Automatic Mapping Lexical Resources: A Lexical Unit as the Keystone

Eduard Bejček, Václava Kettnerová and Markéta Lopatková

Charles University in Prague, Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics
Malostranské náměstí 25, 118 00 Prague 1, Czech Republic
{bejcek,kettnerova,lopatkova}@ufal.mff.cuni.cz

Abstract
This paper presents the fully automatic linking of two valency lexicons of Czech verbs: VALLEX and PDT-VALLEX. Despite the same theoretical background adopted by these lexicons and the same linguistic phenomena they focus on, the fully automatic mapping of these resources is not straightforward. We demonstrate that converting these lexicons into a common format represents a relatively easy part of the task whereas the automatic identification of pairs of corresponding valency frames (representing lexical units of verbs) poses difficulties. The overall achieved precision of 81% can be considered satisfactory. However, the higher number of lexical units a verb has, the lower the precision of their automatic mapping usually is. Moreover, we show that especially (i) supplementing further information on lexical units and (ii) revealing and reconciling regular discrepancies in their annotations can greatly assist in the automatic merging.

Keywords: linking lexical resources, valency lexicons, lexical units of verbs

1. Introduction
Lexical resources play a crucial role in various NLP applications. In the last decades, a great number of lexical resources (LRs)—encoding various linguistic information described at different levels of granularity—have emerged. However, for many languages, large-scale LRs complying with a wide range of NLP tasks have not been developed so far as their manual building is highly time and effort consuming (and automatically extracted information is not of sufficient quality for deeper layers of the language description). As a result, the focus of the current computational lexicography has shifted from building new LRs to the integration and harmonization of the existing ones. The aim is to create extensive LRs on the basis of the current ones. Let us mention at least the project SemLink interlinking four prominent English LRs—PropBank¹, VerbNet², FrameNet³, and WordNet⁴ (Loper et al., 2007).

Mapping projects typically share the basic assumption that merging linguistic data performs on the level of lexical units (LUs) as prominent form-meaning units in a language. In this paper, we demonstrate that fully automatic linking of LUs represents a tricky task, especially in case of verbs with a higher number of LUs. As an illustrative example, we take two close LRs: the VALLEX⁵ and PDT-Vallex⁶ lexicons. We show that even though (i) both lexicons have been built within the same theoretical framework, (ii) they both are monolingual lexicons describing Czech and (iii) they both encode the same type of information, their automatic mapping is not straightforward.

The mapping task is usually preceded by converting merged LRs into a common format. To save such work, standardized representational formats have been proposed, e.g., Linguistic Annotation Framework (ISO:24612:2012), Morpho-syntactic Annotation Framework (ISO:24611:2012), Syntactic Annotation Framework (ISO:24615:2010), Semantic Annotation Framework (ISO:24617-1:2012, ISO:24617-2:2012). Some of them focus on the mapping of the existing formats, e.g., Lexical Markup Framework (ISO:24613) (Francopoulo et al., 2006). These common formats has already been applied in several projects: UBY, (Gurevych et al., 2012), BOOTT-Strep, (Quochi et al., 2008), KYOTO, (Vossen et al., 2010), etc.

However, our experiment on the linking of VALLEX and PDT-Vallex has suggested that when merging LRs, designing a common format represents a relatively easy part of the task in comparison with identifying pairs of corresponding LUs. In this paper, the major difficulties in the automatic identification of pairs of corresponding LUs are analyzed and the possibilities how to overcome them are proposed.

The paper is structured as follows. First, the two lexicons, VALLEX and PDT-Vallex, involved in the merging experiment are briefly characterized in Section 2. Second, the description of the proposed common format is introduced in Section 3. Third, the mapping procedure is throughly described and evaluated and main sources of mismatches are identified in Section 4.
2. Two Czech Valency Lexicons: PDT-Vallex and VALLEX

Let us introduce two LRs involved in our experiment: the PDT-Vallex and VALLEX lexicons.

PDT-Vallex, see esp. (Hajič et al., 2003) and (Urešová, 2011), has been built on the basis of the Prague Dependency Treebank 2.0 (PDT), see (Hajič et al., 2006), with the aim to ensure data consistency in PDT. It became an important resource for further linguistic research as well as for computational processing of the Czech language. It stores the information on valency frames (primarily) of Czech verbs that occur at least once in PDT. All valency frames in PDT-Vallex are linked with verb occurrences in PDT 2.0. The description of valency in PDT-Vallex stems from the Functional Generative Description, see esp. (Sgall et al., 1986), (Paněová, 1994).

The same theoretical framework is used also in VALLEX, see esp. (Žabokrtský and Lopatková, 2007). In contrast to PDT-Vallex, VALLEX aims at the complex description of valency behavior of selected Czech verbs in each of their senses (corresponding to LUs), i.e., at providing wider syntactic analysis of whole verb lexemes (not only those lexical units that appear in a particular corpus). The lexicon covers more than 98% of verb occurrences in the Czech National Corpus; however, corpus evidence is not provided in VALLEX.

Thus the linking of PDT-Vallex and VALLEX fortifies both resources involved: PDT-Vallex profits from complex syntactic information encoded in VALLEX and the latter obtains corpus evidence from PDT.

3. Common Format

As the involved lexicons, VALLEX and PDT-Vallex, are not based on any standardized format, converting them into a common format must precede their automatic linking.

As a preliminary step of this task, the structure of lexical entries of the lexicons was carefully compared, with emphasis put especially on the status of LUs. In both lexicons, the valency information is encoded in the form of valency frames, i.e., sequences of slots for obligatory and/or optional valency complementations comprising information on their (semantic) type and possible morphemic expressions.

However, the VALLEX lexicon is characterized by more complex lexical entries than PDT-Vallex: in VALLEX, each lexical entry stores the information on the whole lexeme. The lexeme is an abstract data structure associating lexical form(s) with LU(s):

(i) Different lexical forms, as aspectual counterparts and orthographic variants (e.g., namdhat‘inf, namoci/namoct‘pf ‘to strain’) are treated within a single lexeme; in such cases, a lexeme is represented by more than one lemma.

(ii) Each LU, roughly corresponding to a verb in a given meaning, is associated with a set of attributes—the valency frame encoding the information on valency behavior of the given LU represents the most important attribute; in addition, further syntactic information (on control, reflexivity, reciprocity etc.) is rendered there.

Unlike VALLEX, neither aspectual counterparts, nor orthographic variants are clustered together in PDT-Vallex. The lexical entry in PDT-Vallex consists of a set of valency frame(s) roughly corresponding to LU(s) associated with a verb lemma.

3.1. vallex.pml

Since the VALLEX and PDT-Vallex lexicons employ similar XML-based formats, there is no need to use a standardized format in this case. Instead, the Prague Markup Language (PML), see (Pajas and Štěpánek, 2009), an XML-based format close to both of them is used. This choice is supported by further advantages of PML: there is a query language and a search engine as well as a viewer and an editor available for any data encoded in PML.

The common vallex.pml, a PML-based format, is straightforwardly derived from the PDT-Vallex format. Then also VALLEX is converted into vallex.pml: each cluster of lemmas had to be split in order to create separate entries for the respective lemmas. The information on aspectual counterparts or orthographic variants and their corresponding frames are encoded by means of references. More detailed description can be found in (Bejček et al., 2010).

For our task, the vallex.pml format offers a functionality comparable to universal formats. Considering the costs of converting lexicons into a common format, the vallex.pml solution proved to be more efficient than using any of universal formats.

4. Mapping of VALLEX and PDT-Vallex

A mapping of LRs can rely on that linguistic information on LUs that is included in both resources. In principle, the more linguistic information of the same type is covered in both lexicons, the easier the mapping task is. However, the same linguistic information can be encoded in a different way even in similar resources. Thus, first of all, it is necessary to carefully compare annotation principles applied in both resources.

In case of VALLEX and PDT-Vallex, the following information on each LU identified as covered by both lexicons can be extracted:

(i) valency frame (4.1.1.),
(ii) a set of lemmas (4.1.2.),
(iii) reciprocity (4.1.3.), and
(iv) control (4.1.4.).

4.1. Automatic Mapping

Having the information on an individual LU of a given verb lemma extracted from one lexicon, it is confronted with the information extracted for all LU(s) of the given lemma from the other lexicon. A set of rules is applied automatically on each pair of confronted LUs. Then a Score representing the level of similarity for the respective pair is summed: each rule, as described in following Sections 4.1.1.–4.1.4., can contribute to the resulting Score. To put it differently, the Score is an overall sum of weights of rules applied to the following types of information on LUs.
### 4.1.1. Valency Frames

The application of each of proposed rules is conditioned by constraints on the form of VALLEX and PDT-Vallex valency frames—these constraints concern: number of valency complementations, their type, and their possible morphemic forms. For this purpose, two variants of a valency frame are taken into account: (i) the standard variant that comprises all obligatory and optional arguments and all obligatory adjuncts and (ii) the extended variant that consists of optional adjuncts in addition to the standard variant; both variants include morphemic forms. These pairs of manually written constraints (one constraint for VALLEX, one for PDT-Vallex) specify pairs of valency frames—these constraints concern: number of valency complementations, their type, and their possible morphemic forms. For this purpose, two variants of a valency frame are taken into account: (i) the standard variant that comprises all obligatory and optional arguments and all obligatory adjuncts and (ii) the extended variant that consists of optional adjuncts in addition to the standard variant; both variants include morphemic forms. These pairs of manually written constraints (one constraint for VALLEX, one for PDT-Vallex) specify pairs of valency frames that can represent the same LU, and as such they should be linked together. Each constraint gets its (manually estimated) weight which contributes to the Score (if the constraint is applied). A greedy algorithm is used whenever more constraints are satisfied, i.e., only the highest weight is added to the Score.

**Example:** Lexical units B and G in Figure 1—displaying automatic linking of the verb obracet ‘to turn’—are linked on the basis of the similarity of valency frames:

| VALLEX          | PDT-Vallex        |
|-----------------|-------------------|
| ACT(1) PAT(4) +^DIR() | ACT(1) PAT(4) EFF(v+4) |
| obracet vše vzhůru nohama / naruby | obracejí naše výroky v právý opak |
| ACT(1) PAT(4) EFF(k+3, na+4, v+4) | ACT(1) PAT(4) DPHR(naruby) |
| obracet vše naruby | obracet pozornost jinam |
| ACT(1) PAT(4) MANN() | ACT(1) PAT(4) DIR3(*) |
| obracel svůj vztek na dávného přítele | obracel skříň otočit. k věci ke zdí DIR3 |
| ACT(1) PAT(4) DPHR([pozornost,...].4) DIR3(*) | ACT(1) PAT(4) ADDR(na+4) |
| obracet pozornost / zájem / zřetel | obracela svůj vztek na dávného přítele |

Figure 1: The verb obracet ‘to turn’ in VALLEX and PDT-Vallex:
an illustrative example of the automatic mapping (the width of lines corresponds to their weights).

### 4.1.2. Lemmas

All autosemantic words used in an individual lexical unit—in gloss(es), example(s) and reciprocity example(s) documenting usages of the given LU, as well as words in valency frames of verbal multiword expressions (idioms and light verbs)—are extracted. The obtained word forms are lemmatized (and the verb itself, if it is used, is removed). Then the resulting sets of lemmas associated with the processed LU candidates for linking are compared. If the sets of lemmas have a non-trivial intersection, the Score is increased by the arithmetic mean of the ratios of shared lemmas to all obtained lemmas.

**Example:** Lexical units H and D in Figure 1 are linked thanks to an intersecting lemma: (i) from the extraction and lemmatization of the words from the example obracet vše naruby ‘to turn everything inside out’ and DPHR(naruby) from the valency frame H, the lemma naruby ‘inside out’ is obtained; (ii) the example obracel svůj vztek na dávného přítele in D is reduced and lemmatized to the following three lemmas vzhůru, noha, naruby. The only shared lemma is naruby ‘inside out’, thus the ratios are $\frac{1}{4}$ for H and $\frac{1}{4}$ for D. Then their arithmetic mean 0.67 is added to the Score.

### 4.1.3. Reciprocity

Additional information that can be obtained from both lexicons is reciprocity, i.e., the information on possible symmetric usages of two (or even three) valency complementations. In VALLEX, reciprocity is explicitly encoded in the form of VALLEX. For PDT-Vallex, reciprocity is explicitly encoded in the description of LUs: those valency complementations that can be expressed reciprocally are listed in the special attribute -rcp. Although the information on reciprocity is not explicitly recorded in PDT-Vallex, it can be extracted from the syntactic annotation of PDT—in cases where valency complementations are expressed reciprocally in the corpus data.
complementations of compared LUs exhibit reciprocity. In case that compared LUs exhibit reciprocity of some of their valency complementations, however, they do not match with respect to types of the given complementations, 0.5 is added to the Score. Further, a full match of reciprocally used valency complementations contributes 1 to the Score and a partial match gives 0.7 to the Score.

Example: The valency frames B and F from VALLEX are linked with the single frame G in PDT-Vallex, see Figure 1. There is no clue to decide whether both B and F or only one of them correspond(s) to G. In VALLEX, the valency frame B (and not the frame F) is marked with the attribute \(_{-rcp}\) providing the information on possible reciprocal usage of ACT and PAT. If there were an instance of the valency frame G in the PDT data with reciprocally used ACT and G would be facilitated.

4.1.4. Control

Finally, the information on control is also included in both mapped lexicons. The control applies to a certain type of verbs (verbs of control) that have one of their valency slots expressed by infinitive, see esp. (Davies and Dubinsky, 2004). The coreferential relations can be observed between a ‘controller’ and a ‘controllee’: the controllee is a syntactic element in the subject position of the infinitive (structurally excluded in the surface) and the controller is a co-indexed expression typically filling one of valency slots of the verb of control. Similarly as reciprocity, the information on control is explicitly recorded in VALLEX; although it is not explicitly encoded in PDT-Vallex, it can be obtained from the PDT data.

In sum, the pairs of matched LUs with the highest Score (a sum of the maximum constraint weight, lemmas intersection, and reciprocity) represent candidate LUs for the linking. In the second phase of the experiment, candidate LUs trustworthy enough to be mapped are selected from a huge sum of pairs of candidate LUs obtained from the first phase of this experiment, as it is described in the following Section 4.2.

4.2. Pruning the Graph

The whole mapping can be seen as adding edges into a bipartite graph: one group of vertices represents LUs in VALLEX and the other one represents LUs in PDT-Vallex. An edge between two vertices (from different groups) represents a candidate for the linking.

Note that (from the perspective of the graph theory) the whole task lies basically in searching for a subset of edges in a complete bipartite graph (under some linguistic conditions). However, the result does not need to be either an edge cover (as a LU may be omitted in one lexicon and as a consequence the corresponding LU should remain as an isolated vertex in the other lexicon), nor a matching (one LU in one lexicon may correspond to several LUs in the other lexicon). Even so, a perfect matching (i.e., one-to-one mapping between LUs) represents obviously the most satisfactory result, whenever the lexicons make it possible. The mapping procedure described above results in a biadjacency matrix of a bipartite graph, i.e., a matrix where the entry \(a_{ij}\) is the Score for linking LU \(i\) from VALLEX with LU \(j\) from PDT-Vallex. This matrix needs to be pruned: the aim is to retain only those edges that are trustworthy enough, i.e., the edges with weight implying a sufficient match of compared LUs. The aim of the pruning procedure is to keep those edges that allow us to achieve the highest possible recall with the highest possible precision. Thus the retained edges follow these two principles:

- each vertex should have positive degree, if possible (= disapprove isolated vertices: get close to an edge cover), and
- no vertex should have degree (much) bigger than one, if possible (= disapprove multiple edges for one vertex: get close to a (perfect) matching).

Table 1 summarizes the overall number of the automatically merged data.

4.3. Evaluation

4.3.1. Testing Data

We carried out a complementary experiment with manual mapping of a sample of 200 verb lemmas. First, 90 verb lemmas were randomly selected according to a number of their lexical units in VALLEX: Let \(v_i\) be the set of all verbs with \(i\) LUs; ten verbs were selected from each set \(v_i, i = 1 \ldots 9\). Second, 110 verbs were selected following the frequency distribution of verbs with respect to their complexity (= a number of their LUs), i.e., the proportion of verbs with \(i\) LUs in the sample data reflects the proportion of \(v_i\) in VALLEX.

4.3.2. Evaluation Measures

As for the complexity of the task, let us focus on the number of correct edges out of all the edges that can be assigned for the verbs from the sample data. There can be 2,721 edges between LUs of all 200 verbs (i.e., all the edges in the complete bipartite graphs for 200 verbs). Annotators A and B marked 529 and 493 edges as correct, respectively. Naturally, these edges are distributed unevenly among verbs with different numbers of LUs. Most edges (both potential and correct ones) are connected to verbs with nine LUs (approx. 700 potential and 70 correct ones) whereas most verbs in the lexicon has only one LU. That is the reason why the precision and the recall are counted separately for each verb. Then they are averaged with weights corresponding to the cardinality of a set \(|v_i|\) to which the verb belongs (a bigger weight for a larger set: e.g., \(v_1\) has a bigger weight as it contains more verbs than \(v_9\) even though verbs with 9 LUs contain much more edges in our annotated sample). It reflects the fact that we are more interested in the number of the correct matching of verbs than in the number of correctly paired LUs of these verbs.

4.3.3. Results

Two human annotators were asked to manually identify correct edges on 200 selected verb lemmas. The achieved inter-annotator agreement (IAA) is rather satisfactory, which implies that the mapping LUs is a feasible task for human annotators, see Table 2. The IAA for the first 90 verbs is similar to the one calculated on the second 110 verbs. It is implied by the measure used (see 4.3.2.).
results given separately for each set of verbs together: it is an average of the values \( v_i \) weighted by the \( |v_i| \) (see Section 4.3.2.).

The overall achieved precision and recall are satisfactory. However, the precision of more complex verbs, see especially rows \( v_7 \) and \( v_8 \), indicates that the mapping task is much harder for verbs with a higher number of LUs.

Let us focus on reciprocity and its possible contribution to the task. This phenomenon has turned up to be too sparse in the PDT data: only 493 cases of reciprocity of verbal valency complementations appear in the PDT data; these cases belong to 144 LUs represented by 130 verb lemmas. Since it represents 1.2% of all LUs in PDT-Vallex, it is obvious that it cannot improve the results considerably — the reciprocity information is used only for three out of 90 selected verbs.\(^9\)

Similarly, the information on control is sparse too (although it is not so sparse as reciprocity: in PDT 4,195 cases of control that correspond to 280 LUs represented by 240 verb lemmas occur). Our linguistic inquiry confirms that the information on control has only limited potential to improve our mapping procedure. If several LUs of a verb allow for control, the type of control is typically the same. Thus for the mapping procedure, the information on a type of control (provided by a type of the valency complementation that represents the ‘controller’ (see Section 4.1.4.)) is not so beneficial as the information on control itself. However, this information is already implied by the morphemic form of infinitive present in a valency frame. For this reason, we have not used the information on the types of control explicitly in the mapping algorithm.

Let us conclude. The evaluation of the automatic mapping of VALLEX and PDT-Vallex performed on 200 manually annotated verbs has suggested that the automatic mapping represents a tricky task, especially in case of verbs with a higher number of LUs, even in case of merging similar lexical resources.

### 4.3.4. Error Analysis

According to our observation, there are the following sources of mismatches between VALLEX and PDT-Vallex which can result in wrong edges, compare also with Figure 1.

(i) **Insufficient information on LUs.** Additional information on lexical units will be beneficial especially for am-

---

\(^9\)When comparing to the test with reciprocity ignored, one verb reaches better results, one verb reaches worse results and one verb has a correct mapping in both cases.
Table 3: The statistics on LUs for selected 200 verbs. The numbers are given for the manual annotation (annotators A and B) and for the automatic procedure (auto).

| Number of LUs in VALLEX | Precision | Recall | F-measure |
|-------------------------|-----------|--------|-----------|
| v1                      | 95        | 77     | 85        |
| v2                      | 84        | 72     | 77        |
| v3                      | 69        | 82     | 75        |
| v4                      | 66        | 75     | 70        |
| v5                      | 57        | 88     | 69        |
| v6                      | 47        | 83     | 60        |
| v7                      | 45        | 68     | 54        |
| v8                      | 40        | 73     | 52        |
| v9                      | 54        | 76     | 63        |
| Average weighted over all 200 verbs | 81 | 77 | 79 |

Table 4: The precision and recall of the automatic linking averaged over the set of 200 verbs.

5. Conclusion

To sum up, we have provided a detailed outline of the automatic procedure for mapping two close valency lexicons of Czech verbs, VALLEX and PDT-Vallex. We have focused on different types of information that can be exploited for the task. We have demonstrated that converting VALLEX and PDT-Vallex into a common data format has represented a relatively easy task in their automatic merging. In contrast, the automatic identification of corresponding lexical units has faced severe difficulties. The automatic mapping has been evaluated on a dataset of 200 manually linked verbs with the resulting precision of 81%. However, the precision of more complex verbs is much lower which indicates that the automatic mapping of LUs of complex verbs is a hard task which can rely only on phenomena covered in both mapped resources that are frequent and varied enough.

As to the future work, an edge threshold should be established. Only such edges between lexical units that have the Score above the threshold will be accepted as candidate LUs for the linking. The experiment revealed that the proposed automatic method allows a relatively high number of superfluous edges. The threshold could be the means how to remove them.

6. Acknowledgements

The research reported in this paper has been supported by the Czech Science Foundation GA ČR, grant No. GA P406/12/0557 and by grant COST CZ.LD14117 of the Ministry of Education, Youth and Sports of the Czech Republic.
This work has been using language resources developed and stored and distributed by the LINDAT/CLARIN project of the Ministry of Education, Youth and Sports of the Czech Republic (project LM2010013).

7. References

Eduard Bejček, Václava Kettnerová, and Markéta Lopatková. 2010. Advanced searching in the valency lexicons using PML-TQ search engine. In The Proceedings of the Text, Speech and Dialogue Conference, pages 51–58, Berlin/Heidelberg. Springer Verlag.

William Davies and Stanley Dubinsky. 2004. The grammar of raising and control. Blackwell, Malden, MA.

Gil Francopoulo, Monte George, Nicoletta Calzolari, Monica Monachini, Nuria Bel, and Mandy Pet Claudia Soria. 2006. Lexical markup framework (LMF). In The Proceedings of the International Conference on Language Resources and Evaluation – LREC 2006, pages 223–226, Genoa. LREC.

Iryna Gurevych, Judith Eckle-Kohler, Silvana Hartmann, Michael Matuschek, Christian M. Meyer, and Christian Wirth. 2012. UBY – A large-scale unified lexical-semantic resource based on LMF. In Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2012), pages 580–590, Avignon, France, April.

Jan Hajič, Jarmila Panevová, Zdeňka Urešová, Alevtina Běmová, Veronika Kolářová, and Petr Pajas. 2003. PDT-VALLEX: Creating a Large-Coverage Valency Lexicon for Treebank Annotation. In Proceedings of The Second Workshop on Treebanks and Linguistic Theories, volume Vol. 9, pages 57–68.

Jan Hajič, Eva Hajičová, Jarmila Panevová, Petr Sgall, Petr Pajas, Jan Štěpánek, Jiří Havelka, and Marie Mikulová. 2006. Prague Dependency Treebank 2.0. Linguistic Data Consortium, Philadelphia, PA, USA. http://hdl.handle.net/11858/00-097C-0000-0001-B098-5, LDC2006T01, ISBN 1-58563-370-4.

Edward Loper, Szu ting Yi, and Martha Palmer. 2007. Combining lexical resources: Mapping between PropBank and VerbNet. In Proceedings of the 7th International Workshop on Computational Semantics, pages 118–128, Tilburg, the Netherlands. Iceland Centre for Language Technology (ICLT).

Petr Pajas and Jan Štěpánek. 2009. System for querying syntactically annotated corpora. In The Proceedings of the ACL-IJCNLP, pages 33–36, Suntec, Singapore. Association for Computational Linguistics.

Jarmila Panevová. 1994. Valency Frames and the Meaning of the Sentence. In Philip A. Luelsdorff, editor, The Prague School of Structural and Functional Linguistics, pages 223–243. John Benjamins Publishing Company, Amsterdam, Philadelphia.

Valeria Quochi, Monica Monachini, Riccardo Del Gratta, and Nicoletta Calzolari. 2008. A lexicon for biology and bioinformatics: the BOOTStrep experience. In Nicoletta Calzolari, Khalid Choukri, Bente Mægaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, and Daniel Tapia, editors, Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC’08), Marrakech, Morocco. European Language Resources Association (ELRA).

Petr Sgall, Eva Hajičová, and Jarmila Panevová. 1986. The Meaning of the Sentence in Its Semantic and Pragmatic Aspects. Reidel, Dordrecht.

Zdeňka Urešová. 2011. Valence sloves v Pražském závislostním korpusu. Studies in Computational and Theoretical Linguistics. Institute of Formal and Applied Linguistics, Praha, Czechia. ISBN 978-80-904571-0-2.

P. Vossen, E. Agirre, F. Bond, W. Bosma, C. Fellbaum, A. Hicks, S. Hsieh, H. Isahara, Ch.Huang, K. Kanzaki, A. Marchetti, G. Rigau, F. Ronzano, R. Segersand, and M. Tesconi. 2010. KYOTO: a Wiki for Establishing Semantic Interoperability for Knowledge Sharing across Languages and Cultures. In E. Blanchard and D. Allard, editors, Handbook of Research on Culturally-Aware Information Technology: Perspectives and Models, pages 265–294. IGI Global.

Zdeněk Žabokrtský and Markéta Lopatková. 2007. Valency information in VALLEX 2.0: Logical structure of the lexicon. The Prague Bulletin of Mathematical Linguistics, (87):41–60. ISSN 0032-6585.