NON-SYSTEMIC VALENCY BEHAVIOR OF CZECH DEVERBAL NOUNS BASED ON THE NomVallex LEXICON

VERONIKA KOLÁŘOVÁ[1] – ANNA VERNEROVÁ[1] – JONATHAN VERNER[2]
[1]Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic
[2] Faculty of Arts, Charles University, Prague, Czech Republic

Abstract: In order to describe non-systemic valency behavior of Czech deverbal nouns, we present results of an automatic comparison of valency frames of interlinked noun and verbal lexical units included in valency lexicons NomVallex and VALLEX. We show that the non-systemic valency behavior of the nouns is mostly manifested by non-systemic forms of their actants, while changes in the number or type of adnominal actants are negligible as for their frequency. Non-systemic forms considerably contribute to a general increase in the number of forms in valency frames of nouns compared to the number of forms in valency frames of their base verbs. The non-systemic forms are more frequent in valency frames of non-productively derived nouns than in valency frames of productively derived ones.

Keywords: adnominal morphemic forms, Czech deverbal nouns, non-systemic valency behavior, valency, valency lexicon

1 INTRODUCTION

When describing valency behavior of Czech deverbal and deadjectival nouns, valency of their base verbs or adjectives can be taken into consideration in order to see whether the nouns' valency properties are derivable from their base words. If this is the case, they can be understood as a result of a regular process. When the valency properties of a noun are more or less independent of its base word, these can be considered to be a result of an irregular process. Applying such a comparative approach, systemic (typical) and non-systemic (special) valency behavior of nouns is distinguished.

Up to now, the distinction between systemic and non-systemic valency behavior has been intensively studied on the material of Czech deverbal nouns (Section 3), focusing on non-systemic forms of their valency complementations [4], e.g., *varovat koho*.Acc 'to warn sb' → *varování komu*.Dat 'warning to sb', i.e., *warning addressed to sb.* In this paper, we show how the non-systemic valency behavior is represented in the current version of the NomVallex lexicon (Section 2), drawing an automatic

comparison between valency frames of nouns included in NomVallex and their base verbs included in the VALLEX lexicon (Section 4). This comparison represents the first attempt to provide statistical data on the non-systemic valency behavior of Czech deverbal nouns.[1]

2   THE NOMVALLEX LEXICON

NomVallex is a valency lexicon of Czech deverbal nouns, created within the theoretical framework of the Functional Generative Description (FGD, [11]) and based on corpus data (*Czech National Corpus*, subcorpus *SYNv6* [8], and *Araneum Bohemicum Maximum* [1]).[2] Applying the valency theory of the FGD [10], valency properties of a noun lexical unit (LU) are captured in a valency frame which is modeled as a sequence of valency slots, supplemented with their morphemic forms. The following types of complementations may fill in the individual slots of valency frames of most deverbal nouns: obligatory or optional actants, i.e., Actor (ACT), Patient (PAT), Addressee (ADDR), Effect (EFF) and Origin (ORIG), e.g., *balení dárků*.PAT *rodiči*.ACT 'wrapping of the presents by parents', and obligatory free modifications, especially those with the meaning of direction, e.g., *chlapcův*.ACT *pozdní příchod do školy*.DIR3 'boy's late arrival to the school'. Nouns denoting quantity (a container) usually only have one valency slot in their valency frame, an actant called Material (MAT), which is in the form of prepositionless genitive, cf. *jedno balení léků*.MAT 'one package of medicine'.

Up to now, NomVallex has focused on deverbal nouns belonging to three semantic classes, i.e. Communication (e.g. *dotaz* 'question'), Mental Action (e.g. *plán* 'plan') and Psychological Noun (e.g. *nenávist* 'hatred'), see [7]. The lexicon captures all lexical meanings of the nouns, differentiating also basic "notional" meanings, i.e. action (e.g., *žádání* 'asking', *dovtípení se* 'inferring'), abstract result of action (e.g., *žádost* 'request'), quality (e.g., *důvtip* 'ingenuity'), substance (e.g., *komunikace (silnice)* 'road'), and quantity (a container, e.g., *soubor* 'collection'). Currently, it contains more than 400 noun lexical units.[3]

NomVallex relates to VALLEX [9], created within the same theoretical framework. NomVallex adopts VALLEX annotation scheme and in relevant cases it also splits the lexems into lexical units and assigns them to the relevant semantic classes according to the base verbal lexical units captured in VALLEX. As both lexicons are available as machine readable data, an automatic comparison of any valency characteristics annotated in the lexicons is possible. The links between the pairs of corresponding verbal and noun lexical units are recorded in the noun

---

[1] Although there are two other valency lexicons containing Czech deverbal nouns ([12] and [2]), none of them links information on valency of the nouns to their base verbs.

[2] The aim of the lexicon is to cover also other nominals such as adjectives and deadjectival nouns.

[3] `https://logic.ff.cuni.cz/nomvallex-beta/`

entries, by indicating the verb's identification code in the attribute *derivedV*, cf. (1) and (2).

(1) *žádat* 'to ask'
   - id: blu-v-žádat-2
   ACT(Nom) ADDR(Acc) PAT(*o*+Acc,inf,*aby,ať,zda,že*)

(2) *žádost* 'request'
   - derivedV: blu-v-žádat-2
   ACT(Gen,poss,*od*+Gen) ADDR(Dat,*k*+Dat)   PAT(Gen,*k*+Dat,*na*+Acc,*o*+Acc, *po*+Loc,inf,*aby,ať,zda,že*)

NomVallex covers both types of Czech deverbal nouns that can denote action or an abstract result of action, namely:

i.   nouns derived from verbs by productive suffixes *-ní/-tí*, e.g., *dotazování* 'asking''*, namítnutí* 'objecting', called productively derived nouns;

ii. nouns derived from verbs by non-productive suffixes, such as *-ba, -a, -ka*, e.g., *námitka* 'objection', or by the zero suffix, e.g., *dotaz* 'question'; these nouns are called non-productively derived nouns.[4]

In order to be able to compare valency behavior of these two types of nouns, NomVallex aims at creating a lexicon entry for both the productively and the non-productively derived nouns derived from one base verb, e.g., *žádání* 'asking' as well as *žádost* 'request' derived by different suffixes from the verb *žádat* 'to ask'.

3   SYSTEMIC AND NON-SYSTEMIC VALENCY BEHAVIOR OF CZECH DEVERBAL NOUNS

In this section, we specify how the systemic and non-systemic valency behavior of Czech deverbal nouns is manifested.

3.1   Systemic valency behavior

The valency behavior referred to as systemic can be observed especially with Czech productively derived nouns*,* if they denote action, cf. *kontrolování* 'checking' in (4). Non-productively derived nouns manifest systemic valency behavior less frequently [3], cf. *kontrola* 'check' in (4).

(3) *policista.*Nom *kontroluje vozidla.*Acc
   'a policeman is checking vehicles'

(4) *kontrolování / kontrola vozidel.*Gen *policistou.*Ins
   'checking / check of vehicles by a policeman'

---

[4] The term „non-productive" reflects esp. the fact that not all verbs have an counterpart in nouns derived by these suffixes, cf. *přemlouvat* 'to persuade' – *přemlouvání* 'persuading' – **přemluva* 'persuasion'.

When determining their valency frames, the nouns are expected to inherit all participants that are present in the valency frame of their base verbal lexical unit, including the "verbal" character of the participants such as Actor, Patient or Addressee.

Forms of adnominal participants change in case the base verbal form is either Nom or prepositionless Acc, cf. (3) and (4), or – less frequently – if a noun or an adjective in Nom or Acc are a part of an expression containing the word *jako* 'as', see Table 1. We also consider a possessive form corresponding to verbal prepositionless Gen to be a systemic change, cf. (5).

(5) *zanechat studia*.Gen 'to quit the school'→
   jeho.poss *zanechání* 'its quitting'

Verbal form	Adnominal systemic form
Nom	Gen, Ins, poss, *od* 'from'+Gen
Gen	poss
Acc	Gen, poss
jako 'as' + Nom	*jako* 'as' + Gen
jako 'as' + Acc	*jako* 'as' + Gen
jako 'as' + adj-Acc	*jako* 'as' + adj-Acc

Tab. 1. Systemic changes

All forms which do not change their form are regarded to be systemic. These include prepositionless cases Gen, Dat and Ins, an infinitive, prepositional groups (PGs, e.g., *k* 'to'+Dat), conjunctions (e.g. *že* 'that'), content clauses, an adjective in prepositionless Ins, expressions containing the word *jako* 'as' (*jako*+Gen, *jako*+adj-Gen, *jako*+PG), and expressions containing preposition *za* 'as/for' plus an adjective in prepositionless Acc (*za*+adj-Acc).

3.2 Non-systemic valency behavior

Non-systemic valency behavior of deverbal nouns is most often and most distinctly manifested by changes in properties of its valency complementations [4]. They involve three phenomena:

i. non-systemic forms of valency complementations (e.g., Gen → Dat, *otázat se kolegy*.Gen 'to ask a colleague' → *otázka kolegovi*.Dat 'a question to-a-colleague'; Sections 4.1 and 4.2);[5]

ii. a change (esp. a reduction) of the number of slots in the valency frame of a noun (e.g., the noun *velení* in *vrchní velení* 'the supreme headquarters' denotes a group of people rather than a process of commanding, as in *jeho*.ACT *velení armádě*.PAT 'his commanding the army', and thus loses ACT from its valency frame; Section 4.3);

---

[5] Various factors contributing to usage of non-systemic forms, including an influence of a form of a valency complementation of a light verb in light verb constructions (e.g. *dát otázku kolegovi* 'to address a question to a colleague'), are discussed in [3].

iii. a change of the character of a valency complementation to exclusively nominal, e.g., Material modifying nouns denoting quantity, as in *jedno balení léků*.MAT 'one package of medicine', in contrast to Patient in *balení kufrů*.PAT 'packing of bags', denoting action. This case is however extremely rare in the NomVallex data and is not dealt with in the paper.

We assume the notional meaning of a deverbal noun that displays non-systemic valency behavior is always different from action, and thus the noun denotes an abstract result of action, quality, substance or quantity.

4 AN AUTOMATIC COMPARISON OF VERBAL AND NOUN VALENCY FRAMES

Our automatic comparison of valency frames of nouns in NomVallex and valency frames of their base verbs in VALLEX obviously only covers nouns that provide a link to their base verbal lexical unit in VALLEX. First, an automatic procedure checks whether the valency frame of the given noun lexical unit corresponds to systemic valency behavior (i.e., if the number and type of valency slots is the same as in the corresponding verbal valency frame, and if their forms are either the same or correspond to a systemic change, see Section 3.1). Second, any difference from the systemic valency behavior is indicated as non-systemic one and is captured in the noun entry in the attribute *framediff* (difference in frame).

In this Section, we only focus on differences in the number or forms of actants, leaving out free valency modifications. Comparing verb-noun pairs with equal actants, we provide the general statistics on the number of morphemic forms in noun valency frames (Section 4.1), and we present distribution of non-systemic adnominal forms across the NomVallex data (Section 4.2). A difference in the number of actants is in focus of Section 4.3.

4.1 An increase in the number of adnominal forms

Noun valency structures show various limitations compared with verbal ones:

i. Adnominal prepositionless Gen and possessive forms may be syntactically ambiguous, being a result of different systemic changes or even some non-systemic ones (cf. Table 1 and Table 4).

ii. Noun valency patterns are subject to certain restrictions on combinations of actants expressed by particular morphemic forms, e.g., double postnominal genitives [5], including their word order, e.g., all incongruent attributes come after the noun, prepositionless Gen comes the first, then come the other forms [13].

iii. Morphemic forms of particular actants modifying nouns denoting an abstract result of action are rather often non-systemic, e.g., *návrh na reformy* 'a proposal for reforms', see [4]. However, the adnominal actants often keep the systemic forms as well, e.g., *návrh reforem* 'a proposal of reforms'.

As a result, deverbal nouns show a strong tendency to have at their disposal more morphemic forms of their actants than their base verbs, cf. (1) and (2), which enables them to use the appropriate form depending on the syntactic structure they occur in or depending on their notional meaning.

A general statistics on the number of morphemic forms in verbal and corresponding noun valency frames is given in Table 2, showing an apparent increase in the number of adnominal forms. The verbal lexical units which correspond to several noun lexical units (as in *žádat* 'to ask' – *žádání* 'asking', *žádat* 'to ask' – *žádost* 'request') are figured in the statistics as many times as many links to noun lexical units they have. The opposite case, nouns with more than one link to a verbal lexical unit, is not included in the statistical data given in Table 2.

On average, the total number of adnominal forms is more than 43% higher than the total number of the verbal forms. Counting the number of forms per a valency frame, noun valency frames contain on average 2.2 more forms than valency frames of the corresponding verbal lexical units. The most considerable increase in number of adnominal forms can be seen in valency frames of non-productively derived nouns of Communication (the total number of adnominal forms is more than 53% higher than the total number of the forms of base verbs of Communication, which brings on average 3.3 more forms in noun valency frames).

Class	Noun's suffix	Verb-noun pairs	Base verb's forms		Adnominal forms			
			Total	Number of forms per LU	Number of forms		Number of forms per LU	
					Total	Increase of %, comp. to verbs	Total	Increase, comp. to verbal LUs
Communication	prod.	71	446	6.3	569	27.6	8.0	1.7
	non-prod.	34	207	6.1	318	53.6	9.4	3.3
Mental Action	prod.	72	330	4.6	501	51.8	7.0	2.4
	non-prod.	34	151	4.4	229	51.7	6.7	2.3
Psych. Verb / Noun	prod.	32	124	3.9	190	53.2	5.9	2
	non-prod.	13	60	4.6	87	45.0	6.7	2.1
Total	prod.	175	900	5.1	1260	40.0	7.2	2.1
	non-prod.	81	418	5.2	634	51.7	7.8	2.6
	both / average	256	1318	5.2	1894	43.7	7.4	2.2

Tab. 2. An increase in the number of adnominal forms

The same data is used in Table 3 in order to pinpoint the distribution of systemic and non-systemic forms in valency frames of nouns. Looking at the average numbers of all systemic and non-systemic forms, we can see that non-systemic forms account for 15% of the total number of adnominal forms. However, taking into consideration

whether the nouns are productively or non-productively derived, the non-systemic forms account for 24.9% of all adnominal forms in valency frames of non-productively derived nouns, while non-systemic forms in the valency frames of productively derived nouns only account for 10.1%. The most significant difference in number of non-systemic forms can be seen in valency frames of nouns of Communication; while the percentage of non-systemic forms in valency frames of non-productively derived nouns is 25.2%, these forms in valency frames of productively derived nouns only account for 5.3%. The statistical data given in Table 3 confirms results of previous manual analysis carried out on corpus data [3], showing clearly that non-productively derived nouns tend to use non-systemic forms to a higher extent than productively derived nouns. At the same time, it follows from the data that in some cases also productively derived nouns use non-systemic forms and so their valency behavior cannot be considered to be purely systemic either.

Class	Noun's suffix	Verb-noun pairs	Base verb's forms	Adnominal forms				Total / 100%
				Systemic		Non-systemic		
					%		%	
Communication	prod.	71	446	539	94.7	30	5.3	569
	non-prod.	34	207	238	74.8	80	25.2	318
Mental Action	prod.	72	330	441	88.1	60	11.9	501
	non-prod.	34	151	177	77.3	52	22.7	229
Psych. Verb / Noun	prod.	32	124	153	80.5	37	19.5	190
	non-prod.	13	60	61	70.1	26	29.9	87
Total	prod.	175	900	1133	89.9	127	10.1	1260
	non-prod.	81	418	476	75.1	158	24.9	634
	both	256	1318	1609	84.9	285	15.1	1894

Tab. 3. The number of systemic and non-systemic adnominal forms

4.2   A distribution of non-systemic forms of actants

Analyzing the adnominal forms in more detail, all non-systemic forms were classified by the actant they express and ordered according to their frequency in the lexicon data (see Table 4).[6] Our data shows that while ACT, EFF and ORIG are only exceptionally expressed by a non-systemic form, PAT and ADDR use these forms quite often, though ADDR only with nouns of Communication. Concerning PAT and ADDR, only PAT can be expressed by an infinitive or by a content clause, starting either with a conjunction (C) or without it (CONT). Regardless this difference, we can see that the most frequent non-systemic form of both PAT and ADDR is

---

[6] In Table 4, NA stands for *non applicable*, i.e. for the case when no such an actant is present in valency frames of nouns representing the particular semantic class, and the number 0 means that such an actant exists but there is no non-systemic form it is expressed by. The numbers after slash signs refer to the number of particular forms.

a prepositional group, cf. (6) and [6]. The second most frequent non-systemic form of ADDR is prepositionless Dat, cf. (6). As for the second most frequent form of PAT, there is no clear tendency for the examined groups of nouns to use some common non-systemic forms; their valency behavior is rather idiosyncratic and should be studied case-by-case, considering their individual valency frames.

(6) *žádat obec*.ADDR(Acc) 'to ask the village' →
   žádost k obci.ADDR(k+Dat) 'request (addressed) to-the-village'
   žádost obci.ADDR(Dat) 'request (addressed) to-the-village'

Functor	Noun's suffix	Class		
		Communication	Mental Action	Psychological Noun
ACT	prod.	0	Gen/1; poss/1	0
	non-prod.	0	0	0
ADDR	prod.	PG/6, Dat/3, poss/2, Gen/1 The most frequent PGs: *pro*+Acc, *k*+Dat	0	NA
	non-prod.	PG/18, Dat/8, Gen/3, poss/3 The most frequent PGs: *k*+Dat, *pro*+Acc	0	NA
EFF	prod.	0	0	NA
	non-prod.	C/1	0	NA
ORIG	prod.	NA	INS/1	NA
	non-prod.	NA	0	NA
PAT	prod.	PG/17, CONT/2, inf/2, C/1 The most frequent PGs: *o*+Loc, *na*+Acc, *k*+Dat	PG/17, Gen/7, poss/4, CONT/3, Dat/3, inf/2, C/2 The most frequent PGs: *o*+Loc, *k*+Dat, *nad*+Ins	PG/22, C/2, Ins/2, poss/1, inf/1 The most frequent PGs: *z*+Gen, *nad*+Ins
	non-prod.	PG/34, C/15, Gen/7, CONT/7, inf/2, poss/1 The most frequent PGs: *na*+Acc, *k*+Dat, *proti*+Dat	PG/31, C/5, CONT/5, inf/1 The most frequent PGs: *o*+Loc, *k*+Dat, *nad*+Ins	PG/18, inf/4, C/3, CONT/1 The most frequent PGs: *z*+Gen, *před*+Ins, *nad*+Ins

Tab. 4. A distribution of non-systemic adnominal forms

4.3 A difference in the number of actants

Our automatic comparison of verb-noun pairs of valency frames also marks cases of a change in the number of actants in noun valency frames. Table 5 shows that these cases are rather rare. An addition of an actant is often just a result of a decision to annotate the valency frame of the noun in a different way, compared to the base verbal lexical unit in VALLEX, rather than a manifestation of non-systemic valency behavior of the particular noun lexical unit. However, besides the case of a different annotation, a deletion of an actant can indicate non-systemic valency behavior indeed (namely a change in the notional meaning of the noun, leading to loosing an actant).

Most frequently, the notional meaning of a noun changes from action to a substance (a person or a group of people as in *neschopné vrchní velení armády.*PAT 'an incompetent army's supreme command', loosing ACT from its valency frame, or a thing as in *jednosměrná komunikace* 'one-way road', loosing all actants of its base verb). A deletion of PAT, accompanying 'action → quality' change in the notional meaning of the noun, can be exemplified by the noun *důvtip* 'ingenuity', cf. the verbal construction in (7) and the nominal construction in (8), out of which the latter cannot be modified by PAT in any morphemic form.

(7) *generál.*ACT *se dovtípil něčeho/že.*PAT
    'a general has inferred sth/that'
(8) *důvtip generála.*ACT
    'the general's ingenuity'

Actant	No change	An actant added	An actant deleted
ACT	282	0	11
ADDR	95	9	12
EFF	33	3	3
ORIG	8	3	4
PAT	282	3	6

Tab. 5. Changes in the number of actants in noun valency frames

5   CONCLUSION

We have presented results of the first automatic comparison of valency frames of interlinked noun and verbal lexical units, included in valency lexicons NomVallex and VALLEX.

Our data shows that the non-systemic valency behavior of Czech deverbal nouns is mostly manifested by non-systemic forms of their actants, most frequently by a prepositional group. The non-systemic forms considerably contribute to a general increase in the number of forms in valency frames of nouns compared to

the number of forms in valency frames of their base verbs. In line with our expectations, the data shows that non-systemic forms are more frequent in valency frames of non-productively derived nouns than in valency frames of productively derived ones.

ACKNOWLEDGMENTS

R e f e r e n c e s

[1]   Benko, V. (2014). Aranea: Yet Another Family of (Comparable) Web Corpora. In Sojka, P. et al., editors, TSD 2014. LNAI 8655, pages 247–256. Springer International Publishing.

[2]   Hajič, J. et al. (2003). PDT-VALLEX: Creating a Largecoverage Valency Lexicon for Treebank Annotation. In Proceedings of The Second Workshop on Treebanks and Linguistic Theories, pages 57–68. Vaxjo University Press.

[3]   Kolářová, V. (2010). Valence deverbativních substantiv v češtině (na materiálu substantiv s dativní valencí). Praha, Karolinum.

[4]   Kolářová, V. (2014a). Special valency behavior of Czech deverbal nouns. In Spevak, O., editor, Noun Valency, pages 19–60, Amsterdam, John Benjamins Publishing Company.

[5]   Kolářová, V. (2014b). Nominalizované struktury se dvěma aktanty ve formě bezpředložkového genitivu. Naše řeč, 97(4–5), pages 286–299.

[6]   Kolářová, V., Vernerová, A., and Klímová, J. (2018). Předložková vyjádření adnominálních valenčních doplnění. Prace Filologiczne, 72, pages 211–223.

[7]   Kolářová, V., Vernerová, A., Klímová, J., and Kolář, J. (2017). Possible but not probable: A quantitative analysis of valency behaviour of Czech nouns in the Prague Dependency Treebank. Jazykovedný časopis, 68(2), pages 208–218.

[8]   Křen, M. et al. (2017). Korpus SYN, verze 6 z 18. 12. 2017. Praha, Ústav Českého národního korpusu FF UK. Accessible at: http://www.korpus.cz.

[9]   Lopatková, M., Kettnerová, V., Bejček, E., Vernerová, A., Žabokrtský, Z., and Barančíková, P. (2018). VALLEX 3.5 – Valenční slovník českých sloves. Praha, Karlova univerzita. Accessible at: http://ufal.mff.cuni.cz/vallex/3.5/.

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

[11] Sgall, P., Hajičová, E., and Panevová, J. (1986). The Meaning of the Sentence in its Semantic and Pragmatic Aspects. Reidel, Dordrecht.

[12] Svozilová, N., Prouzová, H., and Jirsová, A. (2005). Slovník slovesných, substantivních a adjektivních vazeb a spojení. Praha, Academia.

[13] Uhlířová, L. (2017). Slovosled nominální skupiny. In Nový encyklopedický slovník češtiny. Accessible at: https://www.czechency.org/slovnik.