Czech deverbal nouns: issues of their valency in linear and dependency corpora

Veronika Řezníčková
Center for Computational Linguistics
Charles University, Prague
rez@ufal.mff.cuni.cz

1 Introduction

For Czech language, corpora of two types exist: (i) a databank of linear texts, i.e. the Czech National Corpus (CNC in the sequel) at the Faculty of Arts, Charles University, Prague (this is a representative corpus of contemporary written Czech, a part of which, called SYN2000, contains about 100 million wordforms), and (ii) a dependency based treebank, i.e. the Prague Dependency Treebank, PDT in the sequel, which is a part of CNC annotated at several layers. The shallow-parsed shape of the latter, so-called analytical level, contains approx. 90 000 sentences. Both these corpora are annotated (by morphological tags in CNC, by syntactic functions in PDT) and both of them are available by search procedures (the Graphic interface for the Corpus Query Processor, GCQP in the sequel, was developed for searching strings of lemmas, words or morphemic tags through the CNC, and the NetGraph serves for searching sentence structures in PDT). The linear corpus is very useful for searching morphemic and lexical phenomena, including information about their frequency, but the dependency treebank is irreplaceable whenever one investigates syntactic relations in the sentence. Due to the high degree of "free" word order in Czech, many modifications can occur as either preceding or following their governors (exceptions of course exist, such as e.g. the non-congruent attributes with nouns). Thus it is very difficult to formulate a query about syntactic relations in the linear corpus. Even a very simple query about congruent (aggreeing) attributes with nouns (e.g. ([tag="A.MS1.*"] [tag="N.MS1.*"])) can give an incorrect result (e.g. *Po bitvě je každý* (Subject) *generál* (Nominal part of the predicate), lit. 'After the battle, everyone is a general', i.e. 'Everyone is wise after the event'). It is surely more favourable to retrieve syntactical relations in the dependency treebank, but, comparing the size of both corpora, we see that it is still important to extract occurrences also from the linear one.

A verb occupies the central position in the sentence structure, so it is clear that one of the key syntactic relations is the valency of verbs, as well as the valency of deverbal nouns and of adjectives. In the present paper, we compare our results of verification of valency properties of Czech deverbal nouns in both mentioned corpora and we describe problems of the search. We evaluate just syntactic relations, not incorrect morphological tags.

2 Valency of deverbal nouns

Our approach to issues of a valency of deverbal nouns is based on the theory of valency (especially a valency of verbs) elaborated in the framework of the Functional Generative Description (FGD in the sequel, see Sgall, Hajičová and Panevová 1986; Panevová 1980).

In FGD, the valency frame of a verb, stored in the lexicon, can be described as present on the tectogrammatical level, which captures underlying syntactic structures of sentences. The following complementations (i.e. kinds of the dependency relation, or, more precisely, individual dependency relations) are included into the set as being able to fill individual slots of valency frames of verbs:

(i) inner participants or arguments (they can be obligatory or optional): Actor (ACT), Patient (PAT), Addressee (ADDR), Effect (EFF), Origin (ORIG);

(ii) obligatory free modifications or adjuncts (especially those with the meaning of location (e.g. DIR3, LOC) and manner (MANN)).

Most of these complementations can be omitted on the surface layer of the sentence, but some of them must be always present (as PAT with the verb *potkat* 'to meet', MANN with the verb *chovat se* 'to behave', etc.).

While describing valency frames of deverbal nouns, we use the same set of complementations as with verbs. However, in comparison with the frames assigned to the source verb, the process of nominalization (condensation) may be accompanied by a reduction of the number of slots in the valency frames of derived nouns at the underlying layer. Moreover, any complementation of a noun can be omitted on the surface layer. But we do not deal with the issues of deletions in this paper (for more details see Řezníčková 2002).

We can only present a tentative specification of reasons for a reduction of the number of slots in the valency frames of nouns at the underlying layer which may be understood as a key issue, relevant for a further analysis of the valency of nouns (see section 2.1). Another very important issue is the surface expression of valency slots of nouns (see section 2.2).

2.1 The process of nominalization

By deverbal nouns we understand:
i) Nouns derived from verbs
ia) by productive means (e.g. *rozhodnutí* 'decision making', *obžalování* 'accusing');
ib) by non-productive means or by the zero suffix (e.g. *rada* 'advise', *slib* 'promise');
ii) Nouns derived from a predicative adjective (e.g. *on je schopen udělat* 'he is able to do sth' → *jeho schopnost napsat knihu* 'his ability to write a book', *on je povinen udělat* 'he is obliged / required to do sth' → *jeho povinnost vydat majetek* 'his duty / obligation to release possession');
iii) Nouns which primarily are a part of an analytical predicate (e.g. *Petr má šanci vyhrát* 'Peter has a chance to win' → *Petrova šance vyhrát* 'Peter's chance to win', *Petr má právo odvolat se* 'Peter has a right to appeal' → *Petrovo právo odvolat se* 'Peter's right to appeal').

2.1.1 Types of the word-formation process

While describing valency frames of nouns derived from verbs, Panevová differentiates, according to J. Kuryłowicz, two basic types of the word-formative process: syntactic derivation and lexical derivation (cf. Panevová 2000). While in syntactic derivation only the syntactic function of derived word differs from that of the verb, in lexical derivation not only the syntactic function, but also the lexical meaning of the derived word differs (see Kuryłowicz 1936).
(a) Syntactic derivation can be exemplified by the following nouns derived from verbs:
Čtenář si ověřuje danou informaci v jiných zdrojích → Ověřování dané informace čtenářem v jiných zdrojích 'The reader verifies the given information in other sources → Verification of the given information by the reader in other sources';
Odborníci stavějí katedrálu → stavění katedrály odborníky / stavba katedrály odborníky 'Experts build the cathedral → building of the cathedral by experts';
(b) Nouns derived by lexical derivation can be illustrated e.g. by actor names (e.g. *učit → učitel* 'to teach → teacher'), by names denoting the place of action (e.g. *umývat → umývárna* 'to wash → washroom') and by names denoting a tool (e.g. *ořezávat → ořezávátko* 'to sharpen a pencil → a pencil sharpener').

2.1.2 Incorporation of participants or free modifications

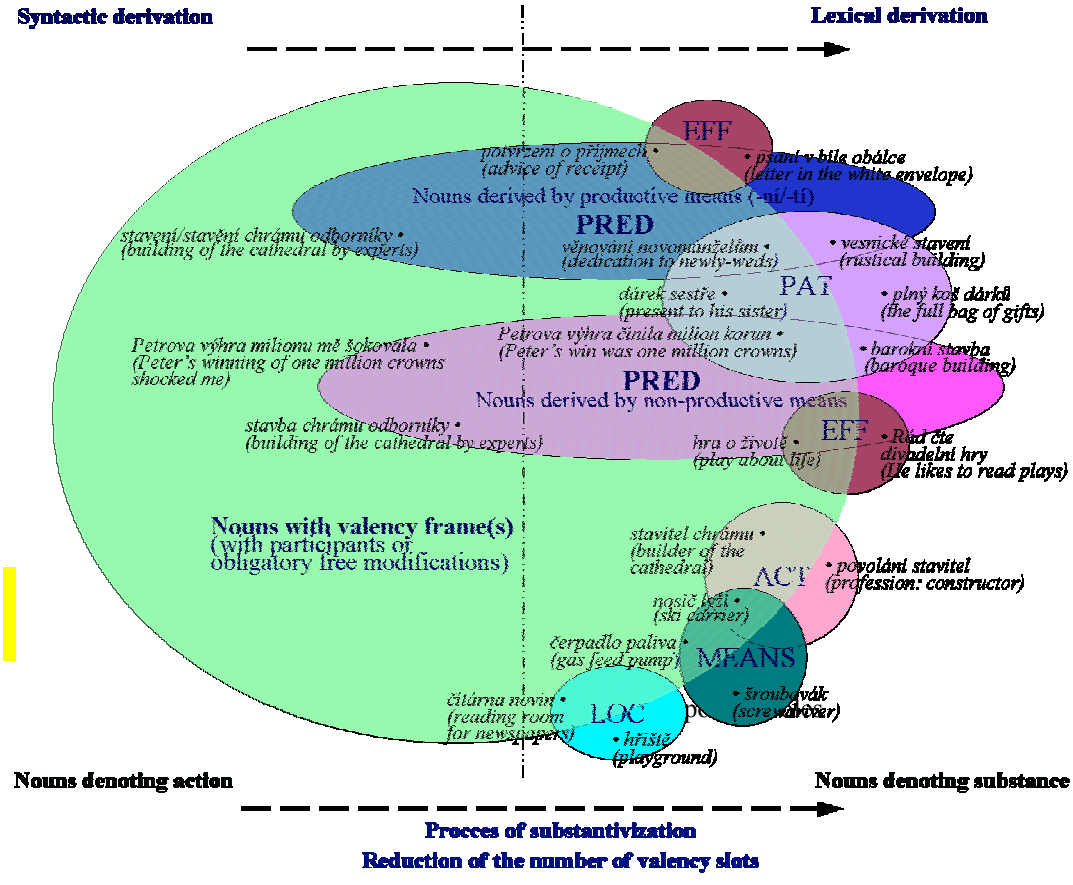
It is typical of nouns derived from verbs by lexical derivation that one of the inner participants or free modifications of the source verb is incorporated into the meaning of the noun itself, as the Actor with actor names, the place of action with names denoting the place and the tool with names denoting a tool (cf. Panevová 2000). When incorporation of inner participants is concerned (not only the incorporation of Actor, but also that of Patient (e.g. *dárek mé sestře*.ADDR 'the gift to my sister', *výrobek firmy*.ACT 'a product of the firm') or Effect (names of artefacts which can be understood as a result of an activity, e.g. *hra o životě*.PAT 'the play about life'), the meaning of the noun itself occupies one of the valency slots of the source verb.

2.1.3 The process of substantivization

The incorporation of participants can be accompanied also by the process called substantivization. While a noun with an incorporated participant can keep other participants from the valency frame of the source verb (cf. the examples in section 2.1.2) and loses only the incorporated one, during the process of substantivization the noun loses gradually all its participants (the syntactic function of the dependent expressed by a possessive pronoun or adjective changes from Actor (ACT) to Appurtenance (APP)).
Thus we can differ several stages of the process of substantivization with a single noun:
1) noun derived from verb by syntactic derivation: noun denoting action (cf. examples (1), (3), and (5));
2) noun derived from verb by lexical derivation:
2a) noun with an incorporated participant (cf. examples (4), (6), and (7));
2b) noun after the process of substantivization (cf. examples (2), (4), and (7)).
(1) *Psaní dopisu.*PAT *Petrem.*ACT *trvalo asi 3 hodiny*
'Writing a letter.PAT by Peter.ACT took him about three hours'
(2) *Petrovo.*APP *psaní leželo na stole* 'Peter's.APP letter lay on the table'
(3) *Petrova.*ACT *výhra milionu.*PAT *šokovala celou rodinu*
'Winning of one million.PAT crowns by Peter.ACT shocked the whole family'
(4) *Petrova.*APP *výhra činila milion korun* 'Peter's.APP win was one million crowns'
(5) *Naše.*ACT *výplata dividend.*PAT *klientům.*ADDR
'Our.ACT payment of dividends.PAT to clients.ADDR'
(6) *Nemá na výplaty zaměstnancům.*ADDR
lit. '(He) has not for payment to employees.ADDR' 'He has no money for employee wages'
(7) *To je moje.*APP *první výplata!* 'This is my.APP first wage-packet!'.

These stages correspond to particular meanings of the lexeme and express the transition of this lexical unit from a noun denoting action to a noun denoting substance.

Figure 1 illustrates the scale of possible incorporated roles (including also those nouns which do not incorporate any special role, but hold the function of a predicate; we label them by the auxiliary symbol PRED as they function similarly as the original predicate), their location on the virtual axis "syntactic derivation → lexical derivation", and with several nouns also the stages of the process of substantivization.



2.1.4 Status of Czech nouns derived from verbs by non-productive means

While nouns derived from verbs by productive means mostly denote action, the status of non-productively derived nouns in the word-formative process is still unclear.

Karlík (Karlík 2000 p. 188; Karlík and Nübler 1998 p. 106) describes formal differences between nouns derived from verbs by productive means and those derived from verbs by non-productive means in Czech; it is important to see that two of them may influence their syntactic behaviour:
i) just productively derived nouns overtake the aspect of the source verb (e.g. *navrhování něčeho* 'proposing (of) sth', *navržení něčeho* 'proposing (of) sth', *návrh něčeho* 'proposal of sth');
ii) just productively derived nouns can overtake the reflexive element "se" (e.g. *koupání* vs. *koupání se* 'bathing', but *koupel* vs. **koupel se* 'bath').

Chomsky differentiates three types of nominalizations in English (Chomsky 1972 p. 60):
i) the gerundive nominals (e.g. John's refusing the offer);
ii) the derived nominals (e.g. John's refusal of the offer);
iii) the "mixed" forms (e.g. John's refusing of the offer).

The most striking differences between (i) and (ii) "have to do with the productivity of the process in question, the generality of the relation between the nominal and the associated proposition, and the internal structure of the nominal phrase" (Chomsky 1972 p. 15). According to his observations "it seems that the transformationalist hypothesis is correct for the gerundive nominals and the lexicalist hypothesis for the derived nominals and perhaps, though much less clearly so, for the mixed forms" (Chomsky 1972 p. 60).

Czech nouns derived from verbs by productive means might correspond to the gerundive nominals in English, at least as far as the type of derivation (transformation) is concerned, but when we take into account also the surface expression of valency slots, then also these nouns correspond to the English "mixed" forms. Czech nouns derived from verbs by non-productive means might correspond to the derived nominals in English. However, considering that the internal structure of the nominal phrase is in Czech also characteristic for nouns derived from verbs by productive means, we can see more similarities between the two mentioned types of nouns in Czech than in English. Thus, even the clear lexical character of the derivation of nouns derived from verbs by non-productive means deserves further discussion. These nouns obviously differ from those derived from verbs by productive means, they in a specific way modify the pure syntactic character of derivation which is typical of nouns derived from verbs by productive means; on the other hand, they certainly have a valency frame, and very often they hold all participants of their source verb; thus, we can assume that they at least in some cases denote action. Of course, they also can incorporate one of its participants (more frequently the PAT, see section 2.1.2), or represent the result of the process of substantivization, and then, even when no participant is incorporated, their valency structure may get "overloaded" and they lose some of its slots (cf. (8)).

(8) (*On*.ACT) *dokázal příteli*.ADDR *svou nevinu*.PAT → *jeho*.ACT *důkaz neviny*.PAT *?příteli*.ADDR
'He.ACT proved his innocence.PAT to (his) friend.ADDR → his.ACT proof of innocence.PAT ?to (his) friend.ADDR'.

Thus, we agree to the observations of Panevová, who presumes a "middle" position of these nouns on the virtual axis: nouns derived from verbs by syntactic derivation – nouns derived from verbs by lexical derivation (cf. Panevová 2000).
The specific valency behaviour of this type of nouns has stimulated us to reexamine their valency properties.

2.2 Surface forms of valency slots of deverbal nouns

The process of nominalization of verbal structure is accompanied by changes of the surface expression of particular slots of a valency frame of a derived noun in comparison with those assigned to the source verbs. The possible transitions are illustrated in figure 2 (cf. also Lopatková, Řezníčková and Žabokrtský 2002).

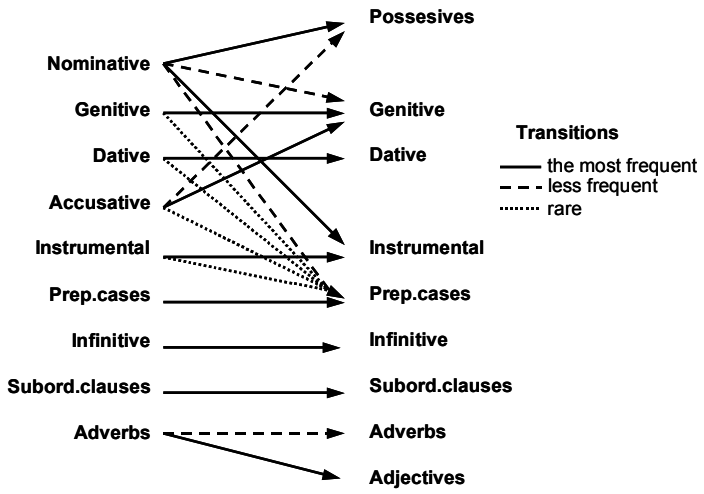


Figure 2: Possible changes of the surface expression

Nominalized structures serve for a condensed expression of the source verbal structures. Especially nominalized structures with more than two participants expressed by prepositionless cases (genitive, dative, instrumental) can even be held as too condensed, "overloaded". However, they are perfectly grammatical structures and they do occur (at least in theoretical texts, cf. Novotný 1980 p. 109), e.g.:

(9) *zdravotní sestra.*ACT *odebrala krev.*PAT *nemocnému.*ADDR → *odběr krve*.PAT *nemocnému*.ADDR *zdravotní sestrou*.ACT 'a nurse.ACT took blood.PAT from the patient.ADDR → taking of a blood-sample.PAT from the patient.ADDR by a nurse.ACT'.

The existence of Czech annotated corpora gives us the unique opportunity to check the usage and the frequency of such structures in Czech. As mentioned above, we focus on nouns derived from verbs by non-productive means.

3 Checking the valency properties of non-productively derived nouns with the use of corpora

Nouns derived from verbs with the valency frame ACT (Nom) PAT (Acc/(embedded clause)) ADDR (Dat) satisfy in the most adequate way the requirements for the surface expression of their participants stated in section 2.2. All three participants can be realized as prepositionless cases with these nouns, i.e. ACT (Instr/(Gen)), Pat (Gen), and ADDR (Dat).

3.1 Expressions of Patient and Addressee

In the present paper, we deal just with two obligatory participants of this type of deverbal nouns, namely with their Patient and Addressee. We have not verified the expression of Actor expressed by instrumental. Occurrences of Actor are searched for just in constructions where Actor is realized as genitive (cf. section 3.2.2).

3.1.1 Expression of Patient (Gen)

In the present section, we concentrate on the valency slot with the function of Patient realized as genitive. According to occurrences extracted from CNC, especially two phenomena may influence the expression of this valency slot:

(i) Incorporation of Patient

The expression of Patient is problematic with nouns that allow for an incorporation of this valency slot:

(10) *daroval knihu* → **dárek knihy* '(he) presented a book → present of a book'
(11) *dluží peníze* → **dluh peněz* '(he) owes money → debt of money'.
However, an expression of Patient with the noun *dluh* 'debt' is also possible, but only as an exception, e.g.: *dluh nájemného* 'debt of the rent'; also in its figurative meaning e.g.: *dluh lásky* 'debt of love', *dluh vděčnosti* 'debt of gratefulness'.

(ii) Patient with nouns derived from verbs of saying and of mental activity

Also an expression of Patient with nouns derived from verbs of saying and from verbs of mental activity is sometimes accompanied by difficulties, e.g.:
(12) *Petr mi slíbil knihu.* → **Slib knihy však nedodržel.*
'Petr promised me the book → However, he did not keep the promise of the book'
(13) *Poradil mi pěknou hloupost.* → **Rada pěkné hlouposti mě ale neuspokojila.*
'(He) advised me a stupid thing → However, this piece of advice of a stupid thing did not satisfy me'.

The problematic character of such an expression of Patient in genitive may be explained by the fact that with the two mentioned types of verbs this valency slot is expressed by an embedded objective clause in the source clauses, which do not exhibit nominalization, e.g.:

(14) *Petr mi slíbil, že mi přinese knihu* lit. 'Petr promised me to bring me the book'
(15) *Poradil mi, co mám číst* '(He) advised me what to read'.

We deal with two possible ways of condensation of Patient expressed by an embedded objective clause:

(a) the process of nominalization of the governing verb of the embedded clause, e.g.:

(16) *Poradil mi, co mám číst* → *Poradil mi četbu*
'(He) advised me what to read → (He) advised me the reading material';

(b) the process of a substitution of the embedded clause by the most important noun in that clause, e.g.:

(17) *Petr mi slíbil, že mi přinese knihu* → *Petr mi slíbil knihu*
lit. 'Petr promised me to bring me the book' → Petr promised me the book'.

We have found in CNC frequencies of occurrences of both mentioned types of slot condensation within valency structure of verbs of saying and verbs of mental activity, and then we compared them with frequencies of occurrences of nominalizations of those verbal structures (altogether 13 verbs and 17 deverbal nouns; eight of these nouns were derived from verbs by non-productive means). According to the results of our search (cf. Kučerová and Řezníčková, in press), the number of occurrences of Patient realized by accusative in verbal structures and the number of occurrences of Patient realized by genitive in nominalized verbal structures are similar, e.g.:

(18) *přikázat něco* 'to order sth' (e.g. *klid* 'rest', *revizi* 'inspection') – 10 occurrences → *příkaz něčeho* 'order of sth' (e.g. *mlčení* 'silence', *úhrady* 'compensation') – 15 occurrences;

(19) *slíbit něco* (e.g. *adresu* 'address', *podporu* 'support') – hundreds of occurrences → *slib něčeho* (e.g. *mlčenlivosti* 'secrecy', *hlasování* 'voting', *vodovodu* 'water main') – also hundreds of occurrences.

The difference between the number of occurrences of nouns with a condensed Patient probably corresponds to the position of particular deverbal nouns on the virtual axis: nominalizations of verbs of mental activity (e.g. *návrh* 'proposal', *nabídka* 'offer', *příslib* 'promise') - nominalizations of verbs of saying *(verba dicendi)* (e.g. *vzkaz* 'message', *rada* 'advice', *odpověď* 'answer'). Patient with nouns

derived from verbs of saying occurred just as an exception, namely in poetry, e.g. *odpověď touhy* 'answer of desire', *vzkaz lásky* 'message of love'. Moreover, these complementations can be understood also as an Actor.

3.1.2 Expression of Addressee (Dat)

The expression of the valency slot with the meaning of Addressee (expressed by a noun in dative) seems to be less problematic. We have extracted from CNC occurrences of this slot for 21 nouns derived from verbs by non-productive means (including also those derived from verbs of saying and from verbs of mental activity). Especially nouns with problematic expression of Patient occur with Addressee relatively often; we found dozens of such constructions with particular nouns.

The expression of Addressee is problematic with nouns that may be understood as the result of the process of substantivization, e.g. *důkaz* 'proof', cf. (8) above. Such collocations seem to be relatively rare. We have found just one example which concerns the noun *důkaz* 'proof' modified by an Addressee in CNC, cf. (20).

(20) *Důkazy slečně Slávce byly v tom, že...* 'Proofs to Miss Slávka consisted in the fact that...'.

3.1.3 Expression of both participants (PAT (Gen), and ADDR (Dat))

Especially nouns derived from verbs with Patient denoting a substance (i.e. originally not expressed by an embedded objective clause, cf. section 3.1.1) occurred with both their participants, e.g.:

(21) *krádež laptopu*.PAT *pracovníkovi*.ADDR *rozvědky* 'theft of the laptop to the member of the scouting patrol'
(22) *dodávky zbraní*.PAT *válčícím zením*.ADDR 'deliveries of weapons to warring countries'
(23) *výdej majetku*.PAT *restituentům*.ADDR 'issue of a property to restituents'.

We obtained dozens of such constructions with particular nouns.

3.2 Problems of the search in the linear corpus (CNC)

While searching through CNC, we used the Graphic interface for the Corpus Query Processor, GCQP in the sequel (cf. Kocek, Kopřivová and Kučera 2000). As has followed from the previous section, we have examined valency properties of individual nouns; also the searched complementation was in a concrete shape (a noun group without a preposition). In contrast to the valency of verbs, most complementations of nouns occupy a position to the right of the given noun, so it is easier to formulate a query about the valency of nouns than that about the valency of verbs in GCQP. Problems we faced during the search are summed up in the following sections.

3.2.1 "Competition" of complementations of two valency structures

The most frequent problem of the search concerns a "competition" between the complementation of the given noun and the complementation of a verb, a deverbal noun or a deverbal adjective occupying a position somewhere before or after the given noun.

When we verify an expression of a participant occupying the position just next to the given noun (using for example the query ([lemma="návrh" 'proposal'] [tag="N...[23].*"])), the possibility of a disagreement between the intention of the query and the obtained result (i.e. a string inside angled brackets) is not so probable with Patient realized as genitive. We have not found any incorrect occurrence, although it obviously could occur, especially when a numeral follows; cf. an imaginary example:

(24) *Na jeho návrh poslanců moc nereagovalo*

lit. ‘To his proposal of-deputies a-lot did-not-respond’; ‘Not so many deputies responded to his proposal’.

When we check an expression of Addressee by dative occupying the position immediately to the right of the given noun, the possibility of problems of the search is very common.
Verbs with their own valency slot expressed by dative can occur before (as in (25)) or after the investigated noun (as in (26)).

(25) *zaklepal a předal <vzkaz muži>* ‘(he) knocked and handed the message to the man’
(26) *úředníci, kteří <dávku studentce> odepřeli* lit. ‘clerks who the benefit to a student refused’.

Adjectives with their own complementation expressed by dative occupy mostly the position after the investigated noun, e.g.:

(27) *Není Miloševičova <nabídka Moskvě> nesympatická?*
lit. ‘Is-not the Miloševič’s offer to Moscow unpleasant?’; ‘Moscow finds the offer of Miloševič unpleasant, doesn’t it?’.

If the head of the competing structures is a noun, it precedes the complementation, e.g.:

(28) *o vyřízení <vzkazu manželce> požádal*
lit. ‘for delivering of the message to the wife (he) asked’; ‘(he) asked for delivering of the message to the wife’.

To make possible also a retrieval of a searched participant modified by its congruent attributes or a retrieval of a given noun with both its participants, i.e. PAT as well as ADDR (cf. (22)), it is necessary to insert some optional positions into the query (e.g. [lemma="návrh" ‘proposal’] [] {0,4} [tag="N...3.*"]). Insertion of optional positions into the query increases, of course, the possibility of problems of the search. In addition to the problems mentioned above (possible also with inserted optional positions), especially the occurrence of a verb, a deverbal noun or a deverbal adjective within the query should be taken into account.

In (29) still another noun with its own valency slot expressed by genitive is present in the query:

(29) *upravený <návrh prodeje státního podílu> v IPB* ‘a modified proposal of sale of the state share in IPB’.

In a similar vein, examples (30), (31) and (32) illustrate problems of the search for a valency slot expressed by dative caused by occurrence of a verb, a deverbal noun and a deverbal adjective with its own valency slot realized as dative within the query.

(30) *má ten <dluh splácet každému cizímu člověku>*
lit. ‘(he) has-to the debt to-pay to every foreign man’ ‘(He) has to pay the debt to every foreign man’
(31) *<návrh konkrétních dotací konkrétním osobám>* ‘proposal of concrete price supports to concrete people’
(32) *objem sociálních <dávek určených dětem>* ‘the extent of social security benefit aimed at children’.

3.2.2 “Competition” of two inner participants within one valency structure

An Actor, which in a verbal structure appears as nominative, may under certain conditions have the surface form of genitive with nouns (cf. figure 2 in section 2.2). Thus, in CNC the query for nouns with Patient expressed by genitive can return also occurrences of nouns followed by an Actor, e.g.:

(33) *dodávky zbraní*.PAT ‘deliveries of weapons’ vs. *dodávky společnosti*.ACT ‘deliveries of a company’
(34) *návrh dotací*.PAT ‘proposal of a price support’ vs. *návrh úřadu*.ACT ‘proposal of an office’.

3.2.3 "Competition" of an inner participant and a free modification within one valency structure

Nouns that allow for an incorporation of Patient, may undergo substantivization and be followed by a free modification with the meaning of Appurtenance, realized also as genitive, e.g.:

(35) *výplaty dividend*.PAT 'payment of dividends' vs. *výplaty zaměstnanců*.APP lit. 'payment of employees'; 'employee wages'
(36) *lup pstruhů*.PAT 'theft of trouts' vs. *V lese je zakopán lup loupežníků*.APP lit. 'In the wood is buried the theft of bandits'; 'The plunder of bandits is buried in the wood'.

3.3 Problems of the search in the dependency corpus (PDT)

The Prague Dependency Treebank is a project of a complex linguistic annotation of sentences from the written Czech National Corpus (for a description of the annotation scheme of PDT see e.g. Hajič et al. 2001). Czech sentences are annotated on three layers: morphological, analytical, and tectogrammatical. While at the shallow-parsed layer of PDT, at the so-called analytical level (AL in the sequel), edges of the tree correspond to (surface) dependency relations, the so-called tectogrammatical layer (TL in the sequel) captures underlying syntactic structures of sentences. Also edge labels (so-called functors) describe underlying syntactic relations between a dependent node and its governing node at TL.

While searching through PDT, we used the search procedure called NetGraph (cf. Mírovský and Ondruška 2002). In the dependency treebank all valency complementations depend on its governing noun, so the most often problem of the search in CNC, i.e. "competition" of complementations of two valency structures (cf. section 3.2.1), should not occur at all in PDT. From this point of view, the success of the search in the dependency treebank is almost 100%. However, in comparison with the number of occurrences extracted from CNC, we of course obtained just a limited set of searched syntactic relations in the smaller PDT; we found dozens of correct occurrences of the noun *dodávka* 'delivery' with Addressee expressed by dative in CNC, but obtained just 5 occurrences of that structure in PDT.

All valency complementations of deverbal nouns are labelled by the analytical function Atr (Attribute)) at AL. Only in syntactically ambiguous structures (cf. (37)) these valency slots are labelled by combined analytical functions, e.g. AtrObj (i.e. Attribute/Object), to mark both possible syntactic relations (i.e. the noun *podnik* 'company' is labelled by the analytical function AtrObj as it can modify the noun *nabídka* 'offer' as well as the verb *zprostředkovat* 'to mediate').

(37) *Zprostředkuje rovněž nabídky dánských firem českým podnikům*
lit. 'Will-mediate also offers of Danish firms to Czech companies'.

Taking into account that underlying syntactic relations between a dependent node and its governing node are captured only at the tectogrammatical layer, it is obvious that cases of both types of slot "competition" within one valency structure (i.e. a "competition" of two inner participants, cf. section 3.2.2, and a "competition" of an inner participant and a free modification within one valency structure, cf. section 3.2.3) can be distinguished only at the underlying syntactic level. Thus, these cases can be regarded as problems of the search in the shallow-parsed shape of the dependency corpus.

4 Conclusion

In the paper we have presented results of our checking of valency properties of Czech deverbal nouns, especially those derived from verbs by non-productive means. By searching through two types of corpora built for Czech, i.e. the databank of linear texts (CNC) and the dependency treebank (PDT), we obtained altogether hundreds of correct occurrences of nominalized verbal structures. It is possible to find considerably more occurrences in CNC than in PDT, but the linear character of CNC causes various types of a disagreement between the intention of the query for the valency structure and the obtained result. Both corpora mentioned provide important information about valency properties of

nouns, usable for a build-up of the valency dictionary of nouns which is useful for further applications of the software tools.

Acknowledgement

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

References

Chomsky N 1972 Remarks on Nominalization. In *Studies on Semantics in Generative Grammar*. Hague, Mouton, pp 11-61.

Hajič J, Hajičová E, Pajas P, Panevová J, Sgall P, Vidová-Hladká B 2001 *Prague Dependency Treebank 1.0* (final production label). CDROM CAT: LDC2001T10., ISBN 1-58563-212-0.

Karlík P 2000 Valence substantiv v modifikované valenční teorii (Valency of nouns in a modified valency theory). In Hladká Z, Karlík P (eds), *Čeština – univerzália a specifika*. Brno, MU, pp181-192.

Karlík P, Nübler N 1998 Poznámky k nominalizaci v češtině (Notes on nominalization in Czech). *Slovo a slovesnost* 59: 105–112.

Kocek J, Kopřivová M, Kučera K (eds) 2000 *Český národní korpus. Úvod a příručka uživatele* (Czech National Corpus. Introduction and the handbook of a user). FF UK, Praha.

Kučerová I, Řezníčková V (in press) Korpus jako výzva k syntaktické analýze. Poznámky k syntaktické derivaci deverbativních substantiv v češtině (Corpus as a challenge to syntactic analysis. Notes on the syntactic derivation of deverbal nouns in Czech). *Slavia*.

Kuryłowicz J 1936 Dérivation lexicale et dérivation syntaxique. In *Bulletin de la Société linguistique de Paris*, 37, pp 79-92.

Lopatková M, Řezníčková V, Žabokrtský Z 2002 Valency Lexicon for Czech: From Verbs to Nouns. In *Proceedings of TSD*, Brno, Springer, pp 147-150.

Mírovský J, Ondruška R 2002 NetGraph System. Searching through the Prague Dependency Treebank. In: *The Prague Bulletin of Mathematical Linguistics*, 77, pp 101-104.

Novotný J 1980 *Valence dějových substantiv v češtině (Valency of deverbal nouns in Czech)*. Praha, SPN.

Panevová J 1980 *Formy a funkce ve stavbě české věty (Forms and Functions in the Sentence Structure of Czech)*. Praha, Academia.

Panevová J 2000 Poznámky k valenci podstatných jmen (Notes on the valency of nouns). In Hladká Z, Karlík P (eds), *Čeština – univerzália a specifika*. Brno, MU, pp 173-180.

Řezníčková V 2002 PDT: Two Steps in Tectogrammatical Annotation with respect to some Issues of Deletion. In *The Prague Bulletin of Mathematical Linguistics*, 78. Prague, Charles University, pp 37-52.

Sgall P, Hajičová E, Panevová J 1986 *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Dordrecht, Reidel, and Prague, Academia.