

Czech-English Bilingual Valency Lexicon Online

Eva Fučíková, Jan Hajič, Jana Šindlerová and Zdeňka Urešová

Institute of Formal and Applied Linguistics
Faculty of Mathematics and Physics
Charles University in Prague

E-mail: {fucikova|hajic|sindlerova|uresova}@ufal.mff.cuni.cz

Abstract

We describe CzEngVallex, a bilingual Czech–English valency lexicon which aligns verbal valency frames and their arguments. It is based on a parallel Czech-English corpus, the Prague Czech-English Dependency Treebank (PCEDT), where for each occurrence of a verb, a reference to the underlying Czech and English valency lexicons (PDT-Vallex and CzEngVallex, respectively) is recorded. The CzEngVallex then pairs the entries (verb senses) of the two lexicons, and allows for detailed studies of verb valency and argument structure in translation and also compare the approaches to valency in the two languages on the background of the same underlying theory, the Functional Generative Description. The CzEngVallex lexicon is now accessible online, and we will also describe here the search interface which makes certain complex queries possible, using the lexicon and accessing the associated examples of verb sense translations, as extracted from the PCEDT corpus.

1 The PCEDT parallel corpus and its lexicons

Valency, or verb argument structure, is an important phenomenon both in linguistic studies as well as in language technology applications, since the verb is considered the core of a clause in (almost) every natural language utterance. Various dictionaries have been built - from Propbank [13] to Framenet [1] as well as various valency lexicons exist for several languages, such as Walenty [16] for Polish, Verbalex [8] or Vallex [9] for Czech, Valence Lexicon for a Treebank of German [3] for German etc. However, there are no truly multilingual valency dictionaries linked to corpora.

The Prague Czech-English Dependency Treebank (PCEDT 2.0) [4] contains the WSJ part of the Penn Treebank [10] and its manual professional translation to Czech, annotated manually using the tectogrammatical representation [11], first used for the Prague Dependency Treebank 2.0 (PDT) [5]. The tectogrammatical representation is in turn based on the Functional Generative Description theory [17].

The PCEDT contains 866,246 English tokens and 953,187 Czech tokens, aligned manually sentence-by-sentence and automatically word-by-word. It is annotated on all three annotation layers of the PDT: morphological, analytical (surface dependency syntax) and tectogrammatical (syntactic-semantic). However, as opposed to the PDT which is annotated fully manually¹, the PCEDT has been annotated for structure and valency at the tectogrammatical representation layer manually, but for POS, morphology and surface syntax only automatically.² Both sides of the tectogrammatical representation have been enriched with valency annotation, using two valency lexicons: PDT-Vallex for Czech and EngVallex for English.

The PDT-Vallex [6, 19, 18] is a valency lexicon originally developed for the PDT annotation. It contains almost 12,000 verb frames for about 7,000 verbs, roughly corresponding to verb senses found during the annotation of the PDT and the PCEDT. For each frame, verb arguments are listed together with their obligatoriness and required morphosyntactic realization(s). Each occurrence of a verb in the PDT (and the Czech side of the PCEDT) is linked to one verb frame in the PDT-Vallex lexicon.

The EngVallex [2] has been created for the English side of the PCEDT annotation. It is a semi-manual conversion of the Propbank frame files [13] into the PDT style of capturing valency information in valency frames, as used for Czech. The correspondence of the original Propbank entries and valency frames in the EngVallex is not necessarily 1:1 - entries have been occasionally merged or split. It contains over 7,000 frames for 4,300 verbs.

2 The CzEngVallex lexicon

The CzEngVallex lexicon [20, 21] is a bilingual valency lexicon with explicit pairing of verb senses (corresponding to valency frames) and their arguments, built upon the PCEDT. It contains 20,835 frame pairs describing the way verbal valency is mapped between languages, in particular between Czech and English.³

The lexicon draws on the Functional Generative Description Valency Theory (FGDVT). In this dependency approach, valency is seen as the property of some lexical items - above all verbs - to select for certain complementations in order to form larger units of meaning (phrase, sentence etc.). The governing lexical unit then governs both the morphosyntactic properties of the dependent elements and their semantic interpretation (roles). The number and realization of the selected

¹With the exception of certain lexical node attributes.

²The surface dependency syntax on the English side has been derived from the Penn Treebank constituent syntax annotation, using head percolation rules, and thus can be considered semi-manual as well.

³This lexicon has been built within the project called “A comparison of Czech and English verbal valency based on corpus material (theory and practice)”, for more information, see <https://ufal.mff.cuni.cz/czengvallex> and <https://ufal.mff.cuni.cz/biblio/?section=grant&id=-5269651103966024613&mode=view>.

dependent elements constituting the valency structure of the phrase (or sentence) can be represented by valency frames, which can be listed in valency dictionaries. The basics of the FGD approach to valency can be found, e.g., in [14] or [7].

The annotation interface for the CzEngVallex is an extension of the tree editor TrEd [12]⁴ environment. It allows displaying and annotating sentential tree structures annotated on multiple linguistic layers with a variety of tags using either the Prague Markup Language (PML) format⁵ or the Treex format.⁶ Treex (formerly TectoMT) [22, 15] is a development framework for general as well as specialized NLP tasks (such as machine translation) working with tectogrammatically annotated structures.

This lexicon is a valuable resource to be used both for linguistically oriented comparative research, as well as for an innovative use in various NLP tasks. Its electronic version⁷ is available from the repository of the Center of linguistic research infrastructure LINDAT/CLARIN in XML format and it is also available using a specific access portal in a searchable version, as described in this paper.⁸

It should be noted that not all verbs from the PCEDT can be found in the CzEngVallex: some verbs have not been translated at all as verbs, and vice versa, and some verb-verb translations have been so structurally different that they have not been included in the CzEngVallex.

Some of these cases can be extracted by inspecting the data where comments have been added by the annotators, and others by simple technical means (finding verbs with no matching alignment, finding verbs aligned to nouns, adjectives, or other structurally divergent tree segments).

| Language | Verb types | Frame types | PCEDT Tokens | |
|----------|------------|-------------|--------------|---------|
| | | | verbs | aligned |
| English | 3,288 | 4,967 | 130,514 | 86,573 |
| Czech | 4,192 | 6,776 | 118,189 | 85,606 |

Table 1: Alignment coverage statistics - CzEngVallex/PCEDT

According to [20], 66% of English verb tokens found in the corpus have been aligned and can be found in the CzEngVallex (for Czech verb occurrences, it is 72%). Also, due to the fact that CzEngVallex is restricted to the parallel corpus only, it also covers only about 2/3rd of the underlying valency lexicons, the PDT-Vallex and the EngVallex. Exact statistics are given in Table 1 (taken from [20]).

While both the underlying lexicons build upon the tectogrammatical representation used for both sides of the Prague Czech-English Dependency Treebank - there are the same five core arguments (ACTor, PATient, ADDRessee, EFFect and ORIGIN, about 40 additional free modifications, which might become obligatory

⁴<http://ufal.mff.cuni.cz/tred>

⁵<http://ufal.mff.cuni.cz/jazz/PML>

⁶<http://ufal.mff.cuni.cz/treex>

⁷<http://hdl.handle.net/11234/1-1512>

⁸<http://lindat.mff.cuni.cz/services/CzEngVallex>

earmark EngVallex : **vyhradit** PDT-Vallex

earmark ACT() **PAT()** ?**AIM()** EngVallex

to reserve or set aside for a particular purpose: passive

- However , several people said that British Air might yet rejoin the bidding group and that the carrier made the statement [-1] to answer questions from British regulators about how it plans [-2] to use proceeds of a securities offering previously earmarked [1] for the UAL buy-out [-T-3] .
- Venezuela has also earmarked \$ 200 million [U] in new money for light-crude exploration .
- The Senate voted [-1] to authorize \$ 4.5 billion [U] for SDI spending in the current fiscal year , but the House , [-2] reflecting a dramatic erosion of support for the program , earmarked only \$ 3.1 billion [U] .

vyhradit ACT() **PAT()** **ADDR()** PDT-Vallex

pridělit, určit

- vyhradit obchodování bankám; v. rubriku sportovním zprávám; v. oblast pro sáňkování

Argument mapping for "earmark" (ev-w1082f1) and "vyhradit" (v-w7924f1):

ACT → **ACT**
PAT → **PAT**
BEN → **ADDR**

Corpus example(s): Close [X]

pcedt Some of the money—**PAT** will go to the Chico Mendes Foundation , but **itPAT** is n't **earmarked for BEN groups BEN** trying to save the rain forest . -RRB-

pcedt Něco—**PAT** z peněz půjde Nadaci Chica Mendese , ale **nejsou vyhrazeny skupinámADDR** usilujícím o záchranu deštného pralesa .)

Hint: Search in real corpus examples is limited to the first 20 examples in the PCEDT corpus. The usual test data from the Penn Treebank corpus, i.e., sections 20-24, are always invisible, reducing the number of browsable examples from the corpus by ca. 20 %.

Once the selected examples appear on the screen, use your mouse to hover over them to get more information on the highlighted words and symbols.

PCEDT example texts extracted and processed on Po srp 24 23:08:12 CEST 2015

Browse lexicon in direction...

[Cz→En](#) [En→Cz](#)

SEARCH VERBS by lemmas and/or functors (or pairs of functors)

English verb: Czech verb (optional):

Slots En→Cz:

| | | | | | |
|----------------------|---|----------------------|----------------------|---|----------------------|
| <input type="text"/> | → | <input type="text"/> | <input type="text"/> | → | <input type="text"/> |
| <input type="text"/> | → | <input type="text"/> | <input type="text"/> | → | <input type="text"/> |
| <input type="text"/> | → | <input type="text"/> | <input type="text"/> | → | <input type="text"/> |
| <input type="text"/> | → | <input type="text"/> | <input type="text"/> | → | <input type="text"/> |

Found: 1 pair of frames (total number is 20835)
(listing is limited to the first 100 pairs)

Earmark.Vyhradit

-OR-

SELECT VERBS
Show English (source side) verbs starting with...

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| A | B | C | D | E | F | G | H | I |
| J | K | L | M | N | O | P | Q | R |
| S | T | U | V | W | Y | Z | | |

Czech verbs in CzEngVallex paired with English verb 'earmark' (click to show):

[ALL Vyčlenit Vyhradit](#)

Figure 1: Main browsing and search form with search result for “earmark” → “vyhradit”

for any given verb, etc. - they also inevitably differ in several respects. First, instead of writing notes and examples to distinguish between verb senses of the individual valency frames, the creators of EngVallex often left a non-obligatory free modification in the valency frame, especially if they also found it in PropBank (where there is no obligatory/non-obligatory distinction being made). Such a free modification thus might be sometimes surprising to someone working only with PDT-Vallex so far. Also, the interpretation of certain label definitions such as ADDR vs. BEN was sometimes slightly different, as well as the conventions for using PAT and EFF with “verba dicendi” (say, explain, write, ...), and also the treatment of idioms and light verbs. These differences often show in the results of the searches as described below, and they do not represent “true“ translation differences, but rather a difference in the application of the FGDVT theory and the tectogrammatical annotation guidelines to the two languages. Nevertheless, we consider that a unique opportunity to discover and study these differences through CzEngVallex (and its online search interface).

3 Online searching and browsing

The CzEngVallex lexicon is available at <http://lindat.cz> among “More Apps”, as the PDT-Vallex and EngVallex lexicons are.⁹ The main search interface asks for a source verb (the direction might be switched using the En→Cz and Cz→En buttons), and either one of possible translations¹⁰ to show all translations found in the CzEngVallex lexicon. If one of the verb input fields is left empty, the list of all translations will be displayed, allowing to directly select only one pair for a full display.¹¹

Once all possible translation pairs are displayed, clicking on one of them shows the linked valency frames and the argument mapping within them. Fig. 1 shows the screen for the source verb “earmark” linked to “vyhradit”, all translations of “earmark” (lower right part of the screenshot), and the linked valency frames in the left column, with the following color coding: olive color is used as the background for the verb pair, dark yellow for the frame headline (with red-coded argument labels in it), and light violet/blue for argument mapping. Comments and examples recorded directly in the lexicon are on a light grey background. Corpus examples are separated by a dark grey bar (they can be hidden or made visible by a single click).

The same color coding is used on the webpages of the two underlying monolingual lexicons, PDT-Vallex and EngVallex, which are accessible by direct links¹² from the CzEngVallex entries.¹³ There are two types of links to these monolingual lexicons - the two links in the headline (with olive color background) lead to the complete entry, while the PDT-Vallex/EngVallex links at the individual valency frame pairs (color-coded dark yellow) lead directly to the particular PDT-Vallex and EngVallex valency frames, respectively. The monolingual lexicons can be used for getting more information, such as more corpus examples or the morphosyntactic information for individual frame slots; for many verbs in PDT-Vallex, there are also additional verb senses, namely those used in other corpora than the PCEDT.

If the user is not satisfied with the selected pair, “ALL” can be selected (see lower right of Fig. 1), at the beginning of the list of plain translation equivalents), and the user is presented with the verb *earmark* and its two possible translations (Fig. 2).

If the user clicks on the button, examples from the

⁹Or directly at <http://lindat.mff.cuni.cz/services/CzEngVallex>.

¹⁰Or any of the fields (but not all) can be left blank.

¹¹The verbs should be given in base form (as a lemma). It is also possible to use a standard regular expression on the verb lemma, e.g. `[a-d].*` for all verbs starting with a, b, c, or d. Full string match is assumed (i.e. the `'^'` and `'$'` characters for string start and end should not be used). Please note also that the list of verb pairs shown is limited to the first 100 pairs only.

¹²Shown as “superscripts” at the displayed verb lemmas and frames in the left column, cf. also Fig. 1.

¹³Except, of course, the underlying lexicons do not contain the argument pairing. They are also accessible at their own websites independently: <https://lindat.mff.cuni.cz/services/PDT-Vallex> and <https://lindat.mff.cuni.cz/services/EngVallex>.



Figure 2: Two sense- and argument-aligned translations of *earmark* to Czech

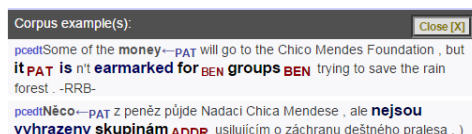


Figure 3: Corpus example for *earmark-vyhradit*

PCEDT parallel corpus are revealed (see Fig. 1 on the lower left, or as shown separately in Fig. 3). In the examples, presented as plain text, the verb and its arguments are highlighted based on the manual syntactic-semantic (tectogrammatical) annotation of the corpus. Arguments are also marked with the argument labels - in the example below, both PATs (including the one annotated only by co-reference) and the corresponding BEN- and ADDR-labelled arguments are shown. Both ACT-Tors are elided in the passive construction used in the sentence and thus not shown, even if annotated in the tectogrammatical representation (and linked in the CzEng-Vallex).

It is also possible to search for particular argument types, specifying either side (Czech or English) or both, and moreover, any combination of argument pairs can be specified.¹⁴ In every pair, either side can be left out (i.e. underspecified); it will then find all verb pairs where there is the entered argument on the specified side (language), and any argument on the other side to which it is linked. On the other hand, specifying a string of dashes -- means that the particular argument must be marked as “not present” in CzEngVallex (same dashed string). For example, if a user wants to search for verb pairs where the English verb has the DPHR argument

¹⁴Up to 7, which is the maximum number of pairs found in the CzEngVallex entries.

SEARCH VERBS by lemmas and/or functors (or pairs of functors)

English verb: Czech verb (optional):

Slots En→Cz:

| | | | | |
|------|---|----|---|--|
| DPHR | → | -- | → | |
| | → | | → | |
| | → | | → | |
| | → | | → | |

Found: 52 pairs of frames (total number is 20835)
(listing is limited to the first 100 pairs)

Be.Probíhat | Break.Překlenout | Come.Dozrát |
Come.Objevit_se | Come.Objevit_se |
Come.Projevovat_se | Come.Přiblížit_se |
Come.Uskutečnit_se | Come.Uvědomovat_si |
Come.Vyrovnat_se | Curry.Podlézat |
Cut.Dojmout | Cut.Odříznout | Cut.Přerušit |

Figure 4: Intermediate result, searching for DPHR→--

while the Czech counterpart has an “empty” argument -- linked to it, i.e., the English verb has a phrasal component while the Czech verb contains the phrasal meaning in the verb itself, these two arguments (in this order, provided that the direction En→Cz is selected) should be entered into a pair of “Slots” windows.

Once the search button is pressed, a list of all verb pairs that fulfill the conditions of the Slots query are displayed below the Search button (see lower part of Fig. 4). The user can visually check which pairs to display in full in the right-hand side of the screen, and then get them there by clicking on the particular pair. The list of pairs displayed might get very long, especially if a weak query (such as “show all pairs with ACT - ACT argument pairings”) is entered; for that purpose, a count of pairs is displayed above the list to alert the user about the size of the list (and provide some statistics at the same time). In our case, if the user selects the “Come.Dozrát” pair, the resulting pair of frames and a corpus example is shown (Fig. 5).

Both the search by concrete verb (or a pair of verbs) and search by arguments can be combined. This is especially useful when searching within very frequent verbs with many verb senses (valency lexicon entries), such as to be or to have.

In addition, one can simply browse the lexicon using the letter-labelled buttons in the lower right part of the search interface (Fig. 1). After clicking on one of those buttons, a list of verbs starting the selected letter is displayed (can be long!), and a particular verb can be selected to see all possible senses of that verb and their pairings.

The screenshot shows a web interface for a bilingual valency dictionary. On the left, there are two main sections: one for the English lemma 'come' (EngVallex) and one for the Czech lemma 'dozrát' (PDT-Vallex). The English section shows the lemma 'come' with its DPHR (DPHR) mapping and a corpus example: 'The attack commercial has come of age in an off-election year.' The Czech section shows the lemma 'dozrát' with its ACT (ACT) mapping and a corpus example: 'čas už dozrá; ve vědomí národa d. přesvědčení'. Below these sections, there is an 'Argument mapping for "come" (ev-w590f7) and "dozrát" (v-w808f1):' section showing 'DPHR → --' and 'ACT → ACT'. A 'Corpus example(s):' section is also present. On the right, there is a search interface with a 'Browse lexicon in direction...' section containing 'Cz→En' and 'En→Cz' buttons. Below that is a 'SEARCH VERBS by lemmas and/or functors (or pairs of functors)' section with input fields for 'English verb:' and 'Czech verb (optional):'. A 'Slots En→Cz:' section contains a table of slots and arrows. A 'Search' button is at the bottom right. Below the search interface, it says 'Found: 1 pair of frames (total number is 20835) (listing is limited to the first 100 pairs)'. The result is 'Come Dozrát'.

Figure 5: Result for *come [of age]→dozrát*, example of DPHR→-- mapping

4 Conclusions

We have described some of the basic features of CzEngVallex, a bilingual valency dictionary created over the Prague Czech-English Dependency Treebank, a parallel corpus of 1 mil. words. The interlinked lexicons and the corpus are now publicly available online and searchable,¹⁵ making it possible for a wide audience to get more insight into the use of verb arguments in translation, benefiting both in linguistic studies as well as in language technology, especially machine translation. The search interface is still under development, and new possibilities will be provided in future versions, such as search based on required morphosyntactic form of arguments, search within examples, “negative” search queries (for exclusions of certain pairings), etc. Moreover, we also consider updating the underlying corpora based on findings using the CzEngVallex, such as unifying the rules for argument labeling across languages; it would consequently improve the quality and consistency of CzEngVallex as well.

Acknowledgments

This work has been supported by the grant No. GP13-03351P of the Grant Agency of the Czech Republic and has been partially supported and has been using language resources developed, stored and distributed by the LINDAT/CLARIN project of the Ministry of Education of the Czech Republic (project LM2010013).

¹⁵And of course, for download: <http://hdl.handle.net/11234/1-1512>

References

- [1] Collin F. Baker, Charles J. Fillmore, and John B. Lowe. The Berkeley FrameNet Project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 1*, ACL '98, pages 86–90, Stroudsburg, PA, USA, 1998. Association for Computational Linguistics.
- [2] Silvie Cinková. From Propbank to Engvallex: Adapting the PropBank-Lexicon to the Valency Theory of the Functional Generative Description. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)*, Genova, Italy, 2006.
- [3] Erhard Hinrichs and Heike Telljohann. Constructing a Valence Lexicon for a Treebank of German. In *Proceedings of the 7th International Workshop on Treebanks and Linguistic Theories*, pages 41–52, 2009.
- [4] J. Hajič, E. Hajičová, J. Panevová, P. Sgall, O. Bojar, S. Cinková, E. Fučíková, M. Mikulová, P. Pajas, J. Popelka, J. Semecký, J. Šindlerová, J. Štěpánek, J. Toman, Z. Urešová, and Z. Žabokrtský. Announcing Prague Czech-English Dependency Treebank 2.0. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, pages 3153–3160, Istanbul, Turkey, 2012. European Language Resources Association.
- [5] Jan Hajič, Jarmila Panevová, Eva Hajičová, Petr Sgall, Petr Pajas, Jan Štěpánek, Jiří Havelka, Marie Mikulová, Zdeněk Žabokrtský, Magda Ševčíková-Razímová, and Zdeňka Urešová. Prague Dependency Treebank 2.0, LDC Cat. No. LDC2006T01, also at <http://hdl.handle.net/11858/00-097c-0000-0001-b098-5>, 2006.
- [6] Jan Hajič, Jarmila Panevová, Zdeňka Urešová, Alevtina Bémová, Veronika Kolářová, and Petr Pajas. PDT-VALLEX: Creating a Large-coverage Valency Lexicon for Treebank Annotation. In Erhard Nivre, Joakim//Hinrichs, editor, *Proceedings of The Second Workshop on Treebanks and Linguistic Theories*, volume 9 of *Mathematical Modeling in Physics, Engineering and Cognitive Sciences*, pages 57—68, Vaxjo, Sweden, 2003. Vaxjo University Press.
- [7] E. Hajičová and P. Sgall. Dependency Syntax in Functional Generative Description. *Dependenz und Valenz—Dependency and Valency*, 1:570–592, 2003.
- [8] Horák, Aleš and Pala, Karel and Hlaváčková, Dana. Preparing VerbaLex Printed Edition. In *Seventh Workshop on Recent Advances in Slavonic Natural Language Processing, RASLAN 2013*, pages 3–11, 2013.
- [9] Markéta Lopatková, Zdeněk Žabokrtský, and Václava Ketnerová. *Valenční slovník českých sloves*. Karolinum, Praha, 2008.

- [10] Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. Building a large annotated corpus of english: The penn treebank. *COMPUTATIONAL LINGUISTICS*, 19(2):313–330, 1993.
- [11] Marie Mikulová, Alevtina Bémová, Jan Hajič, Eva Hajičová, Jiří Havelka, Veronika Kolářová, Markéta Lopatková, Petr Pajas, Jarmila Panevová, Magda Razímová, Petr Sgall, Jan Štěpánek, Zdeňka Urešová, Kateřina Veselá, Zdeněk Žabokrtský, and Lucie Kučová. Anotace na tektogramatické rovině Pražského závislostního korpusu. Anotátorská příručka. Technical Report TR-2005-28, ÚFAL MFF UK, Prague, Prague, 2005.
- [12] Petr Pajas and Peter Fabian. Tred 2.0 - newly refactored tree editor. <http://ufal.mff.cuni.cz/tred>, 2011.
- [13] Martha Palmer, Daniel Gildea, and Paul Kingsbury. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31(1):71–106, 2005.
- [14] Jarmila Panevová. On verbal frames in Functional generative description I. *Prague Bulletin of Mathematical Linguistics*, (22):3–40, 1974.
- [15] Martin Popel and Zdeněk Žabokrtský. TectoMT: modular NLP framework. *Advances in Natural Language Processing*, pages 293–304, 2010.
- [16] Adam Przepiórkowski, Elżbieta Hajnicz, Agnieszka Patejuk, Marcin Woliński, Filip Skwarski, and Marek Świdziński. Walenty: Towards a comprehensive valence dictionary of Polish. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014*, pages 2785–2792, Reykjavík, Iceland, 2014. ELRA.
- [17] Petr Sgall, Eva Hajičová, and Jarmila Panevová. *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Dordrecht, Reidel, and Prague, Academia, Prague, 1986.
- [18] Zdeňka Urešová. *Valence sloves v Pražském závislostním korpusu*. Studies in Computational and Theoretical Linguistics. Ústav formální a aplikované lingvistiky, Praha, Czechia, 2011.
- [19] Zdeňka Urešová. *Valenční slovník Pražského závislostního korpusu (PDT-Vallex)*. Studies in Computational and Theoretical Linguistics. Ústav formální a aplikované lingvistiky, Praha, Czechia, 2011.
- [20] Zdeňka Urešová, Ondřej Dušek, Eva Fučíková, Jan Hajič, and Jana Šindlerová. Bilingual English-Czech Valency Lexicon Linked to a Parallel Corpus. In *Proceedings of the The 9th Linguistic Annotation Workshop (LAW*

IX 2015), Stroudsburg, PA, USA, 2015. Association for Computational Linguistics.

- [21] Zdeňka Urešová, Eva Fučíková, and Jana Šindlerová. Czenvallex: Mapping valency between languages. Technical Report TR-2015-58, ÚFAL MFF UK, 2015.
- [22] Zdeněk Žabokrtský. Treex – an open-source framework for natural language processing. In Markéta Lopatková, editor, *Information Technologies – Applications and Theory*, volume 788, pages 7–14, Košice, Slovakia, 2011. Univerzita Pavla Jozefa Šafárika v Košiciach.