10. Specific valency frames: DPHR and CPHR

Some verbs can have, on top of regular valency frames, also specific frames when used in idiomatic expressions. The verb being assigned such a frame must be a part of an idiomatic construction. One of the frame slots is then labelled by a special functor DPHR (dependent part of a phraseme). The issue of idioms is very complicated and thus it is not easy to find the borderline between metaphorical and non-metaphorical meanings. We use the following principle: if the verb is used in an abstract meaning and has a metaphorical meaning in the given collocation, we mark the remaining part of the collocation as DPHR. Often, the surface realization of such member is complex and the full power of the formal system of describing surface realizations must be used; e.g., the idiom "*dát něčemu zelenou*" (lit.: *[to] give to-something a-green-light*) has the following valency frame: ACT(.1) PAT(.3) DPHR(zelený.FS4).

If a verb is semantically "emptied" (i.e., the semantic content of the verb is reduced or generalized) and it also meets some other requirements (see Cinková, S., Kolářová, V. this volume), we mark one of its valency members as CPHR (compound phraseme), namely the one that gives the collocation the "real" meaning. The specificity of the verbal valency frames with the functor CPHR (called ***support verb constructions***) consists, i.e., in that they do not undergo the process of shifting. For instance: *dostat od otce*.ORIG *příkaz*.CPHR (lit.: *[to] get from [his] father an-order*), *věnovat problému*.ADDR *pozornost*.CPHR (lit.: *[to] pay the-problem attention*).

5. Conclusion

The annotation of the verbal valency on the background of the Prague Dependency Treebank is a valuable contribution to Czech linguistics especially because a large list of verbs (more than 5300 verbs with 8200 valency frames) has been built on the basis of a corpus which has allowed to verify and refine the notion of valency as a substantial part of the Functional Generative Description theory. It was not necessary to make up valency complementation examples in order to fill out the theoretical schemes of valency frames because they were taken from real data. The results that have been achieved are considered to be the first step in this respect providing rich material for further linguistic and computational research. The annotation revealed a number of questions which we have tried to solve. However, many of them stay open to further research and discussion. These open questions as well as the fact that some of the first decisions during the annotation of verbal valency were not correct and also the fact that some rules of the annotation were changed during the annotation can be considered as a positive rather than a negative result of our research. The PDT-VALLEX, which is actually a byproduct of the annotation, is an important source for further linguistic research as well as computational processing of the Czech language. We also hope that it will be a useful source for many different applications and further studies.

Acknowledgement:

The research reported in this paper was supported by the grant of the Czech Ministry of Education LN00A063.

References:

Bémová, A., Buráňová, E., Hajič, J., Kárník, J., Pajas, P., Panevová, J., Štěpánek, J. and Urešová, Z. (1997): *Anotace na analytické rovině: návod pro anotátory.* Technical Report ÚFAL TR-1997-03. Charles University. Prague

Benešová, V. (2002): Delimitace lexií českých sloves z hlediska jejich syntaktických vlastností, diplomová práce, FF UK, Praha

Cinková, S., Kolářová, V. (2004): Nouns as Components of Support Verb Constructions in the Prague Dependency Treebank. This Volume.

Hajič J. (2004): The Prague Dependency Treebank, version 2.0. This volume.

Hajič J., Pajas P. (2004): Zápis valenčních rámců v PDT a jeho sémantika. http://ufal.mff.cuni.cz/pdt.

Hajič, J., Panevová, J., Urešová, Z., Bémová, A., Kolářová, V. (2003): PDT-VALLEX: Creating a Large-coverage Valency Lexicon for Treebank Annotation**.** In *Proceedings of The Second Workshop on Treebanks and Linguistic Theories*, pp. 57-68. Vaxjo University Press.

Hajič J., Urešová Z. (2003): Linguistic Annotation: from Links to Cross-Layer Lexicons**.** In *Proceedings of The Second Workshop on Treebanks and Linguistic Theories*, pp. 69-80. Vaxjo University Press.

Hajičová, E. et al. (in prep.): The Guidelines for Tectogrammatical Annotation. Praha 2004. Part of CD-ROM: Prague Dependency Treebank v. 2.0, LDC, Univ. of Pennsylvania.

Hajičová E., Panevová J., Sgall (2002): P. K nové úrovni anotovaného korpusu, část 1, Slovo a Slovesnost, 63, 161-177.

Kingsbury, P., Palmer, M. (2002): The Proposition Bank: An annotated Corpus of Semantic Roles.

Levin, B. (1993): English Verb Classes and Alternations. Chicago: University of Chicago Press, pp. 348

Lopatková, M., Panevová, J. (2004): Recent development of the theory of valency in the light of the Prague Dependency Treebank. This Volume.

Lopatková, M., Žabokrtský, Z., Skwarska, K., Benešová, V. (2003): Valency Lexicon of Czech Verbs VALLEX 1.0. CKL/UFAL Technical Report TR-2003-18, 2003

Panevová, J. (1974-75): On Verbal Frames in Functional generative Descrption. Part I, The Prague Bulletin of Mathematical Linguistics 22, pp.3-40, Part II, The Prague Bulletin of Mathematical Linguistics 23, pp. 17-52

Panevová, J. (1980): Formy a funkce ve stavbě české věty. Academia, Praha

Panevová, J. (1994): Valency Frames and the Meaning of the Sentence. In: The Prague School of Structural and Functional Linguistics (ed. Ph. L. Luelsdorff), Amsterdam-Philadelphia, John Benjamins, pp.223-243

Panevová, J. (1999): Valence a její univerzální a specifické projevy. In: Hladká Z, Karlík P, Čeština – Univerzália a specifika. Brno, 29-37

Panevová, J. (2003): Some Issues of Syntax and Semantics of Verbal Modifications. In: Proceedings MTT 2003, First International Conference on Meaning – Text Theory. Paris, Ecole Normale Supérieuere, 139-146

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

Sgall, P., Hajičová, E., Panevová, J. (1986): The Meaning of the Sentence in Its Semantic and Pragmatic Aspects (ed. by J. Mey), Dordrecht: Reidel and Prague: Academia.

Straňáková-Lopatková, M., Žabokrtský, Z. (2002): *Valency Dictionary of Czech Verbs: Complex Tectogrammatical Annotation.* In: LREC2002, Proceedings, vol.III. (eds. M. González Rodríguez, C. Paz Suárez Araujo), ELRA, pp. 949-956.

Řezníčková, V. (2003): Czech Deverbal Nouns: Issues of Their Valency in Linear and Dependency Corpora. In: Proceedings of the Workshop on Shallow Processing of Large Corpora (SProLAC 2003), pp. 88-97. UCREL, Lancaster University.