

POSSIBLE BUT NOT PROBABLE:
A QUANTITATIVE ANALYSIS OF VALENCY BEHAVIOUR
OF CZECH NOUNS IN THE PRAGUE DEPENDENCY TREEBANK

VERONIKA KOLÁŘOVÁ¹ – ANNA VERNEROVÁ¹ – JANA KLÍMOVÁ¹ – JAN KOLÁŘ²

¹ Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic

² Institute of Mathematics, Czech Academy of Sciences, Prague, Czech Republic

KOLÁŘOVÁ, Veronika – VERNEROVÁ, Anna – KLÍMOVÁ, Jana – KOLÁŘ, Jan:
Possible but not Probable: A Quantitative Analysis of Valency Behaviour of Czech Nouns in
the Prague Dependency Treebank. *Journal of Linguistics*, 2017, Vol. 68, No 2, pp. 208–218.

Abstract: In order to optimize corpus searches for valency lexicon production, we analyse the relative frequencies of different combinations of valency complementations of Czech deverbal nouns in the Prague Dependency Treebank, considering differences between productively and non-productively derived nouns and their semantic class. We also classify combinations of forms of participants according to their frequency.

Keywords: valency, valency lexicon, Czech nouns, Word Sketch, corpus, Prague Dependency Treebank, quantitative analysis

1 INTRODUCTION

The practice of illustrating words in monolingual dictionaries with quotes goes at least as far back as Samuel Johnson's dictionary of English published in 1755. Until the advent of electronic corpora, quotes had to be collected and organised manually, mostly in the form of slips containing one quotation illustrating a selected keyword. Because of the scale of evidence that needed to be collected, volunteers from the public were involved in the major lexicographic projects such as the *Oxford English Dictionary* [22] (starting in 1858) or *Průruční slovník jazyka českého* [20] (starting after 1905). In the 1960's, first electronic corpora became available for linguistic research, and the first lexicographic project, relying fully on corpus data was started at the University of Birmingham in late 1970's, leading to the publication of the Collins COBUILD English Language Dictionary in 1987. Since then, reliance on corpus evidence is becoming the norm, especially in monolingual lexicography (COBUILD [4], Longman [14]). Due to the rapid growth of the size of the available corpora, the question of the modern-day lexicographer is not "How do I collect enough evidence?", but rather "How do I make sense of the vast amounts of evidence available to me? What tools can I use to discover patterns in the data and what can help me select appropriate quotes for the dictionary?", e.g. SketchEngine [9], GDEX [10], Pralex [13], DeepDict [3].

This paper is concerned with work on NomVallex, a corpus-based valency lexicon of Czech deverbal nouns (Section 3). Thus, the patterns, that we want to find in the data, concern the valency complementations of nouns, their forms and their possible and common combinations (Section 2). Sentences, in which the nouns

appear with several different complementations, are of particular interest due to their syntactic complexity; they are extracted from Czech corpora both automatically and manually (Section 4). Information about the preferred vs. alternative/supplementary forms and a rough indication of the frequency of forms (distinguishing common and rare forms) may be obtained by a quantitative analysis of the Prague Dependency Treebank (Section 5). This analysis may then be used to optimize the search for valency complementations and their combinations using Word Sketches (Section 6).

2 CLASSIFICATION OF ADNOMINAL FORMS AND THEIR COMBINATIONS

Czech is a highly inflectional language; valency complementations of a word are primarily distinguished by their morphological category of case. The forms expressing the individual complementations play a central role in the description of the valency behaviour of Czech nouns. The (im)possibility to use certain forms is even indicative of a meaning shift [12].

Adnominal forms often undergo both systemic and non-systemic changes when compared to forms of valency complementations of base verbs (the changes are also called typical and special shifts in surface forms of participants, see [12]). However, several constraints on adnominal forms and their combinations hold true; the major ones can be formulated as follows:

- an adnominal form is not allowed to have the form of a prepositionless accusative (with the exception of the free modification of duration, i.e. for how long, e.g. *čtení hodinu* ‘reading for an hour’);
- while prepositionless genitive and possessive forms (pronouns or adjectives) usually alternate as expressions of the same complementation, it is not possible to simultaneously express an agent by a prepositionless genitive and a different semantic role by a possessive form (e.g. **jejich bití chlapců* ‘their_{possessive} beating of boys_{genitive}’ cannot express the verbal construction ‘boys are beating them’).

These constraints, semantic shifts and the syntactic complexity of a noun group in general have an impact on the number of possible expressions of nominal modifications which is often higher than the number of corresponding verbal ones. Various forms of valency complementations are available to enable nouns to form grammatical constructions with two or more of their complementations expressed. Typically, an adnominal complementation can be expressed by at least two forms (variants).

However, we can see differences between the respective forms (variants) and their combinations in their usage. Some forms of complementations are theoretically possible and grammatical, yet they only function as alternative or supplementary forms and they are very rare¹. The case of complementation combinations is similar; some combinations of complementations and their forms are preferred, other combinations are only alternative or supplementary. A classification of adnominal forms is presented in Table 1.

¹ Herbst [7, p. xl] works with the following frequency scale: rare, frequent and very frequent.

Aspects of classification	Types of forms	
Grammar	grammatical	ungrammatical
Type of changes (shifts)	systemic	non-systemic
Preference in usage	preferred	alternative or supplementary
Frequency	frequent	rare

Tab. 1. Classification of adnominal forms

3 THE DEVELOPMENT OF NOMVALLEX

The corpus-based² valency lexicon of Czech nouns called NomVallex is a project building upon the theory of valency developed within the Functional Generative Description (FGD; [21]) and extending two existing valency lexicons developed within this tradition, Vallex (a valency lexicon of Czech verbs; [15]; [16]) and PDT-Vallex³ (containing valency patterns of verbs, nouns, adjectives and adverbs as they occurred in the PDT-corpora; [6]; [24]). Vallex provides semantic class membership [8] and valency patterns for all meanings (i.e. lexical units) of verbs included and it was the reason why Vallex was chosen as the base for the NomVallex project. Valency properties of nouns included in NomVallex are captured in the form of a valency frame for each meaning (lexical unit), and an enumeration of combinations of adnominal complementations representing various valency patterns, as extracted from Czech corpora [11].

The valency theory for the theoretical framework of the FGD has been detailed in numerous studies addressing especially valency of verbs [19] and nouns [18]; [12]. The following types of complementations may fill in the individual slots of the valency frames of verbs:

- inner participants or arguments that can be obligatory or optional: Actor (ACT), Patient (PAT), Addressee (ADDR), Effect (EFF), Origin (ORIG) (e.g., *Vláda_{ACT} omezila těžbu_{PAT} uranu ze současných 950 tun_{ORIG} na 500 tun_{EFF} ročně* ‘The government_{ACT} restricted uranium mining_{PAT} from the current 950 tonnes_{ORIG} to 500 tonnes_{EFF} per year’);
- obligatory free modifications or adjuncts, especially those with the meaning of direction (e.g., *přijet někam_{DIR3}* ‘to arrive somewhere’) or location (e.g., *přebývat někde_{LOC}* ‘to dwell somewhere’) and manner (e.g., *chovat se dobře_{MANN}* ‘to behave well’).

The same inventory of valency complementations is assumed for deverbal nouns denoting an action. The inventory of valency complementations of non-deverbal nouns and deverbal nouns undergoing substantial shifts in their meaning is supplemented with some more modifications, especially with a special nominal participant Material (MAT; e.g., *skupina lidí_{MAT}* ‘group of people’, *jedno balení*

² Another approach to valency of Czech nouns (so-called corpus-driven approach) was applied by Čermáková (2009).

³ <http://hdl.handle.net/11858/00-097C-0000-0023-4338-F>

*másla*_{MAT} ‘one package of butter’) and a free modification Appurtenance (APP; e.g., *Petrovo*_{APP} *auto* ‘Peter’s car’, *oddělení odbytu*_{APP} ‘sales department’).

3.1 Semantic Classes in NomVallex and a Preliminary List of Entries

Nouns representing five semantic classes are included in NomVallex, namely Communication (e.g. *odpověď* ‘answer’), Exchange (e.g. *dodávka* ‘delivery’), Contact (e.g. *dotyk* ‘touch’), Mental action (e.g. *dojem* ‘impression’), and Psychological state (e.g. *obava* ‘fear’). The assignment of a semantic class is carried over from Vallex: a noun is supposed to be assigned the same semantic class as its base verb in Vallex, with the exception of nouns that undergo a change in meaning. On the basis of the list of verbs in Vallex (see Table 2 for numbers of lexical units representing particular semantic classes), a preliminary list of noun entries was created. Within these semantic classes, we aim to provide valency patterns of all types of Czech nouns with a meaning denoting an action or an abstract result of an action. These nouns are either derived from verbs by productive means (suffixes *-(e)ní/tí*, as in *vykládání* ‘explaining // unloading’ or *pojetí* ‘conception’) or by non-productive means including the zero suffix (such as *vykládka* ‘unloading’, *výklad* ‘explanation / interpretation’). The preliminary list of candidate entries to be included in NomVallex currently contains 1230 lemmas, cf. Table 3.

	Commu- nication	Exchange	Contact	Mental action	Psych. verbs	Total
Verbs in Vallex	428	182	125	338	143	1216

Tab. 2. Number of verbal lexical units in Vallex.

	Commu- nication	Exchange	Contact	Mental action	Psych. state	Total
Productively deri- ved nouns	335	171	117	257	104	984
Non-productively derived nouns	110	38	14	56	28	246
Total	445	209	131	313	132	1230

Tab. 3. Number of lemmas of nouns included in the NomVallex preliminary list of entries

4 EXTRACTION OF VALENCY PATTERNS FROM CZECH CORPORA: METHODOLOGY

Searching for valency patterns of Czech nouns usually means searching for many various combinations of forms, including word order variants. We use the following Czech lemmatized and morphologically annotated corpora: the synchronic part of the Czech National Corpus (CNC)⁴, the web corpus Araneum Bohemicum Maximum⁵ [2] and corpora from the Prague Dependency Treebank Family,

⁴ <http://korpus.cz/>

⁵ http://ucts.uniba.sk/aranea_about/index.html

especially the Prague Dependency Treebank (PDT 3.0)⁶. The PDT 3.0 [1] contains Czech texts with complex and interlinked morphological (2 million words), syntactic (1.5 MW) and complex semantic annotation (0.8 MW).

Using the CNC and the Araneum corpus, valency patterns of Czech nouns are being extracted either with the help of Sketch Engine's Word Sketches [9], or by sophisticated CQL queries specified in the KonText application⁷. Searching through the PDT 3.0 is carried out by the tool called PML-TQ [23].

A manual syntactic annotation of the PDT 3.0 enables to carry out a precise quantitative analysis of all adnominal forms and their combinations (Section 5). However, as the corpus is rather small, some rare meanings and some non-systemic, supplementary or rare forms of complementations of nouns in their more frequent meanings do not occur in the data at all. In contrast, it is impossible to do a reliable quantitative analysis in the data of the CNC or the Araneum corpus unless a manual syntactic annotation is provided. On the other hand, these big corpora give evidence about assorted adnominal forms and their combinations that, although rare, should be captured in the valency lexicon.

While annotating individual lexemes from a manually prepared list of headwords (see Section 3), we suggest the following procedure:

1. To prepare a list of tentative lexical units by applying the systemic shifts to the units of the base verb; then to adjust this list to reflect meaning shifts and additional or missing lexical units;
2. To specify the most frequent forms and their combinations on the basis of the PDT-corpora data;
3. To use WordSketches with an extended Word Sketch Grammar to discover individual preferred forms of complementations of the lemma together with their most common/relevant lexical realisation; however, the Word Sketches are not sense disambiguated, so their output needs to be manually explored and the forms and examples added to the relevant lexical units as appropriate;
4. The statistical nature of Word Sketches makes them unsuitable for discovering alternative or supplementary forms; so we suggest to manually extend the list of possible realisations of each complementation with forms discovered by inspecting a sample of corpus concordances and/or by introspection and confirmed by manual search in the CNC data and the Araneum corpus;
5. To automatically create and run corpus searches for combinations of two or more expressed complementations, extracting concordances that could be used as dictionary examples;
6. To manually check the concordances extracted in step 5, selecting the appropriate examples.

In this paper, we refer to step 2 (Section 5) and explore the ways how steps 3 and 5 can be optimized in order to produce the least amount of output (Section 6) while providing the most useful evidence for the manual steps 4 and 6.

⁶<http://ufal.mff.cuni.cz/pdt3.0>

⁷http://wiki.korpus.cz/doku.php/en:pojmy:dotazovaci_jazyk

5 A QUANTITATIVE ANALYSIS OF COMBINATIONS OF PARTICIPANTS IN THE PDT 3.0

5.1 Annotation Scheme

The valency theory of the FGD was applied to the PDT-corpora data which resulted in a very complex and detailed annotation scheme [17]. Different meanings of words with valency that occur in the data are differentiated in PDT-Vallex. The annotation of valency consists of:

- determining and assigning a valency frame from PDT-Vallex;
- a lemma and a corresponding semantic role (ACT, PAT, ADDR, etc.) are assigned to the nodes for valency complementations expressed in the surface form of the sentence;
- obligatory valency complementations unexpressed on the surface are captured by an added (newly created) node with an artificial lemma (for example #PersPron), and the corresponding semantic role is also assigned.

First, we carried out a quantitative analysis focusing on relative frequencies of combinations of participants, modifying both productively and non-productively derived nouns in the PDT 3.0 (Section 5.2). Second, we present an analysis of combinations of adnominal forms for all productively derived nouns in PDT 3.0 (Section 5.3).

5.2 Relative Frequencies of Combinations of Adnominal Participants

Using the NomVallex preliminary list of entries (1230 lemmas, see Section 3), we searched through the PDT 3.0 for both productively and non-productively derived nouns representing the five selected semantic classes (Communication, Contact, Exchange, Mental action and Psychological state). 623 such lemmas occurred in the PDT 3.0 in a total of 8273 occurrences (see Table 4).

		Communi- cation	Exchange	Contact	Mental action	Psych. state	Total
Productively derived nouns	Lemmas	145	94	30	107	29	405
	Occurrences	1552	699	128	1236	179	3794
Non-productively derived nouns	Lemmas	102	34	10	54	18	218
	Occurrences	2163	540	16	1256	504	4479
Total	Lemmas	247	128	40	161	47	623
	Occurrences	3715	1239	144	2492	683	8273

Tab. 4. Number of lemmas and occurrences of nouns found in the PDT 3.0

Nodes added for obligatory complementations that are not present on the surface layer of the sentence enable us to search also for the unexpressed elements and to differentiate expressed and unexpressed valency modifications in our searches. We carried out a quantitative analysis focusing on relative frequencies of combinations of adnominal participants (i.e., ACT, PAT, ADDR, EFF, and ORIG). Figures 1 and 2 show that the most frequent combination is the case when only the PAT is expressed (with

the exception of non-productively derived nouns of Contact which represent the least frequent class and so the numbers may be influenced by their rare occurrence). The case when only ACT is expressed is the second most frequent combination, followed by the combinations ACT+PAT or PAT+ADDR, the latter of which is applicable only in the case of nouns that have ADDR in their valency frame. Interestingly, relative frequencies of the combination ACT+PAT are very low with nouns of Exchange and nouns of Contact. Relative frequencies of combinations of three participants – no more than 0.13% – are not shown in the Figures.

5.3 Relative Frequencies of Combinations of Forms of Adnominal Participants
 Analysis of forms of complementations of all productively derived nouns in PDT 3.0 (not only the nouns representing the five selected semantic classes) strongly confirms the intuition that most nouns occurring with a valency complementation occur with a single complementation expressed in prepositionless genitive (*zvyšování ceny* ‘increasing of the price’; relative percentage around 70%). Less common forms are listed in Table 5.

Concerning nouns with two complementations expressed, the most common combination is a complementation in genitive together with a complementation expressed by a prepositional group (*měření sil_{PAT} se Švédy_{ADDR}* ‘pitting (one’s) strength_{PAT} against the Swedes_{ADDR}’; almost 3%).

As mentioned above, combinations of three expressed valency participants are rare; of these, the most common is noun in the genitive and two prepositional groups (*snížení investic_{PAT} z jedné miliardy_{ORIG} na 600 milionů_{EFF}* ‘the fall of investments_{PAT} from 1000 million_{ORIG} to 600 million_{EFF}’; 6 occurrences in PDT 3.0, making up for less than 0.1% of nouns with expressed valency participants).

The PDT 3.0 does not contain any instance of a noun modified by four expressed valency participants.

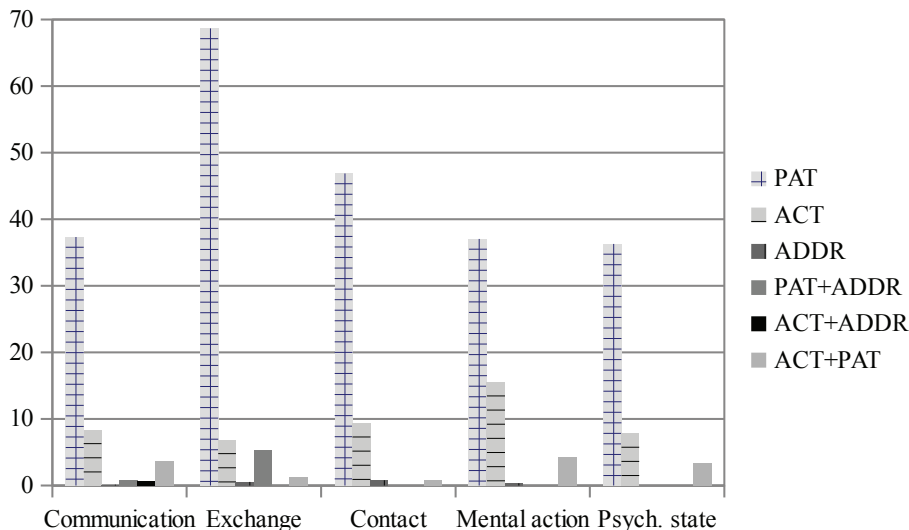


Fig. 1. Relative frequencies of selected combinations of participants modifying *productively* derived nouns in the PDT 3.0

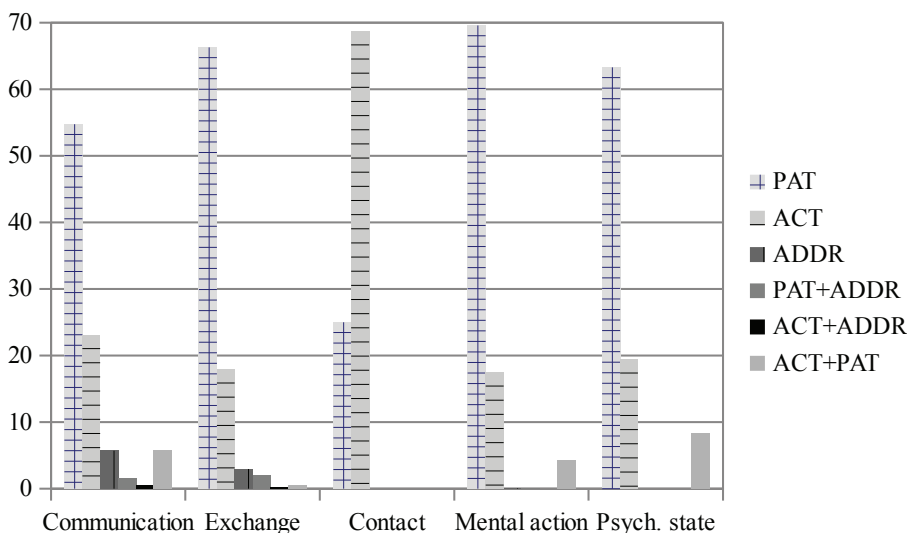


Fig. 2. Relative frequencies of selected combinations of participants modifying *non-productively* derived nouns in the PDT 3.0

Percentage	Expressed complementations	Example
around 70%	prepositionless genitive	<i>zvyšování ceny</i> 'increasing of the price'
almost 9%	prepositional groups	<i>srovnávání s čím</i> 'comparison with sth'
almost 6%	possessive forms	possessive adjectives: <i>Clintonův</i> 'Clinton's' possessive pronouns: <i>náš</i> 'ours'
almost 3%	prepositionless genitive + prepositional group	<i>měření sil_{PAT} se Švédy_{ADDR}</i> 'pitting (one's) strength _{PAT} against the Swedes _{ADDR} '
about 1.5%	indeclinable noun	<i>vedení Oilers</i> 'the leadership of Oilers' acronyms: <i>rozdělení ČSFR</i> 'the division of ČSFR (i.e. Czechoslovakia)'
0.8%	content clause	<i>prohlášení, že ...</i> 'a declaration that ...'
0.5%	infinitive	<i>oprávnění zastavit vozidlo</i> 'authorization to stop a vehicle'
< 0.1%	prepositionless genitive + prepositional group + prepositional group	<i>snížení investic_{PAT} z jedné miliardy_{ORIG} na 600 milionů_{EFF}</i> 'the fall of investments _{PAT} from 1000 million _{ORIG} to 600 million _{EFF} '

Tab. 5. Combinations of forms expressing complementations of a single noun. Of the combinations with percentage below 1%, only those of particular interest are listed.

6 WORD SKETCH AND CORPUS SEARCH OPTIMIZATION

The fact, that a certain form is not common among adnominal complementations modifying a particular noun, does not necessarily mean that such complementation cannot be found by Word Sketches. Quite to the contrary, these forms may be specific to a limited group of nouns and thus forming more statistically significant collocations. For example, the dative complementations appear with less than 1% of noun instances in the examined PDT 3.0 data. However, using our extended SketchGrammar, a Word Sketch of the noun *předání* ‘delivery, handover’ contains several typical lexical realisations of the dative complementation: *exekutorovi* ‘to the executor’, *zdravotníkům* ‘to the paramedics’, *Číně* ‘to China’, *zákazníkovi* ‘to the customer’ etc.

On the other hand, we may try to judge the expected utility of WordSketches by comparing the output of the search for valency complementations only to the output of the search for any complementations of nouns, including the complementations that are not part of the valency frame (i.e. the free modifications). For example, there are twice as many occurrences of the combination of a complementation expressed as a noun in the genitive and complementation expressed by a prepositional group when we allow for non-valency complementations. In other words, if we created a Word Sketch rule for discovering this combination, we may expect it would trigger about half of the time on a combination such that at least one of the two complementations does not actually belong to valency frame of the noun.

Up to now, we have extended the Czech Word Sketch Grammar with some forms of adnominal collocations that are typical of valency, especially with collocations in prepositionless dative and instrumental. Upon further examination of the frequent forms, we plan to experiment with adding ternary relations capturing some of the most common combinations of two adnominal complementations.

7 CONCLUSION

The paper refers to the current work on NomVallex, the corpus-based valency lexicon of Czech nouns. We carried out a quantitative analysis of valency behaviour of Czech deverbal nouns in the PDT 3.0. Reflecting the difference between productively and non-productively derived nouns and their semantic class membership, we show that order of relative frequencies of combinations of adnominal participants is almost the same for all observed types of nouns. We also specify the most frequent and infrequent combinations of adnominal forms, which enables to optimize Word Sketches.

ACKNOWLEDGEMENTS

The research reported in the paper was supported by the Czech Science Foundation under the project GA16-02196S. This work has been using language resources developed, stored and distributed by the LINDAT/CLARIN project of the Ministry of Education, Youth and Sports of the Czech Republic (project LM2015071).

The last author was supported by grant no. RVO 67985840 of the Czech Academy of Sciences.

References

- [1] Bejček, E. et al. (2013). Prague Dependency Treebank 3.0. Data/software, Univerzita Karlova v Praze, MFF, ÚFAL, Prague, Czech Republic.
- [2] 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.
- [3] Bick, E. (2009). DeepDict – A Graphical Corpus-based Dictionary of Word Relations. In Bick, E., Hagen, K., Müürisep, K., and Trosterud, T., editors, *Proceedings of the NODALIDA 2009 workshop Constraint Grammar and robust parsing*. NEALT Proceedings Series, Vol. 8, i-ii, Northern European Association for Language Technology (NEALT), Tartu University.
- [4] Collins *COBUILD Advanced Learner's Dictionary* (2014).
- [5] Čermáková, A. (2009). *Valence českých substantiv*. Lidové noviny, Praha.
- [6] Hajič, J., Panevová, J., Urešová, Z., Bémová, A., Kolářová, V., and Pajas, P. (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, Växjö University Press, Sweden.
- [7] Herbst, T., Heath, D., Roe, I. F., and Götz, D. (2004). A valency dictionary of English: a corpus-based analysis of the complementation patterns of English verbs, nouns, and adjectives. Walter de Gruyter, Berlin, Germany.
- [8] Kettnerová, V., Lopatková, M., and Hrstková, K. (2008). Semantic Classes in Czech Valency Lexicon: Verbs of Communication and Verbs of Exchange. In *Lecture Notes in Computer Science, Vol. 5246, Proceedings of the 11th International Conference, TSD 2008*, pages 109–116, Springer, Berlin – Heidelberg, Germany.
- [9] Kilgariff, A. and Tugwell, D. (2001). WORD SKETCH: Extraction and display of significant collocations for lexicography. In *Proc Collocations workshop*, pages 32–38, ACL, Toulouse, France.
- [10] Kilgariff, A., Husák, M., McAdam, K., Rundell, M., and Rychlý, P. (2008). GDEX: Automatically finding good dictionary examples in a corpus. In *Proceedings of the 13th EURALEX International Congress*, pages 425–432, Spain.
- [11] Klímová, J., Kolářová, V., and Vernerová, A. (2016). Towards a Corpus-based Valency Lexicon of Czech Nouns. In Kernerman, I. et al., editors, *Globalex 2016, Lexicographic Resources for Human Language Technology*, pages 1–7. Accessible at: http://ailab.ijs.si/globalex/files/2016/06/LREC2016Workshop-GLOBALEX_Proceedings-v2.pdf.
- [12] Kolářová, V. (2014). Special valency behavior of Czech deverbal nouns. In Spevak, O., editor, *Noun Valency*, pages 19–60, John Benjamins, Amsterdam, Netherlands.
- [13] Lexikální databáze Pralex: Accessible at: <http://lexiko.ujc.cas.cz/index.php?page=23>.
- [14] Longman Dictionary of Contemporary English. (2014). 6th Edition. Pearson.
- [15] Lopatková, M., Kettnerová, V., Bejček, E., Vernerová, A., and Žabokrtský, Z. (2015). *VALLEX 3.0 – Valenční slovník českých sloves*. Charles University in Prague. Accessible at: <http://ufal.mff.cuni.cz/vallex/3.0/>.
- [16] Lopatková, M., Kettnerová, V., Bejček, E., Vernerová, A., and Žabokrtský, Z. (2016). *Valenční slovník českých sloves VALLEX*. Karolinum, Praha.
- [17] Mikulová, M. et al. (2006). Annotation on the tectogrammatical level in the Prague Dependency Treebank. Annotation manual. Technical Report TR-2006-30, ÚFAL MFF UK, Praha.
- [18] Panevová, J. (2002). K valenci substantiv (s ohledem na jejich derivaci). In *Zborník matice srpske za slavistiku*, vol. 61, pages 29–36.
- [19] Panevová, J. (2014). Contribution of Valency to the Analysis of Language. In Spevak, O., editor, *Noun Valency*, pages 1–18, John Benjamins Publishing Company, Amsterdam, Netherlands.

- [20] *Příruční slovník jazyka českého*. (1935–1957). Státní nakladatelství – Školní nakladatelství – Státní pedagogické nakladatelství, Praha.
- [21] Sgall, P., Hajičová, E., and Panevová, J. (1986). *The Meaning of the Sentence in its Semantic and Pragmatic Aspects*. Reidel, Dordrecht.
- [22] Stevenson A., editor (2010). *Oxford Dictionary of English*. Oxford University Press.
- [23] Štěpánek, J. and Pajas, P. (2010). Querying Diverse Treebanks in a Uniform Way. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, pages 1828–1835, European Language Resources Association (ELRA), Valletta, Malta.
- [24] Urešová, Z. (2012). Building the PDT-VALLEX valency lexicon. In *Proceedings of the fifth Corpus Linguistics Conference*, pages 1–18, University of Liverpool, Liverpool, UK.