Using Parallel Texts and Lexicons for Verbal Word Sense Disambiguation

Ondřej Dušek Eva Fučíková Jan Hajič Martin Popel Jana Šindlerová Zdeňka Urešová

Charles University in Prague, Faculty of Mathematics and Physics

Institute of Formal and Applied Linguistics

Malostranské nám. 25

11800 Prague 1, Czech Republic

`{odusek,fucikova,hajic,popel,sindlerova,uresova}@ufal.mff.cuni.cz`

Abstract

We present a system for verbal Word Sense Disambiguation (WSD) that is able to exploit additional information from parallel texts and lexicons. It is an extension of our previous WSD method (Dušek et al., 2014), which gave promising results but used only monolingual features. In the follow-up work described here, we have explored two additional ideas: using English-Czech bilingual resources (as features only – the task itself remains a monolingual WSD task), and using a "hybrid" approach, adding features extracted both from a parallel corpus and from manually aligned bilingual valency lexicon entries, which contain subcategorization information. Albeit not all types of features proved useful, both ideas and additions have led to significant improvements for both languages explored.

1 Introduction

Using parallel data for Word Sense Disambiguation (WSD) is as old as Statistical Machine Translation (SMT): Brown et al. (1992) analyze texts in both languages before the IBM SMT models are trained and used, including WSD driven purely by translation equivalents.[1] A combination of parallel texts and lexicons also proved useful for SMT at the time (Brown et al., 1993). In our previous experiments (Dušek et al., 2014), we have shown that WSD based on a manually created valency lexicon (for verbs) can achieve encouraging results. Combining the above ideas and previous findings with parallel data and a manually created bilingual valency lexicon, we have moved to add bilingual

features to improve on the previous results on the verbal WSD task. In addition, we have opted for a new machine learning system, the Vowpal Wabbit toolkit (Langford et al., 2007).[2]

In Section 2, we present the annotation framework and the lexicons used throughout this paper. Section 3 describes our experiments, Section 4 summarizes relevant previous works and Section 5 concludes the paper.

2 Verbal word senses in valency frames

2.1 Prague dependency treebanks and valency

The Prague Dependency Treebank (PDT 2.0/2.5) (Hajič et al., 2006) contains Czech texts with rich annotation.[3] Its annotation scheme is based on the formal framework called Functional Generative Description (FGD) (Sgall et al., 1986), which is dependency-based with a "stratificational" (layered) approach: The annotation contains interlinked surface dependency trees and deep syntactic/semantic (*tectogrammatical*) trees, where nodes stand for concepts rather than words. The notion of valency in the FGD is one of the core concepts on the deep layer; for the purpose of our experiments, it is important that the deep layer links each verb node (occurrence) to the corresponding valency frame in the associated valency lexicon, effectively providing verbal word sense labeling.

The parallel Prague Czech-English Dependency Treebank 2.0 (PCEDT 2.0) (Hajič et al., 2012) has been annotated using the same principles as the PDT, providing us with manually disambiguated verb senses on both the Czech and the English side. The texts are disjoint from the PDT; PCEDT contains the Wall Street Journal (WSJ) part of the Penn Treebank (Marcus et al., 1993) and its

---

[1]Given the "automatic" nature of the word senses so derived, no figures on the WSD accuracy within the IBM Candide SMT system had been given in the Brown et al. (1992) paper.

[2]`http://hunch.net/~vw`

[3]`http://ufal.mff.cuni.cz/pdt2.0`

radit[2] `ACT(1) PAT(4;k+3;aby) ADDR(3)`

help[1] `ACT() PAT() ADDR()`

Figure 1: Valency frame examples from PDT-Vallex and EngVallex (Czech *radit* = 'give advice, help').

translation into Czech. Sentences have been manually aligned during the human translation process, and words have been then aligned automatically using GIZA++ (Och and Ney, 2003). We have used valency frame annotation (and other features) of the PCEDT 2.0 in our previous work; however, billingual alignment information has not been used before.

2.2 Valency lexicons

PDT-Vallex[4] (Hajič et al., 2003; Urešová, 2011) is a valency lexicon of Czech verbs (and nouns), manually created during the annotation of the PDT/PCEDT 2.0.

Each entry in the lexicon contains a headword (lemma), according to which the valency frames (i.e., senses) are grouped. Each valency frame includes the valency frame members and the following information for each of them (see Fig. 1):

- its function label, such as `ACT`, `PAT`, `ADDR`, `EFF`, `ORIG`, `TWHEN`, `LOC`, `CAUS` (actor, patient, addressee, effect, origin, time, location, cause),[5]

- its semantic "obligatoriness" attribute,

- subcategorization: its required surface form(s) using morphosyntactic and lexical constraints.

Most valency frames are further accompanied by a note or an example which explains their meaning and usage. The version of PDT-Vallex used here contains 11,933 valency frames for 7,121 verbs.

EngVallex[6] (Cinková, 2006) is a valency lexicon of English verbs based also on the FGD framework, created by an automatic conversion from PropBank frame files (Palmer et al., 2005) and subsequent manual refinement.[7] EngVallex was used for the annotation of the English part of the
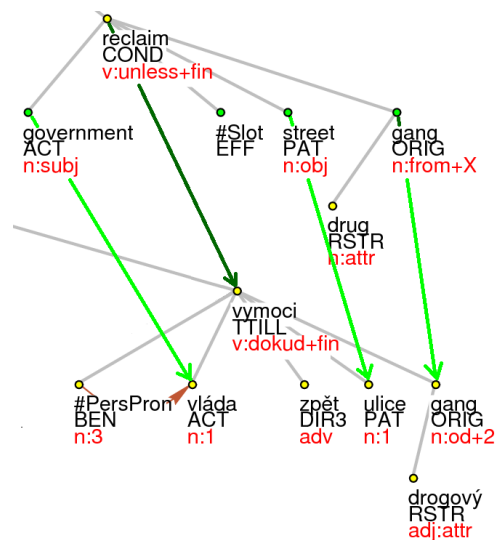
Figure 2: PCEDT trees aligned using the CzEng-Vallex mapping

PCEDT 2.0. Currently, it contains 7,148 valency frames for 4,337 verbs. EngVallex does not contain the explicitly formalized subcategorization information.

2.3 CzEngVallex: Valency lexicon mapping

CzEngVallex (Urešová et al., 2015a; Urešová et al., 2015b) is a manually annotated Czech-English valency lexicon linking the Czech and English valency lexicons, PDT-Vallex and EngVallex. It contains 19,916 frame (verb sense) pairs. CzEng-Vallex builds links not only between corresponding frames but also between corresponding verb arguments. This lexicon thus provides an interlinked database of argument structures for each verb and enables cross-lingual comparison of valency. As such (together with the parallel corpora to which it is linked), it aims to serve as a resource for cross-language linguistic research. Its primary purpose is linguistic and translatology research.

CzEngVallex is based on the treebank annotation of the PCEDT 2.0, covering about 86.000 aligned verbal pairs in it. Fig. 2 shows an example alignment between the English verb *reclaim* (sense: *get back by force*) and its arguments. 3,288 EngVallex and 4,192 PDT-Vallex verbs occur interlinked in the PCEDT 2.0 at least once, amounting to 4,967 and 6,776 different senses, respectively. Token-wise, over 66% of English verbs and 72% of Czech verbs in the PCEDT 2.0 have a verbal translation covered by the CzEngVallex mapping.

3 Verbal WSD experiments

We are focusing here on measuring the influence of parallel features on the WSD performance. In order to compare our results to our previous work, we use the same training/testing data split, i.e., PCEDT 2.0 Sections 02–21 as training data, Section 24 as development data, and Section 23 as evaluation data, and start from the same set of monolingual features. We also include Czech monolingual results on PDT 2.5 (default data split) for comparison. Unlike our previous work using LibLINEAR logistic regression (Fan et al., 2008), we apply Vowpal Wabbit (Langford et al., 2007) for classification.

Note that the input to our WSD system is plain text without any annotation, and we only use the gold verb senses from PCEDT/PDT to train the system. All required annotation for features as well as word alignment for parallel texts is performed automatically.

3.1 Monolingual experiments

We applied the one-against-all cost-sensitive setting of the Vowpal Wabbit linear classifier with label-dependent features.[8] Feature values are combined with a candidate sense label from the valency lexicon. If a verb was unseen in the training data or is sense-unambiguous, we used the first or only sense from the lexicon instead of the classifier.[9]

The training data were automatically analyzed from plain word forms up to the PDT/PCEDT-style deep layer using analysis pipelines implemented in the Treex NLP framework (Popel and Žabokrtský, 2010).[10] The gold-standard sense labels were then projected onto the automatic annotation. This emulates the real-world scenario where no gold-standard annotation is available.

The monolingual feature set of Dušek et al.

(2014) includes most attributes found in the PCEDT annotation scheme:

- the surface word form of the lexical verb and all its auxiliaries,
- their part-of-speech and morphological attributes,
- formemes – compact labels capturing morphosyntactic properties of deep nodes (e.g., `v:fin` for a finite verb, `v:because+fin` for a finite verb governed by a subordinating conjunction, `v:in+ger` for a gerund governed by a preposition),[11]
- syntactic labels given by the dependency parser,
- all of the above properties found in the neighborhood of the verbal deep node (parent, children, siblings, nodes adjacent in the word order).

3.2 Using word alignment

This scenario keeps all the previous settings and includes one more feature type – the translated lemma from the other language as projected through word alignment. This feature is also concatenated with the candidate sense label from the lexicon. We reuse the automatic GIZA++ word alignment from PCEDT 2.0 and project it to the automatic deep layer annotation using rules implemented in the Treex framework.

Since GIZA++ alignment can be obtained in an unsupervised fashion, this still corresponds to a scenario where no previous word alignment is available. Our experience from the CzEngVallex project (see Section 2.3), where GIZA++ alignment links were corrected manually, suggests that the automatic alignment is quite reliable for verbs (less than 1% of alignment links leading from verbs required correction).

3.3 Combining alignment with valency lexicon mapping

This setting includes the aligned lemma features and adds a single binary feature that combines parallel data information from PCEDT 2.0 with the CzEngVallex valency lexicon mapping (see Section 2.3).

For each verbal sense from the PDT-Vallex and EngVallex lexicons, we created a list of all lemmas from the other language corresponding to senses connected to this sense through the CzEngVallex

---

[8]Based on preliminary experiments on development data sets, we used the following options for training: `--passes=4 -b 20 --loss_function=hinge --csoaa_ldf=mc`, i.e., 4 passes over the training data, a feature space size of 2^{20}, the hinge loss function and cost-sensitive one-against-all multiclass reduction with label-dependent features.

[9]Cf. total accuracy vs. classifier accuracy in Tables 1 and 2.

[10]The automatic deep analysis pipelines for both languages are shown on the Treex demo website at `https://lindat.mff.cuni.cz/services/treex-web/run`. They include part-of-speech taggers (Spoustová et al., 2007; Straková et al., 2014) and a dependency parser (McDonald et al., 2005), plus a rule-based conversion of the resulting dependency trees to the deep layer.

[11]See (Dušek et al., 2012) for a more detailed description of formemes.

	Unl-F1	Lab-F1	TotAcc	ClAcc
previous	94.53	80.30	84.95	80.03
Monolingual	95.84	82.39	85.97	81.38
+ aligned lemmas*	95.84	82.59	86.18	81.65
+ val. lexicon**	95.84	82.93	86.53	82.14

Table 1: Experimental results for English
All numbers are percentages. Unl-F1 and Lab-F1 stand for unlabeled and labeled sense detection F1-measure, respectively (see Section 3.4 for details). TotAcc is the total accuracy (including 1st frame from the lexicon in unambiguous verbs), ClAcc is the classifier accuracy (disregarding unambiguous verbs). "*" marks a statistically significant improvement over the Monolingual setting at 95% level, "**" at 99% level.[12]

	Unl-F1	Lab-F1	TotAcc	ClAcc
previous (PDT)	96.90	76.65	79.70	72.41
monoling./PDT	96.94	77.97	80.43	75.64
monoling./PCEDT	97.34	80.22	82.41	78.12
+ aligned lemmas	97.34	80.30	82.50	78.24
+ val. lexicon*	97.34	80.47	82.66	78.45

Table 2: Experimental results for Czech
See Table 1 for a description of labels. We include the performance of our Monolingual setting on PDT 2.5 for comparison with our previous work.

mapping, i.e., a list of "known possible translations" for this verb sense.

The new binary feature exploits the fact that the possible translation lists are typically different for different senses of the same verb: given a verb token and an aligned token from the other language, the feature is set to "true" for those candidate senses that have the aligned token's lemma on the list of their possible translations.

Since the same feature is shared for all verbs (only its value varies), it is guaranteed to occur very frequently, which should increase its usefulness to the classifier.

3.4 Results

The results of the individual settings are given in Tables 1 and 2. The figures include the sense detection F-measure in an unlabeled (just detecting a verb occurrence whose sense must be inferred) and labeled setting (also selecting the correct sense) as well as the accuracy of the sense detection alone (in total and in ambiguous verbs with two or more senses).

We can see that just using the Vowpal Wabbit classifier with the same features provides a substantial performance boost. The aligned lemma features bring a very mild improvement both in English and Czech (not statistically significant for Czech). Using the CzEngVallex mapping feature brings a significant improvement of 0.8% in English and 0.3% in Czech labeled F1 absolute.[12]

The lower gain in Czech from both aligned lemmas and the CzEngVallex mapping can be explained by a higher ambiguity on average of the equivalents used in English (cf. the number of different verbs in PCEDT used in Czech and English in Section 2.3). The aligned English verbs are thus not as helpful for the disambiguation of Czech verbs as is the case in the reversed direction. In addition, the problem itself seems to be harder for Czech on the PCEDT data, given the higher number of senses on average and the higher number of verbs, i.e., greater data sparsity.

The most probable cause for the low gain from aligned lemmas is that the aligned lemma features are relatively sparse (they are different for each lemma and the classifier is not able to connect them). On the other hand, the single binary CzEngVallex feature occurs frequently and can thus then help even in rare verbs with a low number of training examples. A more detailed analysis of the results suggests that this is indeed the case: in both languages, aligned lemma features help mostly for more common verbs whereas the CzEngVallex mapping feature also improves WSD of rarer verbs.

For each language, we examined in detail a sample of randomly selected 30 cases where our three setups gave different results. The positive effect brought about by the aligned lemma features and the CzEngVallex mapping features was evident (examples are shown in Figures 3 and 4 for English and Czech, respectively). We could also find a few cases where the setups using parallel features improved even though there was no helpful aligned translation for the verb in question: even the non-presence of information from the other language can be a hint to the classifier. We have also found cases where the parallel data information introduced noise. This was mostly caused by a translation using an ambiguous verb (see Figure 5), or a verb that would usually suggest a different sense (see Figure 6). In addition, we found in our samples one case of alignment error leading to misclassification and one probable

---

[12] We used paired bootstrap resampling (Koehn, 2004) with 1,000 resamples to assess statistical significance.

PCEDT annotation error. On the whole, the positive effects of using information from parallel data are prevailing.

4 Related work

Within semantic role labeling (SRL) tasks, predicate detection is often part of the task, whereas WSD is not.[13] Due to limited lexicon coverage, we have used verbs only and evaluated on the frame (sense) assigned to the occurrence of the verb in the corpus. While the best results reported for the CoNLL 2009 Shared task are 85.41% labeled F1 for Czech and 85.63% for English (Björkelund et al., 2009), they are not comparable for several reasons, the main being that SRL evaluates each argument separately, while for a frame to be counted as correct in our task, the whole frame (by means of its reference ID) must be correct, which is substantially harder (if only for verbs). Moreover, we have used a newer version of the PDT (including PDT-Vallex) and EngVallex-annotated verbs in the PCEDT, while the English CoNLL 2009 Shared Task is PropBank-based.[14]

Dependency information is also often used for WSD outside of SRL tasks (Lin, 1997; Chen et al., 2009), but remains mostly limited to surface syntax.

WSD for verbs has been tackled previously, e.g. (Edmonds and Cotton, 2001; Chen and Palmer, 2005). These experiments, however, do not consider subcategorization/valency information explicitly.

Previous work on verbal WSD using the PDT Czech data includes a rule-based tool of Honetschläger (2003) and experiments by Semecký (2007) using machine learning. However, they have used gold-standard annotation for features.

The closest approach to ours is by Tufiş et al. (2004), where both a dictionary (WordNet) and a parallel corpus is used for WSD on the Orwell's 1984 novel (achieving a relatively low 74.93% F1).

Generally, the hybrid approach combining manually created dictionaries with machine learning has been applied to other tasks as well; we have already mentioned SMT (Brown et al., 1993). Dic-

tionaries have been used in POS tagging (Hajič, 2000). More distant is the approach of, e.g., Brown et al. (1992) and Ide et al. (2002), where parallel text is used for learning supervision, but not for feature extraction; Diab and Resnik (2002) use an unsupervised method.

We should also mention the idea of using parallel corpora as hidden features, a task first performed by (Brown et al., 1992) for WSD and subsequently in many other tasks, such as named entity recognition (Kim et al., 2012), dependency parsing (Haulrich, 2012; Rosa et al., 2012) or coreference resolution (Novák and Žabokrtský, 2014). Cross-language annotation projection is also a related method: see, for instance, (van der Plas and Apidianaki, 2014).

5 Conclusions and future work

We can conclude that the "hybrid" system combining the use of a parallel treebank and manually created bilingual valency lexicon described herein significantly outperformed the previous results, where only monolingual data and features have been used. We compared that to the case where only lemmas projected through word alignment are used (to distinguish the contribution of the parallel corpus alone vs. the manual lexicon), and the lemma features alone brought a very mild improvement (not statistically significant for Czech).

While it shows the usefulness of manually created lexical resources in this particular task,[15] we are planning to extend our WSD system in the future in two ways: first, to use automatically translated texts (instead of a manually translated parallel corpus), and second, to use automatically extracted valency alignments based on our Czech-English "manual" experience with CzEngVallex. In both cases, we would also like to test our approach on other language pairs (most likely with English as the one of the languages due to its rich resources). Both extensions are certainly possible, and they would allow a fair comparison against a truly monolingual WSD task without any additional resources at runtime, but of course it will have to be seen whether the noise introduced by these two automatic steps overrides the positive effects reported here.

---

[13]Predicate identification has not been part of the CoNLL 2009 shared task (Hajič et al., 2009), though.

[14]Please recall that EngVallex is a manually refined PropBank with different labeling scheme and generally $m : n$ mapping between PropBank and EngVallex frames.

[15]For POS tagging, a "hybrid" combination of a dictionary and a statistical tagger have also proved successful (Hajič, 2000).

EN: But those machines are still *considered* novelties, [. . .]

CS: Ale tyto stroje [. . .] jsou stále *považovány* ('believe to be') za novinky.

- Wrongly classified as **consider**[1] ('think about') in the monolingual setting, corrected as **consider**[2] ('believe to be') with aligned lemmas and val. lexicon.

EN: This *feels* more like a one-shot deal.

CS: Teď to *vypadá* ('looks like') spíš na jednorázovou záležitost.

- Wrongly classified as **feel**[4] ('have a feeling') in the monolingual and aligned lemma settings, corrected as **feel**[5] ('look like') with val. lexicon.

Figure 3: Examples of English WSD improved by information from Czech parallel texts (top: aligned lemma features help with a verb that is relatively frequent in the training data, bottom: the CzEngVallex mapping feature helps with a rarer verb).

CS: [. . .] čemu lidé z televizního průmyslu *říkají* ('call') stanice „s nejvyšší spontánní znalostí".

EN: [. . .] what people in the television industry *call* a "top of mind" network.

- Wrongly classified as **říkat**[7] ('say') in the monolingual setting, corrected as **říkat**[4] ('call') with aligned lemmas and val. lexicon.

CS: Jestliže investor *neposkytne* ('does not provide, give, lend') dodatečnou hotovost [. . .]

EN: If the investor doesn't *put up* the extra cash [. . .]

- Wrongly classified as **poskytnout**[2] ('light verb, give (chance, opportunity etc.)') in the monolingual and aligned lemma settings, corrected as **poskytnout**[1] ('provide, lend') with val. lexicon.

Figure 4: Examples of Czech WSD improved by information from English parallel text (top: a relatively frequent verb, bottom: less frequent verb).

EN: Laptops [. . .] have become the fastest-growing personal computer segment , with sales *doubling* this year .

CS: Laptopy [. . .] se staly, díky letošnímu *zdvojnásobení* objemu prodeje, nejrychleji rostoucím segmentem mezi osobními počítači .

- Correctly classified as **double**[3] ('become twice as large') in the monolingual setting, misclassified as **double**[2] ('make twice as large') with aligned lemmas and val. lexicon. The Czech word *zdvojnásobení* is ambiguous and allows both senses.

CS: Výrobek firmy Atari Corp . Portfolio [. . .] stojí pouhých 400 $ a *běží* na třech AA bateriích [. . .]

EN: Atari Corp. 's Portfolio [. . .] costs a mere $ 400 and *runs* on three AA batteries [. . .]

- Correctly classified as **běžet**[6] ('work, function') in the monolingual and aligned lemmas setting, misclassified as **běžet**[3] ('move on foot') with val. lexicon. The English translation *run* allows both senses.

Figure 5: Examples of translations using ambiguous verbs which did not help in WSD (top: English, bottom: Czech).

EN: "We didn't even get a chance to *do* the programs we wanted to do."

CS: „Nedali nám žádnou šanci *uskutečnit* plány, které jsme měli připravené."

- Correctly classified as **do**[6] ('perform (a function), run (a trade)') in the monolingual and aligned lemmas setting, misclassified as **do**[2] ('perform an act') with val. lexicon. The Czech word *uskutečnit* ('accomplish') suggests an incorrect reading.

CS: [...] například Iowa *zaznamenala* [...] nárůst populace o 11000 lidí [...]

EN: Iowa , for instance , *saw* its population grow by 11,000 people [...]

- Correctly classified as **zaznamenat**[5] ('light verb, experience (rise, difficulty, gain etc.)') in the monolingual and val. lexicon setting, misclassified as **zaznamenat**[1] ('notice') with aligned lemmas. The English verb *see* would usually suggest the latter sense.

Figure 6: Examples of translations using verbs that would typically suggest a different sense than the correct one.

Acknowledgments

References

A. Björkelund, L. Hafdell, and P. Nugues. 2009. Multilingual semantic role labeling. In *Proceedings of CoNLL 2009: Shared Task*, pages 43–48, Boulder, Colorado, United States.

P. F. Brown, S. A. Della Pietra, V. J. Della Pietra, J. D. Lafferty, and R. L. Mercer. 1992. Analysis, statistical transfer, and synthesis in machine translation. In *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation*, pages 83–100.

P. F. Brown, S. A. Della Pietra, V. J. Della Pietra, M. J. Goldsmith, J. Hajič, R. L. Mercer, and S. Mohanty. 1993. But dictionaries are data too. In *Proceedings of the Workshop on Human Language Technology*, HLT '93, pages 202–205.

J. Chen and M. Palmer. 2005. Towards robust high performance word sense disambiguation of English verbs using rich linguistic features. In *Natural Language Processing–IJCNLP 2005*, pages 933–944. Springer.

P. Chen, W. Ding, C. Bowes, and D. Brown. 2009. A fully unsupervised word sense disambiguation method using dependency knowledge. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 28–36. Association for Computational Linguistics.

S. Cinková. 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*.

M. Diab and P. Resnik. 2002. An unsupervised method for word sense tagging using parallel corpora. In *Proceedings of the 40th ACL*, pages 255–262.

O. Dušek, Z. Žabokrtský, M. Popel, M. Majliš, M. Novák, and D. Mareček. 2012. Formemes in English-Czech deep syntactic MT. In *Proceedings of the Seventh Workshop on Statistical Machine Translation*, page 267–274.

O. Dušek, J. Hajič, and Z. Urešová. 2014. Verbal valency frame detection and selection in Czech and English. In *The 2nd Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, pages 6–11, Baltimore. Association for Computational Linguistics.

P. Edmonds and S. Cotton. 2001. Senseval-2: Overview. In *The Proceedings of the Second International Workshop on Evaluating Word Sense Disambiguation Systems*, SENSEVAL '01, pages 1–5.

R. E Fan, K. W Chang, C. J Hsieh, X. R Wang, and C. J Lin. 2008. LIBLINEAR: a library for large linear classification. *The Journal of Machine Learning Research*, 9:1871–1874.

J. Hajič, M. Ciaramita, R. Johansson, D. Kawahara, M. A. Martí, L. Màrquez, A. Meyers, J. Nivre, S. Padó, J. Štěpánek, P. Straňák, M. Surdeanu, N. Xue, and Y. Zhang. 2009. The CoNLL-2009

shared task: Syntactic and semantic dependencies in multiple languages. In *Proceedings of CoNLL-2009*, Boulder, Colorado, USA.

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. 2012. Announcing Prague Czech-English Dependency Treebank 2.0. In *Proceedings of LREC*, pages 3153–3160.

J. Hajič. 2000. Morphological tagging: Data vs. dictionaries. In *Proceedings of NAACL*, pages 94–101.

J. Hajič, J. Panevová, Z. Urešová, A. Bémová, V. Kolářová, and P. Pajas. 2003. PDT-VALLEX: creating a large-coverage valency lexicon for treebank annotation. In *Proceedings Of The 2nd Workshop on Treebanks and Linguistic Theories*, volume 9, page 57–68.

J. Hajič, J. Panevová, E. Hajičová, P. Sgall, P. Pajas, J. Štěpánek, J. Havelka, M. Mikulová, Z. Žabokrtský, M. Ševčíková Razímová, and Z. Urešová. 2006. *Prague Dependency Treebank 2.0*. Number LDC2006T01. LDC, Philadelphia, PA, USA.

M. W. Haulrich. 2012. *Data-driven bitext dependency parsing and alignment*. Ph.D. thesis, Copenhagen Business School, Department of International Business Communication.

V. Honetschläger. 2003. Using a Czech valency lexicon for annotation support. In *Text, Speech and Dialogue*, pages 120–125. Springer.

N. Ide, T. Erjavec, and D. Tufiş. 2002. Sense discrimination with parallel corpora. In *Proceedings of the ACL-02 Workshop on Word Sense Disambiguation - Volume 8*, WSD '02, pages 61–66.

S. Kim, K. Toutanova, and H. Yu. 2012. Multilingual named entity recognition using parallel data and metadata from Wikipedia. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers - Volume 1*, pages 694–702, Stroudsburg, PA, USA. Association for Computational Linguistics.

P. Koehn. 2004. Statistical significance tests for machine translation evaluation. In *Empirical Methods in Natural Language Processing*, pages 388–395.

J. Langford, L. Li, and A. Strehl. 2007. Vowpal Wabbit online learning project.

D. Lin. 1997. Using syntactic dependency as local context to resolve word sense ambiguity. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*, pages 64–71, Madrid, Spain. Association for Computational Linguistics.

M. P. Marcus, M. A. Marcinkiewicz, and B. Santorini. 1993. Building a large annotated corpus of English: The Penn Treebank. *COMP LING*, 19(2):330.

R. McDonald, F. Pereira, K. Ribarov, and J. Hajič. 2005. Non-projective dependency parsing using spanning tree algorithms. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 523–530.

M. Novák and Z. Žabokrtský. 2014. Cross-lingual coreference resolution of pronouns. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 14–24, Dublin, Ireland, August. Dublin City University and Association for Computational Linguistics.

F. J. Och and H. Ney. 2003. A systematic comparison of various statistical alignment models. *Comput. Linguist.*, 29(1):19–51, March.

M. Palmer, D. Gildea, and P. Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106.

M. Popel and Z. Žabokrtský. 2010. TectoMT: modular NLP framework. *Advances in Natural Language Processing*, pages 293–304.

R. Rosa, O. Dušek, D. Mareček, and M. Popel. 2012. Using parallel features in parsing of machine-translated sentences for correction of grammatical errors. In *Proceedings of the Sixth Workshop on Syntax, Semantics and Structure in Statistical Translation*, SSST-6 '12, pages 39–48, Stroudsburg, PA, USA. Association for Computational Linguistics.

K. K. Schuler. 2005. *VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon*. Ph.D. thesis, Department of Computer and Information Science, University of Pennsylvania, Philadelphia.

J. Semecký. 2007. Verb valency frames disambiguation. *The Prague Bulletin of Mathematical Linguistics*, (88):31–52.

P. Sgall, E. Hajičová, and J. Panevová. 1986. *The meaning of the sentence in its semantic and pragmatic aspects*. D. Reidel, Dordrecht.

D. J. Spoustová, J. Hajič, J. Votrubec, P. Krbec, and P. Květoň. 2007. The Best of Two Worlds: Cooperation of Statistical and Rule-based Taggers for Czech. In *Proceedings of the Workshop on Balto-Slavonic Natural Language Processing: Information Extraction and Enabling Technologies*, pages 67–74, Stroudsburg, PA, USA. Association for Computational Linguistics.

J. Straková, M. Straka, and J. Hajič. 2014. Open-Source Tools for Morphology, Lemmatization, POS Tagging and Named Entity Recognition. In *ACL 2014*, pages 13–18. Association for Computational Linguistics.

D. Tufiş, R. Ion, and N. Ide. 2004. Fine-grained word sense disambiguation based on parallel corpora, word alignment, word clustering and aligned wordnets. In *Proceedings of the 20th COLING'04*.

Z. Urešová, O. Dušek, E. Fučíková, J. Hajič, and J. Šindlerová. 2015a. Bilingual English-Czech valency lexicon linked to a parallel corpus. In *Proceedings of LAW IX - The 9th Linguistic Annotation Workshop*, pages 124–128, Denver, Colorado. Association for Computational Linguistics.

Z. Urešová, E. Fučíková, and J. Šindlerová. 2015b. CzEngVallex: Mapping Valency between Languages. Technical Report TR-2015-58, Charles University in Prague, Institute of Formal and Applied Lingustics, Prague. To appear at `http://ufal.mff.cuni.cz/techrep/tr58.pdf`.

Z. Urešová. 2011. *Valenční slovník Pražského závislostního korpusu (PDT-Vallex)*. Studies in Computational and Theoretical Linguistics. Prague.

L. van der Plas and M. Apidianaki. 2014. Cross-lingual word sense disambiguation for predicate labelling of French. In *TALN-RECITAL, 21ème Traitement Automatique des Langues Naturelles, Marseille, 2014*.