

Creating a Verb Synonym Lexicon Based on a Parallel Corpus

Zdeňka Urešová, Eva Fučíková, Eva Hajičová, Jan Hajič

Charles University, Faculty of Mathematics and Physics

Institute of Formal and Applied Linguistics

Malostranské nám. 25

11800 Prague 1, Czech Republic

{uresova,fucikova,hajicova,hajic}@ufal.mff.cuni.cz

Abstract

This paper presents the first findings of our recently started project of building a new lexical resource called CzEngClass, which consists of bilingual verbal synonym groups. In order to create such a resource, we explore semantic ‘equivalence’ of verb senses (across different verb lexemes) in a bilingual (Czech-English) setting by using translational context of real-world texts in a parallel, richly annotated dependency corpus. When grouping semantically equivalent verb senses into classes of synonyms, we focus on valency (arguments as deep dependents with morphosyntactic features relevant for surface dependencies) and its mapping to a set of semantic “roles” for verb arguments, common within one class. We argue that the existence of core argument mappings and certain adjunct mappings to a common set of semantic roles is a suitable criterion for a reasonable verb synonymy definition, possibly accompanied with additional contextual restrictions. By mid-2018, the first version of the lexicon called CzEngClass will be publicly available.

Keywords: Lexical Resource, Parallel Corpus, Semantics, Syntax, Synonymy, Valency

1. Introduction

While synonym lexicons, such as the Roget’s thesaurus (Associates, 1988),¹ or WordNet (Miller, 1995; Fellbaum, 1998) are well-known resources, used in both research and NLP applications, we believe that there is a significant gap in these resources (which concerns mainly verbs). We see two orthogonal problems: first, a granularity problem—current synonym classes (synsets, entries, frames, ...) as found in these resources are usually too broad (contain words which can be considered synonyms only in very specific contexts), or they use too fine-grained sense distinctions, such as in WordNet, which often distinguish too many and too close verb senses (Palmer et al., 2007); second, a coverage problem - simply there are not enough verbs covered or some verb senses are missing.²

In our project, we attempt to introduce criteria, based mainly on valency or predicate-argument structure (taking into account both semantic and morphosyntactic properties of verbs) to help define both the granularity of verb senses as well as the synonym classes themselves. Along with (Levin, 2013), we refer to predicate-argument structure as to those structures that involve the realization of lexical item’s arguments, including morphosyntactic properties that affect the morphosyntactic realization of arguments. This is similar to our valency approach, the FGD Valency Theory (FGDVT), first described in (Panevová, 1974). As (Levin, 2014) notes, the meaning of a verb may be characterized via the relations that its arguments bear to it and “Semantic Roles” (SRs) can be seen as labels for certain recurring predicate-argument relations.

In addition, we have chosen to use multilingual, rather than monolingual-only evidence to help with both the sense dis-

tinctions and the synonymy relation proper. We have started with Czech and English, since there are resources (both lexical and textual) that allow us to do so.

The goal of the project is thus to group verbs used as synonyms in Czech and English into (cross-lingual) synonym classes representing a cross-lingual meaning of the state or event expressed by the set of verbs assigned to that class. We call the resulting lexicon CzEngClass.

For the purpose of this work, we use the term “synonym” in the “loose” interpretation (Lyons, 1968), i.e., the necessary semantic equivalence takes also wider context into account. We proceed strictly “bottom-up”, i.e. from the corpus and existing lexical resources towards the new lexicon. The novel feature is the use of a richly annotated bilingual corpus (Sect. 2.) to achieve deeper insight into the usage of verbs (together with their arguments) in translation.

The results reported in our paper are based on a sample of 60 classes manually processed and entered into the CzEngClass lexicon. We also describe their linkage to additional existing resources, the relevant features of which are described, too (Sect. 3.). The tool used for creating this sample is described elsewhere in this volume (Urešová et al., 2018).

Our approach to synonymy, meaning and sense is captured in Section 4.. Section 5. deals with the design of CzEngclass lexicon. In Section 6., we exemplify the main (albeit first) findings. We comment on the CzEngClass’ criteria for grouping synonymous, based on samples that have been annotated and processed while creating the first entries. Section 7. summarizes our approach and points to future work.

2. Corpus Resources

For our work we use primarily the parallel Prague Czech-English Dependency Treebank 2.0 (PCEDT 2.0) (Hajič et al., 2012). This corpus contains the Wall Street Journal (WSJ) part of the Penn Treebank (Marcus et al., 1993) and its manual (human) Czech translations. Each language part is enhanced with a rich manual linguistic an-

¹<http://www.roget.org>

²It is left for future investigation whether other parts of speech, especially nouns, would be subject to similar problems, but now we concentrate on verbs since they are generally considered the core of every sentence (clause). We will extend our approach to nouns and adjectives later.

notation in The Prague Dependency Treebank (PDT 2.0) style (Hajič et al., 2006; Hajič et al., 2018), which is in turn based on the Functional Generative Dependency (FGD) framework (Sgall et al., 1986). The PDT annotation uses a “stratificational” (layered) approach containing multiple layers (morphology, surface syntax and deep syntax). The main annotated phenomena are (surface) dependency structure and (deep) syntactico-semantic labeling of predicate-argument structure. For the purpose of our study, it is important that the annotation contains interlinked surface dependency trees and deep syntactico-semantic (tectogrammatical) trees, both vertically and also horizontally across the two languages on sentence and node levels. At the deep layer, each verb node (occurrence) is additionally linked to the corresponding valency frame in the associated valency lexicons, PDT-Vallex and EngVallex (Sect. 3.), effectively providing also word sense labeling for all verb occurrences in the bilingual corpus.

3. Lexical Resources

PDT-Vallex (Urešová et al., 2014), (Urešová, 2011) is a Czech valency lexicon, manually created in a bottom-up way during the annotation of the PDT/PCEDT 2.0. Each entry in the lexicon contains a headword (lemma) with associated valency frames. A valency frame typically corresponds to one sense of the verb, even though for close verb senses³ and identical valency frames only one valency frame may exist in the lexicon. Each valency frame includes labeled valency frame members, or valency “slots” (i.e., ACT for Actor, PAT for Patient, ADDR for Addressee, etc.), semantic “obligatoriness” attribute, and subcategorization information, i.e., required surface form(s) of the individual valency frame members. Most valency frames include a note or an example explaining their meaning and usage. The version of PDT-Vallex used here contains 11,933 valency frames for 7,121 verbs.

EngVallex (Cinková et al., 2014), (Cinková, 2006) is a valency lexicon of English verbs created on the same principles as PDT-Vallex by an automatic conversion from PropBank frame files (Palmer et al., 2005) which was manually refined afterwards.⁴ EngVallex was used for the annotation of the English part of the PCEDT 2.0. Currently, it contains 7,148 valency frames for 4,337 verbs. For the most part, EngVallex does not contain explicitly formalized subcategorization information.

CzEngVallex (Urešová et al., 2015), (Urešová et al., 2016) is based on the treebank annotation of the PCEDT 2.0 corpus, covering about 86,000 aligned verbal pairs. It is a manually annotated Czech-English valency lexicon linking verbal entries of PDT-Vallex and EngVallex. Over 66% of English verbs and 72% of Czech verbs⁵ in the PCEDT 2.0 have a verbal translation covered by the CzEngVallex mapping. CzEngVallex builds links not only between corresponding verbal frames but also between corresponding verb arguments for each pair of verb senses, providing an

³For explanation of the term “sense”, as it is used in this paper, please see Sec. 4..

⁴EngVallex preserves most of the links to PropBank.

⁵The remaining pairings are verb-noun or verbs translated as structurally different constructions.

interlinked database of argument structures available for each verb and documenting a cross-lingual comparison of Czech and English valency behavior.

VALLEX⁶ (Lopatková et al., 2016) is closely related to PDT-Vallex because it is built on the same theoretical framework. This lexicon is much more elaborated, however it is not based on the PDT data.

Among other resources we use, there are **FrameNet** (Baker et al., 1998; Fillmore et al., 2003), **FrameNet+** (Pavlick et al., 2015), **VerbNet** (Schuler, 2006), **SemLink** (Palmer, 2009; Bonial et al., 2012), **PropBank** (Palmer et al., 2005) and **English WordNet**⁷ (Miller, 1995; Fellbaum, 1998) as well as **Czech WordNet** (Pala et al., 2011), (Pala and Smrž, 2004). These resources have been used for an initial set of semantic roles (taken mostly from FrameNet and VerbNet),⁸ and their entries will be referred to explicitly from all the corresponding entries in the CzEngClass lexicon, if possible to the exact lexical units/frames/sense groups/synsets.

4. Synonymy

Before we address synonymy, we make a short digression to the term “meaning” and term “sense” as we interpret them in our work, due to sometimes wildly different understandings of these terms.

We understand the term “sense” in the same way as e.g., (Hofmann, 1993) saying “when a form has several different concepts associated with it, we sometimes call them different senses or readings of the word.” (Wierzbicka, 1996) has the same approach, explaining the term “sense” on four different senses of the word *spring*. We also differentiate a single verb (type, lemma, possibly multiword) into one or more “senses,” represented by its valency frame; in our lexical valency resources (PDT-Vallex and EngVallex) the individual senses (i.e., the valency frames) of a verb are technically represented by a unique ID. The term (lexical) “meaning” is understood here with regard to a context, i.e., syntactic and semantic surroundings of the word; similarly to Wittgenstein’s understanding that “the meaning of an expression is a function of its use in a particular context” (Frawley, 1992). In our use of the terms “meaning” and “sense”, the following holds:

- “sense” is used only to distinguish different meanings within one verb type (lemma), e.g., the verb *leave* has (at least) two senses, *leave sth somewhere* (leave book on the table) and *leave someplace [for some other place]* (leave Paris [for London]),
- “meaning” is not applied to verbs (lemmas, verb types), but only to their distinguished senses (lexical units), and can be compared across such units: e.g., *leave* in the sense *leave someplace [for some other place]* has similar meaning to *depart* in its sense *depart from somewhere*,
- consequently, two senses of the same verb can never be totally equivalent, i.e., they never have the same

⁶<http://ufal.mff.cuni.cz/vallex/3.0/>

⁷<https://wordnet.princeton.edu/>

⁸The current theories of SRs are reviewed in more detail in (Levin and Hovav, 2005).

meaning (they would not be separated if they were equivalent),

- meanings can thus only be (non-trivially) compared *across* verbs (more precisely, across lexical units defined for (“within”) different verbs).

In the working definition for establishing the CzEngClass entries, we use “**contextually-based**” synonymy. For two verbs (verb senses) to be considered contextually synonymous, and therefore be members of the same class, they have to convey the same or similar meaning, both monolingually and cross-lingually, and they must share the same Semantic Roles (SRs), albeit they can be expressed by different morphosyntactic realizations as well as subject to additional restrictions.

5. Structure of CzEngClass Lexicon

For our study, the following lexicon structure (Fig. 1) has been designed.

The CzEngClass lexicon builds upon the existing resources, as described above: CzEngVallex, PDT-Vallex and Eng-Vallex and the PCEDT parallel corpus. In addition, the other lexicons listed (VALLEX, FrameNet, VerbNet, Prop-Bank and WordNet(s)) are used as additional sources, and links will be kept between their entries and the CzEngClass entries.

At the core of the CzEngClass lexicon, there are Synonym Classes, which are, for the purpose of this project, defined as (multilingual, or rather cross-lingual)⁹ groups of verb senses (of different words) that have the same meaning *and* the arguments of which can be mapped to a common set of SRs (cf. the purple boxes Agent, Item, Change in Fig. 1). SRs have been reviewed (with FrameNet’s core roles for the linked-to entry(ies) providing inspiration) and determined for all the members of the group, and mapped to arguments from EngVallex (and PDT-Vallex for Czech verbs), using also the pairings from the CzEngVallex lexicon, which is in turn linked to the PCEDT parallel Czech-English corpus (lower part of Fig. 1).

6. CzEngClass Examples

We capture the common background information of one synonym class realized through a set of common SRs in one “frame,” called SynSemFrame (synonym semantic frame). We present two examples here.

In the relatively straightforward synonym verb class in Table 1, exemplified in the three sentences below, we consider all verbs to be synonyms and group them in the SynSem-Frame COMPLAIN, since the valency pattern of the source verbs *complain*, *gripe*, *grumble* correlates (for almost all arguments) 1:1 with the valency pattern of the translational equivalent, i.e., verb *stěžovat si* (lit. *complain*). The exception is the Czech verb itself, where both ADDR and LOC can be mapped to the role Addressee. The reason is, however, simply the conventions applied in the FGDVT to cases where the surface expression is location rather than (= in place of) a (true) addressee (as an animate agent); typically, this happens for offices (*to the clerk.ADDR* vs. *at the court*

office.LOC), government seats (*to the governor.ADDR* vs. *in Annapolis.LOC*), etc. Consequently, in all these cases, functors from the valency lexicons and the deep dependencies can easily be mapped to a common set of SRs, taken from FrameNet in this case.

Examples (Czech translations double as examples for the Czech verb *stěžovat si* in the class):

En: *Mrs. Yeargin.Complainer/ACT never complained to school officials.Addressee/ADDR [that the standardized test was unfair].Complaint/PAT*

Cz: *Yearginová.Complainer/ACT si nikdy na školském úřadu.Location/LOC nestěžovala na nespravedlnost.Complaint/PAT standardizovaných testů.*

En: [*“For \$10 million, you can move \$100 million of stocks.”*].Complaint/PAT *a specialist.Complainer/ACT ... gripes.* NULL.Addressee/ADDR

Cz: [*“Za 10 milionů dolarů můžete přesunout akcie za 100 milionů dolarů.”*].Complaint/PAT *stěžuje si ... odborník..Complainer/ACT* NULL.Addressee/ADDR

En: *Soviet consumers.Complainer/ACT grumble at the exorbitant black-market prices.Complaint/PAT* NULL.Addressee/ADDR

Cz: *Sovětsští spotřebitelé.Complainer/ACT si stěžují na nehorázné ceny.Complaint/PAT za takové zboží na černém trhu.* NULL.Addressee/ADDR

	Roles		
	Complainer	Addressee	Complaint
stěžovat si	ACT	ADDR,LOC	PAT
complain	ACT	ADDR	PAT
gripe	ACT	ADDR	PAT
grumble	ACT	ADDR	PAT

Table 1: Mappings for COMPLAIN class

Table 2 shows the class OFFER, with additional complex examples. While the verbs *offer*, *bid*, *proffer* and *tender*, as well as the Czech verb *nabídnout* do not pose any problem in mapping their valency arguments to the four SRs associated with this class, the other aligned translations, described below, are more complex.

The verb *extend* (in the sense “extend an offer”) seems not to fit the pattern; without the Entity Offered/PAT being a noun phrase meaning exactly what the class is about, namely *offers*, *bids*, *aid*, etc., its meaning is more similar to “hand over” than to “offer”. However, looking up all the possible deep objects to this meaning of *extend*, it is clear that it cannot be complemented by a direct object *not* being an *offer*, *bid* or something similar. If such a noun is further modified, one of its dependents might describe the Entity Offered (e.g., *employment: employment offer* or *offer of employment*). If it is not present, this construction simply says that there was an *offer* while not specifying anything more specific about it, unless the word *offer* or its equivalent is left out and the Entity Offered is specified as a dependent directly on the verb *extend* itself, as in *the banks ... extended [company] up to \$90 million in revolving loans*. Therefore, both the PAT as well as the restrictive attribute of the PAT (denoted as PAT(RSTR) can be the Entity Offered.

The verbal phrase *make available* is also used as a translation of the Czech verb *nabídnout*. In this case, *make* behaves almost as a light verb, which typically keeps some

⁹For the time being, bilingual: in Czech and English.

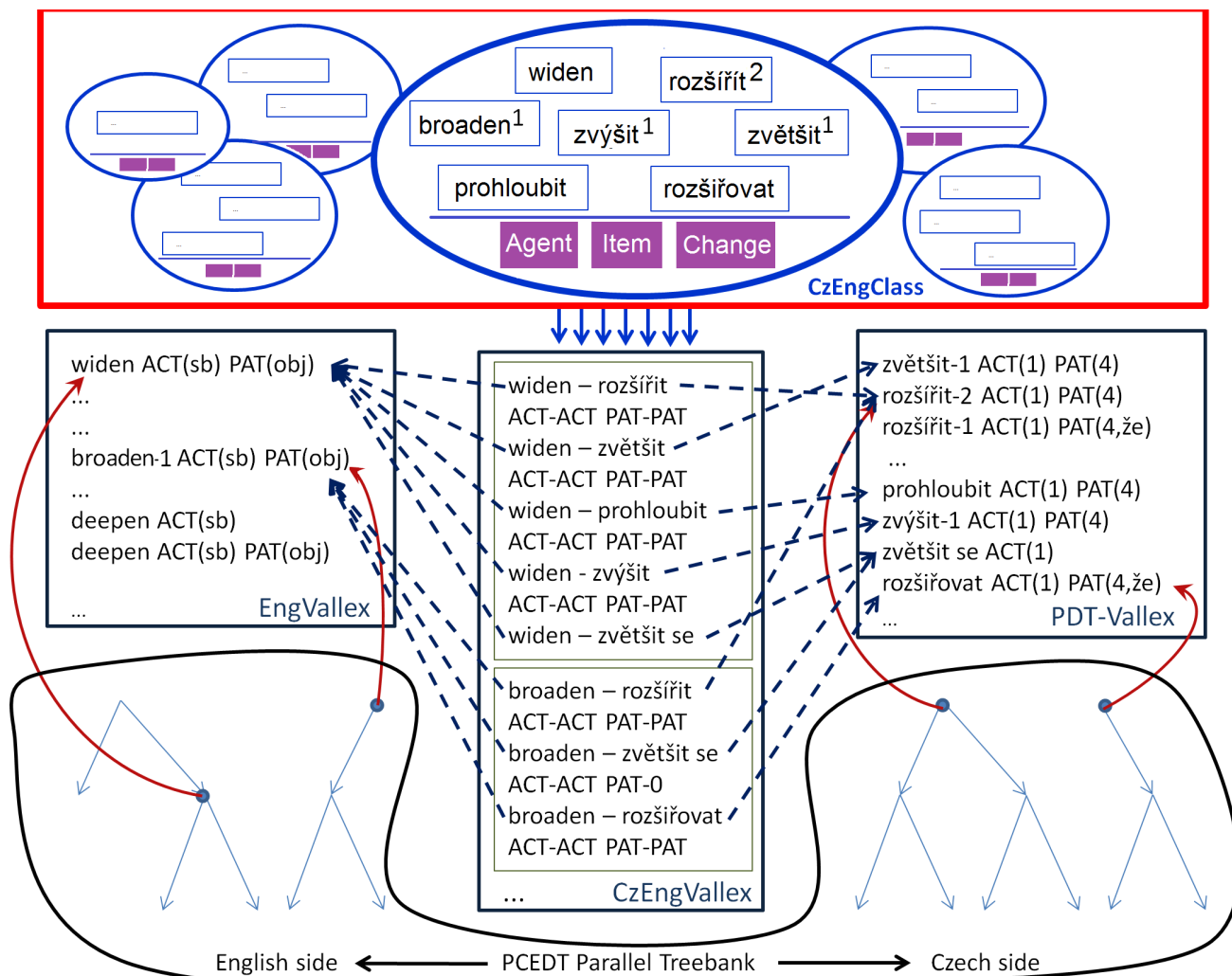


Figure 1: CzEngClass lexicon & other resources

arguments (mostly ACT) and “pushes” some other ones to the nominal part. In this case, the situation is similar, without *make* being considered a light verb (otherwise, a different functor would have been used for *available*). From the example below, it is clear that the translation was quite adequate and thus *make available* should be kept in the synonym class OFFER (see the deep dependency annotation of this example captured in Figure 2) where the Recipient is annotated with the relation BEN as a dependent on *available*, denoted as EFF(BEN) in Table 2:

En: *Japanese researchers.Offerer/ACT ... have made available.EFF three possible cures.Entity Offered/PAT to American researchers.Recipient/BEN ...*

Cz: *Japonští výzkumníci nabídli ... americkým výzkumníkům tři možné léčebné postupy...*

Last translation used for *nabídnout* was *place* (a value), e.g., in the sentence:

En: *... buyers who place the highest value on them.*

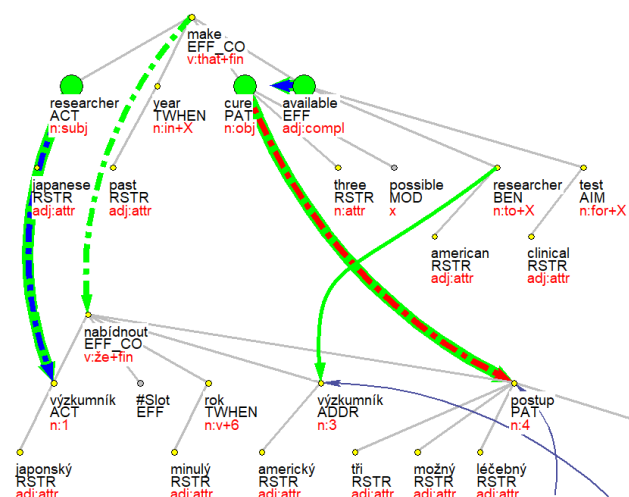
Cz: *...kupujícím, kteří za ně nabídnou nejvyšší hodnotu.*

However, this is considered pure nominalization (other sentences have been found where *offer*, *bid*, etc. as a head of a noun phrase is aligned with a Czech content verb), even if there was no other equivalent word for *place* in the Czech

translation. This is an example where the translated text probably does not have the same effect on the target reader: the translator simply added content to a sentence based on her/his understanding of the context. Therefore, the verb *place* was, at present, not included in the OFFER synonym class. This also demonstrates the difference between *interlingual synonymy* (which defines the CzEngClass lexicon) and *translational equivalence* (not necessarily represented).

	Roles			
	Offerer	Recipient	Entity Offered	Entity Received
nabídnout	ACT	ADDR	PAT	EFF
offer	ACT	ADDR	PAT	EFF
bid	ACT	ADDR	PAT	EFF
proffer	ACT	ADDR	PAT	SUBS
tender	ACT	ADDR	PAT	SUBS
extend	ACT	ADDR	PAT, PAT(RSTR)	SUBS
make available	ACT	EFF(BEN) Restriction: EFF[EXCHANGE]	PAT	SUBS

Table 2: Mapping for OFFER class



En: Japanese researchers... have made available three possible cures to American researchers.
 Cz: Japonští výzkumníci nabídli ... americkým výzkumníkům tři možné léčebné postupy.

Figure 2: Dependency tree annotations for English sentence rooted in *make available* and its Czech translation.

7. Conclusions and Future Work

The CzEngClass lexicon is a contribution to a set of lexical resources important for many NLP tasks, such as event detection and linking, semantic relation extraction, etc. While these tasks often use unlabeled data for training, there is always a need for human-annotated data for evaluation, tuning, etc. Importantly, we believe that such a resource is currently missing in the offerings of lexical resources. In addition, it contains links to all the relevant related resources, such as VALLEX, FrameNet, VerbNet, PropBank and WordNet, making it suitable for comparative studies. We have based our study on the assumption that mapping syntactic structure (verb arguments) of verbs to semantic roles (with some added restrictions) helps us discover their meaning affinity through which we can group these verbs into synonym classes. To help avoid single-language bias, we have used a parallel corpus and the alignments of verbs, previously manually checked and extracted. In such a corpus, aligned translations of a single verb should in principle be, in most cases, synonymous. The resulting sets of cross-lingual synonyms including the explicit mappings between the FGDVT valency functors and the semantic roles assigned to each synonym set are captured in the CzEngClass lexicon. So far, the roles used in FrameNet and VerbNet seem to form a suitable set for each synonym class and the mapping. However some adjustments, e.g., merging the roles, adding non-core roles, have to be made. Once the lexicon reaches reasonable size (approx. 500 classes), we will compare the results with automated synonym discovery methods, such as (van der Plas et al., 2011; van der Plas et al., 2014), either using Deep Learning (looking, e.g., at embeddings based on argument-role mapping) or other previously well-researched methods, such as the LDA which has been already used for Czech, e.g., in (Materna, 2012).

Finally, we plan to publish the resulting CzEngClass lexicon as an open source dataset.

Acknowledgments

This work has been supported by the grant No. GA17-07313S of the Grant Agency of the Czech Republic, and it uses resources hosted by the LINDAT/CLARIN Research Infrastructure, projects No. LM2015071 and CZ.02.1.01/0.0/0.0/16_013/0001781, supported by the Ministry of Education of the Czech Republic. The last co-author has also been supported in part by the project VIA-DAT, No. DG16P02R019, of the Ministry of Culture of the Czech Republic.

8. Bibliographical References

Associates, P. S. I. . (1988). *Roget's thesaurus of synonyms and antonyms*. P.S.I. & Associates.

Baker, C. F., Fillmore, C. J., and Lowe, J. B. (1998). 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. Association for Computational Linguistics.

Bonial, C., Feely, W., Hwang, J. D., and Palmer, M. (2012). Empirically Validating VerbNet Using SemLink. In *Seventh Joint ACL-ISO Workshop on Interoperable Semantic Annotation*, Istanbul, Turkey, May.

Cinková, S. (2006). From PropBank to EngValLex: adapting the PropBank-Lexicon to the valency theory of the functional generative description. In *Proceedings of LREC 2006, Genova, Italy*.

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. Language, Speech, and Communication. MIT Press, Cambridge, MA.

Fillmore, C. J., Johnson, C. R., and L.Petruck, M. R. (2003). Background to FrameNet: FrameNet and Frame Semantics. *International Journal of Lexicography*, 16(3):235–250.

Frawley, W. (1992). *Linguistic Semantics*. Lawrence Erlbaum, Hillsdale.

Hofmann, T. R. (1993). *Realms of Meaning. An Introduction to Semantics*. Longman, Oxford.

Levin, B. and Hovav, M. R. (2005). *Argument realization*. Cambridge Univ. Press, Cambridge.

Levin, B. (2013). Argument Structure. *Oxford Bibliographies in Linguistics*.

Levin, B. (2014). Semantic roles. *Oxford Bibliographies in Linguistics*.

Lopatková, M., Kettnerová, V., Bejček, E., Vernerová, A., and Žabokrtský, Z. (2016). *Valenční slovník českých sloves VALLEX*. Nakladatelství Karolinum, Praha, Czechia.

Lyons, J. (1968). *Introduction to Theoretical Linguistics*. Cambridge University Press.

Marcus, M. P., Santorini, B., and Marcinkiewicz, M. A. (1993). Building a Large Annotated Corpus of English: The Penn Treebank. *COMPUTATIONAL LINGUISTICS*, 19(2):313–330.

- Materna, J. (2012). LDA-Frames: An Unsupervised Approach to Generating Semantic Frames. In *Proceedings of the 13th International Conference on Computational Linguistics and Intelligent Text Processing - Volume Part I, CICLing'12*, pages 376–387, Berlin, Heidelberg. Springer-Verlag.
- Miller, G. A. (1995). WordNet: A Lexical Database for English. *Commun. ACM*, 38(11):39–41, November.
- Pala, K. and Smrž, P. (2004). Building Czech Wordnet. *Romanian Journal of Information Science and Technology*, 7(1-2):79–88.
- Palmer, M., Gildea, D., and Kingsbury, P. (2005). The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31(1):71–106, March.
- Palmer, M., Dang, H., and Fellbaum, C. (2007). Making Fine-grained and Coarse-grained sense distinctions, both manually and automatically. *Journal of Natural Language Engineering*, 13(2):137–163.
- Palmer, M. (2009). Semlink: Linking PropBank, VerbNet and FrameNet. In *Proceedings of the Generative Lexicon Conference*, page 9–15.
- Panevová, J. (1974). On verbal Frames in Functional Generative Description. *Prague Bulletin of Mathematical Linguistics*, 22:3–40.
- Pavlick, E., Wolfe, T., Rastogi, P., Callison-Burch, C., Dredze, M., and Van Durme, B. (2015). Framenet+: Fast paraphrastic tripling of framenet. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 408–413, Beijing, China, July. Association for Computational Linguistics.
- Schuler, K. K. (2006). *VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon*. Ph.D. thesis, University of Pennsylvania.
- Sgall, P., Hajičová, E., and Panevová, J. (1986). *The meaning of the sentence in its semantic and pragmatic aspects*. D. Reidel, Dordrecht.
- Urešová, Z., Fučíková, E., and Šindlerová, J. (2016). CzEngVallex: a bilingual Czech-English valency lexicon. *The Prague Bulletin of Mathematical Linguistics*, 105:17–50.
- Urešová, Z., Fučíková, E., Hajičová, E., and Hajič, J. (2018). Tools for Building an Interlinked Synonym Lexicon Network. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC'18)*, Miyazaki, Japan, May. European Language Resources Association (ELRA).
- Urešová, Z. (2011). *Valenční slovník Pražského závislostního korpusu (PDT-Vallex)*. Studies in Computational and Theoretical Linguistics. Ústav formální a aplikované lingvistiky, Praha, Czechia.
- van der Plas, L., Merlo, P., and Henderson, J. (2011). Scaling up cross-lingual semantic annotation transfer. In *Proceedings of ACL/HLT*, pages 299–304, Portland, US.
- van der Plas, L., Apidianaki, M., and Chen, C. (2014). Global methods for cross-lingual semantic role and predicate labelling. In *Proceedings of the 25th International Conference on Computational Linguistics (COLING)*, pages 1279–1290, Dublin, Ireland.
- Wierzbicka, A. (1996). *Semantics: primes and universals*. Oxford Univ. Press, Oxford & New York.

9. Language Resource References

- Cinková, S., Fučíková, E., Šindlerová, J., and Hajič, J. (2014). *EngVallex - English Valency Lexicon*. LINDAT/CLARIN PID: <http://hdl.handle.net/11858/00-097C-0000-0023-4337-2>.
- Hajič, J., Hajičová, E., Panevová, J., Sgall, P., Cinková, S., Fučíková, E., Mikulová, M., Pajas, P., Popelka, J., Semecký, J., Šindlerová, J., Štěpánek, J., Toman, J., Urešová, Z., and Žabokrtský, Z. (2012). *Prague Czech-English Dependency Treebank 2.0*, <https://catalog.ldc.upenn.edu/LDC2004T25>. Institute of Formal and Applied Linguistics, LINDAT/CLARIN PID: <http://hdl.handle.net/11858/00-097C-0000-0015-8DAF-4>.
- Hajič, J., Bejček, E., Bémová, A., Buráňová, E., Hajičová, E., Havelka, J., Homola, P., Kárník, J., Kettnerová, V., Klyueva, N., Kolářová, V., Kučová, L., Lopatková, M., Mikulová, M., Mírovský, J., Nedoluzhko, A., Pajas, P., Panevová, J., Poláková, L., Rysová, M., Sgall, P., Spoustová, J., Straňák, P., Synková, P., Ševčíková, M., Štěpánek, J., Urešová, Z., Vidová Hladká, B., Zeman, D., Zikánová, Š., and Žabokrtský, Z. (2018). *Prague Dependency Treebank 3.5*. Institute of Formal and Applied Linguistics, LINDAT/CLARIN, Charles University, LINDAT/CLARIN PID: <http://hdl.handle.net/11234/1-2621>.
- Hajič, J., Panevová, J., Hajičová, E., Sgall, P., Pajas, P., Štěpánek, J., Havelka, J., Mikulová, M., Žabokrtský, Z., Ševčíková Razímová, M., and Urešová, Z. (2006). *Prague Dependency Treebank 2.0*. LDC, number LDC2006T01, LINDAT/CLARIN PID: <http://hdl.handle.net/11858/00-097C-0000-0015-8DAF-4>.
- Pala, K., Čapek, T., Zajíčková, B., Bartůšková, D., Kulková, K., Hoffmannová, P., Bejček, E., Straňák, P., and Hajič, J. (2011). *Czech WordNet 1.9 PDT*. LINDAT/CLARIN PID: <http://hdl.handle.net/11858/00-097C-0000-0001-4880-3>.
- Urešová, Z., Štěpánek, J., Hajič, J., Panevová, J., and Mikulová, M. (2014). *PDT-Vallex: Czech Valency lexicon linked to treebanks*. LINDAT/CLARIN PID: <http://hdl.handle.net/11858/00-097C-0000-0023-4338-F>.
- Urešová, Z., Fučíková, E., Hajič, J., and Šindlerová, J. (2015). *CzEngVallex - Czech English Valency Lexicon*. Charles University in Prague, LINDAT/CLARIN PID: <http://hdl.handle.net/11234/1-1512>.