EUROSENSE: Automatic Harvesting of Multilingual Sense Annotations from Parallel Text

Claudio Delli Bovi, Jose Camacho-Collados, Alessandro Raganato and **Roberto Navigli**
Department of Computer Science
Sapienza University of Rome
{dellibovi,collados,raganato,navigli}@di.uniroma1.it

Abstract

Parallel corpora are widely used in a variety of Natural Language Processing tasks, from Machine Translation to cross-lingual Word Sense Disambiguation, where parallel sentences can be exploited to automatically generate high-quality sense annotations on a large scale. In this paper we present EUROSENSE, a multilingual sense-annotated resource based on the joint disambiguation of the Europarl parallel corpus, with almost 123 million sense annotations for over 155 thousand distinct concepts and entities from a language-independent unified sense inventory. We evaluate the quality of our sense annotations intrinsically and extrinsically, showing their effectiveness as training data for Word Sense Disambiguation.

1 Introduction

One of the long-standing challenges in Natural Language Processing (NLP) lies in automatically identifying the meaning of words in context. Various lines of research have been geared towards achieving this goal, most notably Word Sense Disambiguation (Navigli, 2009, WSD) and Entity Linking (Rao et al., 2013, EL). In both tasks, supervised approaches (Zhong and Ng, 2010; Melamud et al., 2016; Iacobacci et al., 2016; Kågebäck and Salomonsson, 2016) tend to obtain the best performances over standard benchmarks but, from a practical standpoint, they lose ground to knowledge-based approaches (Agirre et al., 2014; Moro et al., 2014b; Weissenborn et al., 2015), which scale better in terms of scope and number of languages. In fact, the development of supervised disambiguation systems depends crucially on the availability of re-

liable sense-annotated corpora, which are indispensable in order to provide solid training and testing grounds (Pilehvar and Navigli, 2014). However, hand-labeled sense annotations are notoriously difficult to obtain on a large scale, and manually curated corpora (Miller et al., 1993; Passonneau et al., 2012) have a limited size. Given that scaling the manual annotation process becomes practically unfeasible when both lexicographic and encyclopedic knowledge is addressed (Schubert, 2006), recent years have witnessed efforts to produce larger sense-annotated corpora automatically (Moro et al., 2014a; Taghipour and Ng, 2015a; Scozzafava et al., 2015; Raganato et al., 2016). Even though these automatic approaches produce noisier corpora, it has been shown that training on them leads to better supervised and semi-supervised models (Taghipour and Ng, 2015b; Raganato et al., 2016; Yuan et al., 2016; Raganato et al., 2017), as well as to effective embedded representations for senses (Iacobacci et al., 2015; Flekova and Gurevych, 2016).

A convenient way of generating sense annotations is to exploit parallel corpora and word alignments (Taghipour and Ng, 2015a): indeed, parallel corpora exist in many flavours (Tiedemann, 2012) and are widely used across the NLP community for a variety of different tasks. In this paper we focus on Europarl (Koehn, 2005)[1], one of the most popular multilingual corpora, originally designed to provide aligned parallel text for Machine Translation (MT) systems. Extracted from the proceedings of the European Parliament, the latest release of the Europarl corpus comprises parallel text for 21 European languages, with more than 743 million tokens overall.

Apart from its prominent role in MT as a training set, the Europarl corpus has been used

---

[1] http://opus.lingfil.uu.se/Europarl.php

for cross-lingual WSD (Lefever and Hoste, 2010, 2013), including, more recently, preposition sense disambiguation (Gonen and Goldberg, 2016), and widely exploited to develop cross-lingual word embeddings (Hermann and Blunsom, 2014; Gouws et al., 2015; Coulmance et al., 2015; Vyas and Carpuat, 2016; Vulić and Korhonen, 2016; Artetxe et al., 2016) as well as multi-sense embeddings (Ettinger et al., 2016; Šuster et al., 2016).

In this paper, our aim is to augment Europarl with sense-level information for multiple languages, thereby constructing a large-scale sense-annotated multilingual corpus which has the potential to boost both WSD and MT research.

We follow an approach that has already proved effective in a definitional setting (Camacho-Collados et al., 2016a): unlike previous cross-lingual approaches, we do not rely on word alignments against a pivot language, but instead leverage all languages at the same time in a joint disambiguation procedure that is subsequently refined using distributional similarity. We draw on the wide-coverage multilingual encyclopedic dictionary of BabelNet (Navigli and Ponzetto, 2012)[2], which enables us to seamlessly cover lexicographic and encyclopedic knowledge in multiple languages within a unified sense inventory.

As a result of our disambiguation pipeline we obtain and make available to the community EU-ROSENSE, a multilingual sense-annotated corpus with almost 123 million sense annotations of more than 155 thousand distinct concepts and named entities drawn from the multilingual sense inventory of BabelNet, and covering all the 21 languages of the Europarl corpus. As such EUROSENSE constitutes, to our knowledge, the largest corpus of its kind.

2 Related Work

Extending sense annotations to multiple languages is a demanding endeavor, especially when manual intervention is required. Despite the fact that sense-annotated corpora for a number of languages have been around for more than a decade (Petrolito and Bond, 2014), they either include few samples per word sense, or only cover a restricted set of ambiguous words (Passonneau et al., 2012); as a result, multilingual WSD was until recently almost exclusively tackled using knowledge-based approaches (Agirre et al., 2014;

Moro et al., 2014b). Nowadays, however, the rapid development of NLP pipelines for languages other than English has been opening up the possibilities for the automatic generation of multilingual sense-annotated data. Nevertheless, the few approaches that have been proposed so far are either focused on treating each individual language in isolation (Otegi et al., 2016), or limited to short and concise definitional text (Camacho-Collados et al., 2016a).

On the other hand, the use of parallel text to perform WSD (Ng et al., 2003; Lefever et al., 2011; Yao et al., 2012; Bonansinga and Bond, 2016) or even Word Sense Induction (Apidianaki, 2013) has been widely explored in the literature, and has demonstrated its effectiveness in producing high-quality sense-annotated data (Chan and Ng, 2005). This strategy, however, requires word alignments for each language pair to be taken into account, with alignment errors that might propagate and hamper subsequent stages unless human supervision is employed to correct erroneous annotations (Taghipour and Ng, 2015a). Moreover, cross-language disambiguation using parallel text requires a language-independent annotation framework that goes beyond monolingual WordNet-like sense inventories (Lefever et al., 2011) in order for the annotations obtained to be used effectively within end-to-end applications.

With EUROSENSE, instead, the key idea is to exploit at best parallel sentences to provide enriched context for a joint multilingual disambiguation. Using BabelNet, a unified multilingual sense inventory, we obtain language-independent sense annotations for a wide variety of concepts and named entities, which can be seamlessly mapped to individual semantic resources (e.g WordNet, Wikipedia, DBpedia) via Babel-Net's inter-resource mappings.

3 Building EUROSENSE

Following Camacho-Collados et al. (2016a), our fully automatic disambiguation pipeline for constructing EUROSENSE couples a graph-based multilingual joint WSD/EL system, Babelfy (Moro et al., 2014b)[3], and a language-independent vector representation of concepts and entities, NASARI (Camacho-Collados et al., 2016b).[4] It comprises two stages: multilingual disambigua-

---

[2] http://babelnet.org

[3] http://babelfy.org
[4] http://lcl.uniroma1.it/nasari

tion (Section 3.1) and refinement based on distributional similarity (Section 3.2).

3.1 Stage 1: Multilingual Disambiguation

As a preprocessing step, we part-of-speech tag and lemmatize the whole corpus using TreeTagger (Schmid, 1995)[5]. We perform disambiguation at the sentence level. However, instead of disambiguating each sentence in isolation, language by language, we first identify all available translations of a given sentence and then gather these together into a single multilingual text.

Then, we disambiguate this multilingual text using Babelfy. Given that Babelfy is capable of handling text with multiple languages at the same time, this multilingual extension effectively increases the amount of context for each sentence, and directly helps in dealing with highly ambiguous words in any particular language (as the translations of these words may be less ambiguous in some different language). Moreover, given the multilingual nature of our sense inventory, Babelfy's high-coherence approach favors naturally sense assignments that are consistent across languages at the sentence level (i.e. those having fewer distinct senses shared by more translations of the same sentence).

As a result, we obtain a full, high-coverage version of EUROSENSE where each disambiguated word or multi-word expression (*disambiguated instance*) is associated with a coherence score.[6]

3.2 Stage 2: Similarity-based Refinement

In this stage we aim at improving the sense annotations obtained in the previous step (Section 3.1), with a procedure specifically targeted at correcting and extending these sense annotations. In general, graph-based WSD systems, such as Babelfy, have been shown to be heavily biased towards the Most Common Sense (MCS) (Calvo and Gelbukh, 2015). In order to get a handle on this bias and improve our pipeline's disambiguation accuracy we adopt a refinement based on distributional similarity, which is not affected by the MCS.

To this end, we exploit the 300-dimensional embedded representations of concepts and entities of NASARI to discard or refine disambiguated instances that are less semantically coherent. These NASARI vector representations were constructed by combining structural and distributional knowledge from Wikipedia and Word-Net with Word2Vec word embeddings (Mikolov et al., 2013) trained on textual corpora.

For each sentence, we first identify a subset D of high-confidence disambiguations[7] from among those given by Babelfy in the previous step. Then, we calculate the centroid of all the NASARI vectors corresponding to the elements of D, and we re-disambiguate the mentions associated with the remaining low-confidence disambiguated instances (i.e. those not in D), by picking, for each mention w, the concept or entity $\hat{s}$ whose NASARI vector[8] is closest to the centroid of the sentence:

$$\hat{s} = \underset{s \in S_w}{\operatorname{argmax}} \, cos\left(\frac{\sum_{d \in D} \vec{d}}{|D|}, \vec{s}\right) \qquad (1)$$

where S_w is the set of all candidate senses for mention w according to BabelNet. Cosine similarity (cos) is used as similarity measure. Finally, in order to discard less confident annotations, we consider the cosine value associated with each refined disambiguation as confidence score, and use it to compare each disambiguated instance against an empirically validated threshold of 0.75.

As a result, we obtain the refined high-precision version of EUROSENSE, where each disambiguated instance is associated with both a coherence score and a distributional similarity score.

4 Corpus and Statistics

Table 1 reports general statistics on EUROSENSE regarding both its high-coverage (cf. Section 3.1) and high-precision (cf. Section 3.2) versions. Joint multilingual disambiguation with Babelfy generated more than 215M sense annotations of 247k distinct concepts and entities, while similarity-based refinement retained almost 123M high-confidence instances (56.96% of the total), covering almost 156k distinct concepts and entities. 42.40% of these retained annotations were corrected or validated using distributional similarity. As expected, the distribution over parts of speech is skewed towards nominal senses (64.79% before refinement and 81.79% after refinement)

---

[5]We rely on the internal preprocessing pipeline of Babelfy for those languages not supported by TreeTagger.

[6]As in Camacho-Collados et al. (2016a), coherence score is given by the normalized number of connections of a given concept within the sentence.

[7]We follow Camacho-Collados et al. (2016a) and consider disambiguated instances with a coherence score above 0.125.

[8]Given a concept or entity s we indicate with $\vec{s}$ its corresponding NASARI vector.

		Total	EN	FR	DE	ES
Full	# Annotations	215 877 109	26 455 574	22 214 996	16 888 108	21 486 532
	Distinct lemmas covered	567 378	60 853	30 474	66 762	43 892
	Distinct senses covered	247 706	138 115	65 301	75 008	74 214
	Average coherence score	0.19	0.19	0.18	0.18	0.18
Refined	# Annotations	122 963 111	15 441 667	12 955 469	9 165 112	12 193 260
	Distinct lemmas covered	453 063	42 947	23 603	50 681	31 980
	Distinct senses covered	155 904	86 881	49 189	52 425	52 859
	Average coherence score	0.29	0.28	0.25	0.28	0.27

Table 1: General statistics on EUROSENSE before *(full)* and after refinement *(refined)* for all the 21 languages. Language-specific figures are also reported for the 4 languages of the intrinsic evaluation.

followed by verbs (19.26% and 12.22%), adjectives (11.46% and 5.24%) and adverbs (4.48% and 0.73%). We note that the average coherence score increases from 0.19 to 0.29 after refinement, suggesting that distributional similarity tends to favor sense annotations that are also consistent across different languages. Table 1 also includes language-specific statistics on the 4 languages of the intrinsic evaluation, where the average lexical ambiguity ranges from 1.12 senses per lemma (German) to 2.26 (English) and, as expected, decreases consistently after refinement.

Interestingly enough, if we consider all the 21 languages, the total number of distinct lemmas covered is more than twice the total number of distinct senses: this is a direct consequence of having a unified, language-independent sense inventory (BabelNet), a feature that sets EUROSENSE apart from previous multilingual sense-annotated corpora (Otegi et al., 2016). Finally we note from the global figures on the number of covered senses that 109 591 senses (44.2% of the total) are not covered by the English sense annotations: this suggests that EUROSENSE relies heavily on multi-linguality in integrating concepts or named entities that are tied to specific social or cultural aspects of a given language (and hence would be underrepresented in an English-specific sense inventory).

5 Experimental Evaluation

We assessed the quality of EUROSENSE's sense annotations both intrinsically, by means of a manual evaluation on four samples of randomly extracted sentences in different languages (Section 5.1), as well as extrinsically, by augmenting the training set of a state-of-the-art supervised WSD system (Zhong and Ng, 2010) and showing that

it leads to consistent performance improvements over two standard WSD benchmarks (Section 5.2).

5.1 Intrinsic Evaluation: Annotation Quality

In order to assess annotation quality directly, we carried out a manual evaluation on 4 different languages (English, French, German and Spanish) with 2 human judges per language. We sampled 50 random sentences across the subset of sentences in EUROSENSE featuring a translation in all 4 languages, totaling 200 sentences overall.

For each sentence, we evaluated all sense annotations both before and after the refinement stage, along with the sense annotations obtained by a baseline that disambiguates each sentence in isolation with Babelfy. Overall, we manually verified a total of 5818 sense annotations across the three configurations (1518 in English, 1564 in French, 1093 in German and 1643 in Spanish). In every language the two judges agreed in more than 85% of the cases, with an inter-annotator agreement in terms of Cohen's kappa (Cohen, 1960) above 60% in all evaluations (67.7% on average).

Results, reported in Table 2, show that joint multilingual disambiguation improves consistently over the baseline. The similarity-based refinement boosts precision even further, at the expense of a reduced coverage (whereas both Babelfy and the baseline attempt an answer for every disambiguation target). Over the 4 languages, sense annotations appear to be most reliable for German, which is consistent with its lower lexical ambiguity on the corpus (cf. Section 4).

5.2 Extrinsic Evaluation: Word Sense Disambiguation

We additionally carried out an extrinsic evaluation of EUROSENSE by using its refined sense an-

	EN		FR		DE		ES	
	Prec.	**Cov.**	**Prec.**	**Cov.**	**Prec.**	**Cov.**	**Prec.**	**Cov.**
Babelfy	76.1	100	59.1	100	80.4	100	67.5	100
EUROSENSE (full)	80.3	100	67.9	100	84.6	100	76.7	100
EUROSENSE (refined)	**81.5**	75.0	**71.8**	63.5	**89.3**	53.8	**82.5**	62.9

Table 2: Precision *(Prec.)* and coverage *(Cov.)* of EUROSENSE, manually evaluated on a random sample in 4 languages. Precision is averaged between the two judges, and coverage is computed assuming each content word in the sense inventory to be a valid disambiguation target.

notations for English as a training set for a supervised all-words WSD system, It Makes Sense (Zhong and Ng, 2010, IMS). Following Taghipour and Ng (2015a), we started with SemCor (Miller et al., 1993) as initial training dataset, and then performed a subsampling of EUROSENSE up to 500 additional training examples per word sense. We then trained IMS on this augmented training set and tested on the two most recent standard benchmarks for all-words WSD: the SemEval-2013 task 12 (Navigli et al., 2013) and the SemEval-2015 task 13 (Moro and Navigli, 2015) test sets. As baselines we considered IMS trained on SemCor only and OMSTI, the sense-annotated dataset constructed by Taghipour and Ng (2015a) which also includes SemCor. Finally, we report the results of UKB, a knowledge-based system (Agirre et al., 2014).[9] As shown in Table 3, IMS trained on our augmented training set consistently outperforms all baseline models, showing the reliability of EUROSENSE as training corpus, even against sense annotations obtained semi-automatically (Taghipour and Ng, 2015a).

6 Release

EUROSENSE is available at `http://lcl.uniroma1.it/eurosense`. We release two different versions of the corpus:

- A *high-coverage version*, obtained after the first stage of the pipeline, i.e. multilingual joint disambiguation with Babelfy. Here, each sense annotation is associated with a coherence score (cf. Section 3.1);

- A *high-precision version*, obtained after the similarity-based refinement with NASARI. In this version, sense annotations are associated

	SemEval-2013	SemEval-2015
IMS$_{SemCor}$	65.3	69.3
IMS$_{OMSTI}$	65.0	69.1
IMS$_{EUROSENSE}$	**66.4**	**69.5**
UKB	59.0	61.2
UKB$_{w2w}$	62.9	63.3
MCS	63.0	67.8

Table 3: F-Score on all-words WSD.

with both a coherence score and a distributional similarity score (cf. Section 3.2).

7 Conclusion

In this paper we presented EUROSENSE, a large multilingual sense-annotated corpus based on Europarl, and constructed automatically via a disambiguation pipeline that exploits the interplay between a joint multilingual disambiguation algorithm and a language-independent vector-based representation of concepts and entities. Crucially, EUROSENSE relies on the wide-coverage unified sense inventory of BabelNet, which enabled the disambiguation process to exploit at best parallel text and enforces cross-language coherence among sense annotations. We evaluated EUROSENSE both intrinsically and extrinsically, showing that it provides reliable sense annotations that improve supervised models for WSD.

---

[9]We include its two implementations using the full Word-Net graph and the disambiguated glosses of WordNet as connections: default and word by word (*w2w*).

References

Eneko Agirre, Oier Lopez de Lacalle, and Aitor Soroa. 2014. Random Walks for Knowledge-Based Word Sense Disambiguation. *Computational Linguistics* 40(1):57–84.

Marianna Apidianaki. 2013. LIMSI: Cross-lingual Word Sense Disambiguation using Translation Sense Clustering. In *Proc. of SemEval*. pages 178–182.

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2016. Learning Principled Bilingual Mappings of Word Embeddings while Preserving Monolingual Invariance. In *Proc. of EMNLP*. pages 2289–2294.

Giulia Bonansinga and Francis Bond. 2016. Multilingual Sense Intersection in a Parallel Corpus with Diverse Language Families. In *Proc. of the 8th Global WordNet Conference*. pages 44–49.

Hiram Calvo and Alexander Gelbukh. 2015. Is the most frequent sense of a word better connected in a semantic network? In *International Conference on Intelligent Computing*. Springer, pages 491–499.

José Camacho-Collados, Claudio Delli Bovi, Alessandro Raganato, and Roberto Navigli. 2016a. A Large-Scale Multilingual Disambiguation of Glosses. In *Proc. of LREC*. pages 1701–1708.

José Camacho-Collados, Mohammad Taher Pilehvar, and Roberto Navigli. 2016b. Nasari: Integrating explicit knowledge and corpus statistics for a multilingual representation of concepts and entities. *Artificial Intelligence* 240:36–64.

Yee Seng Chan and Hwee Tou Ng. 2005. Scaling Up Word Sense Disambiguation via Parallel Texts. In *Proc. of AAAI*. volume 5, pages 1037–1042.

J. A. Cohen. 1960. A coefficient of agreement of nominal scales. *Educational and Psychological Measurement* 20(1):37–46.

Jocelyn Coulmance, Jean-Marc Marty, Guillaume Wenzek, and Amine Benhalloum. 2015. Transgram, Fast Cross-lingual Word-embeddings. In *Proc. of EMNLP*. pages 1109–1113.

Allyson Ettinger, Philip Resnik, and Marine Carpuat. 2016. Retrofitting Sense-Specific Word Vectors Using Parallel Text. In *Proc. of NAACL-HLT*. pages 1378–1383.

Lucie Flekova and Iryna Gurevych. 2016. Supersense Embeddings: A Unified Model for Supersense Interpretation, Prediction and Utilization. In *Proc. of ACL*. pages 2029–2041.

Hila Gonen and Yoav Goldberg. 2016. Semi Supervised Preposition-Sense Disambiguation using Multilingual Data. In *Proc. of COLING*. pages 2718–2729.

Stephan Gouws, Yoshua Bengio, and Greg Corrado. 2015. BilBOWA: Fast Bilingual Distributed Representations without Word Alignments. In *Proc. of ICML*. pages 748–756.

Karl Moritz Hermann and Phil Blunsom. 2014. Multilingual Distributed Representations without Word Alignment. In *Proc. of ICLR*. pages 1–9.

Ignacio Iacobacci, Mohammad Taher Pilehvar, and Roberto Navigli. 2015. SensEmbed: Learning Sense Embeddings for Word and Relational Similarity. In *Proc. of ACL*. pages 95–105.

Ignacio Iacobacci, Mohammad Taher Pilehvar, and Roberto Navigli. 2016. Embeddings for Word Sense Disambiguation: An Evaluation Study. In *Proc. of ACL*. pages 897–907.

Mikael Kågebäck and Hans Salomonsson. 2016. Word Sense Disambiguation using a Bidirectional LSTM. In *Proc. of CogALex*. pages 51–56.

Philipp Koehn. 2005. Europarl: A Parallel Corpus for Statistical Machine Translation. In *Proc. of MT summit*. volume 5, pages 79–86.

Els Lefever and Veronique Hoste. 2010. SemEval-2010 Task 3: Cross-lingual Word Sense Disambiguation. In *Proc. of SemEval*. pages 15–20.

Els Lefever and Véronique Hoste. 2013. SemEval-2013 task 10: Cross-lingual Word Sense Disambiguation. In *Proc. of SemEval*. pages 158–166.

Els Lefever, Véronique Hoste, and Martine De Cock. 2011. ParaSense or How to Use Parallel Corpora for Word Sense Disambiguation. In *Proc. of ACL-HLT*. pages 317–322.

Oren Melamud, Jacob Goldberger, and Ido Dagan. 2016. context2vec: Learning Generic Context Embedding with Bidirectional LSTM. In *Proc. of CONLL*. pages 51–61.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. *CoRR abs/1301.3781*.

George A. Miller, Claudia Leacock, Randee Tengi, and Ross T. Bunker. 1993. A semantic concordance. In *Proc. of HLT*. pages 303–308.

Andrea Moro and Roberto Navigli. 2015. SemEval-2015 Task 13: Multilingual All-Words Sense Disambiguation and Entity Linking. *Proc. of SemEval* pages 288–297.

Andrea Moro, Roberto Navigli, Francesco Maria Tucci, and Rebecca J. Passonneau. 2014a. Annotating the MASC Corpus with BabelNet. In *Proc. of LREC*. pages 4214–4219.

Andrea Moro, Alessandro Raganato, and Roberto Navigli. 2014b. Entity Linking meets Word Sense Disambiguation: a Unified Approach. *Transactions of the Association for Computational Linguistics* 2:231–244.

Roberto Navigli. 2009. Word Sense Disambiguation: A survey. *ACM Computing Surveys* 41(2):1–69.

Roberto Navigli, David Jurgens, and Daniele Vannella. 2013. SemEval-2013 Task 12: Multilingual Word Sense Disambiguation. In *Proc. of SemEval*. pages 222–231.

Roberto Navigli and Simone Paolo Ponzetto. 2012. BabelNet: The Automatic Construction, Evaluation and Application of a Wide-Coverage Multilingual Semantic Network. *Artificial Intelligence* 193:217–250.

Hwee Tou Ng, Bin Wang, and Yee Seng Chan. 2003. Exploiting Parallel Texts for Word Sense Disambiguation: An Empirical Study. In *Proc. of ACL*. pages 455–462.

Arantxa Otegi, Nora Aranberri, Antonio Branco, Jan Hajic, Steven Neale, Petya Osenova, Rita Pereira, Martin Popel, Joao Silva, Kiril Simov, and Eneko Agirre. 2016. QTLeap WSD/NED Corpora: Semantic Annotation of Parallel Corpora in Six Languages. In *Proc. of LREC*. pages 3023–3030.

Rebecca J. Passonneau, Collin Baker, Christiane Fellbaum, and Nancy Ide. 2012. The MASC Word Sense Sentence Corpus. In *Proc. of LREC*. pages 3025–3030.

Tommaso Petrolito and Francis Bond. 2014. A Survey of WordNet Annotated Corpora. In *Proc. of the 7th Global WordNet Conference*. pages 236–245.

Mohammad Taher Pilehvar and Roberto Navigli. 2014. A large-scale pseudoword-based evaluation framework for state-of-the-art Word Sense Disambiguation. *Computational Linguistics* 40(4):837–881.

Alessandro Raganato, Claudio Delli Bovi, and Roberto Navigli. 2016. Automatic Construction and Evaluation of a Large Semantically Enriched Wikipedia. In *Proc. of IJCAI*. pages 2894–2900.

Alessandro Raganato, Jose Camacho-Collados, and Roberto Navigli. 2017. Word Sense Disambiguation: A Unified Evaluation Framework and Empirical Comparison. In *Proc. of EACL*. pages 99–110.

Delip Rao, Paul McNamee, and Mark Dredze. 2013. Entity Linking: Finding extracted entities in a knowledge base. *Multi-Source, Multilingual Information Extraction and Summarization* 11:93–115.

Helmut Schmid. 1995. Improvements In Part-of-Speech Tagging With an Application To German. In *Proc. of the ACL SIGDAT-Workshop*. pages 47–50.

Lenhart Schubert. 2006. Turing's Dream and the Knowledge Challenge. In *Proc. of AAAI*. pages 1534–1538.

Federico Scozzafava, Alessandro Raganato, Andrea Moro, and Roberto Navigli. 2015. Automatic identification and disambiguation of concepts and named entities in the multilingual Wikipedia. In *AI*IA*, Springer, pages 357–366.

Simon Šuster, Ivan Titov, and Gertjan van Noord. 2016. Bilingual Learning of Multi-sense Embeddings with Discrete Autoencoders. In *Proc. of NAACL-HLT*. pages 1346–1356.

Kaveh Taghipour and Hwee Tou Ng. 2015a. One Million Sense-Tagged Instances for Word Sense Disambiguation and Induction. In *Proc. of CoNLL*. pages 338–344.

Kaveh Taghipour and Hwee Tou Ng. 2015b. Semi-Supervised Word Sense Disambiguation Using Word Embeddings in General and Specific Domains. *Proc. of NAACL-HLT* pages 314–323.

Jörg Tiedemann. 2012. Parallel Data, Tools and Interfaces in OPUS. In *Proc. of LREC*. pages 2214–2218.

Ivan Vulić and Anna Korhonen. 2016. On the Role of Seed Lexicons in Learning Bilingual Word Embeddings. In *Proc. of ACL*. pages 247–257.

Yogarshi Vyas and Marine Carpuat. 2016. Sparse Bilingual Word Representations for Cross-lingual Lexical Entailment. In *Proc. of NAACL-HLT*. pages 1187–1197.

Dirk Weissenborn, Leonhard Hennig, Feiyu Xu, and Hans Uszkoreit. 2015. Multi-Objective Optimization for the Joint Disambiguation of Nouns and Named Entities. In *Proc. of ACL*. pages 596–605.

Xuchen Yao, Benjamin Van Durme, and Chris Callison-Burch. 2012. Expectations of Word Sense in Parallel Corpora. In *Proc. of NAACL-HLT*. pages 621–625.

Dayu Yuan, Julian Richardson, Ryan Doherty, Colin Evans, and Eric Altendorf. 2016. Semi-supervised Word Sense Disambiguation with Neural Models. In *Proc. of COLING*. pages 1374–1385.

Zhi Zhong and Hwee Tou Ng. 2010. It Makes Sense: A Wide-Coverage Word Sense Disambiguation System for Free Text. In *Proc. of ACL System Demonstrations*. pages 78–83.