Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 5759–5765
Marseille, 11–16 May 2020

A Short Survey on Sense-Annotated Corpora

Tommaso Pasini[1], **Jose Camacho-Collados**[2]
[1]SapienzaNLP Group, Department of Computer Science, Sapienza University of Rome, Italy
[2]School of Computer Science and Informatics, Cardiff University, United Kingdom
[1]pasini@di.uniroma1.it, [2]camachocolladosj@cardiff.ac.uk

Abstract

Large sense-annotated datasets are increasingly necessary for training deep supervised systems in Word Sense Disambiguation. However, gathering high-quality sense-annotated data for as many instances as possible is a laborious and expensive task. This has led to the proliferation of automatic and semi-automatic methods for overcoming the so-called knowledge-acquisition bottleneck. In this short survey we present an overview of sense-annotated corpora, annotated either manually- or (semi)automatically, that are currently available for different languages and featuring distinct lexical resources as inventory of senses, i.e. WordNet, Wikipedia, BabelNet. Furthermore, we provide the reader with general statistics of each dataset and an analysis of their specific features.

Keywords: Word Sense Disambiguation, Semantics, Corpus Creation, Multilinguality

1. Introduction

Word Sense Disambiguation (WSD) is a key task in Natural Language Understanding. It consists in assigning the appropriate meaning from a pre-defined sense inventory to a word in context(Navigli, 2009). While knowledge-based approaches to this task have been proposed (Agirre et al., 2014; Moro et al., 2014; Butnaru et al., 2017; Chaplot and Salakhutdinov, 2018), supervised approaches (Zhong and Ng, 2010; Melamud et al., 2016; Iacobacci et al., 2016; Kågebäck and Salomonsson, 2016; Luo et al., 2018) have been more effective in terms of performance when sense-annotated corpora are available (Raganato et al., 2017a; Huang et al., 2019). Unfortunately, obtaining such data is heavily time-consuming and expensive (Schubert, 2006), and a reasonable amount of manually-annotated instances are available for English only (Miller et al., 1993a).

One of the first attempts towards building large sense-annotated corpora was SemCor (Miller et al., 1993a) which contains instances annotated with senses from WordNet (Fellbaum, 1998). Since then, several semi-automatic and automatic approaches have also been proposed (Taghipour and Ng, 2015a; Delli Bovi et al., 2017; Pasini and Navigli, 2018). These automatic efforts tend to produce noisier annotations, but their coverage has been shown to lead to better supervised and semi-supervised WSD systems (Taghipour and Ng, 2015b; Otegi et al., 2016; Raganato et al., 2016; Yuan et al., 2016; Delli Bovi et al., 2017; Pasini and Navigli, 2017), as well as to learn effective embedded representations for senses (Iacobacci et al., 2015; Flekova and Gurevych, 2016; Mancini et al., 2017). Nevertheless, each of the aforementioned datasets come with its own format, hence making it complicated to merge them or moving from one to another. Vial et al. (2018) tackled this specific problem and, following Raganato et al. (2017a), proposed UFSAC, a unified repository of several sense-annotated corpora all with the same format, thus making it easier develop and test WSD models on different datasets.

In this survey we present the main approaches in the literature to build sense-annotated corpora, not only for WordNet but also for multilingual sense inventories, namely Wikipedia and BabelNet. There have been additional constructing sense-annotated data for other resources such as the New Oxford American Dictionary (Yuan et al., 2016) or other language-specific versions like GermaNet (Henrich et al., 2012). While these language-specific resources are certainly relevant, in this paper we have focused on English WordNet and multilingual resources with a higher coverage like Wikipedia and BabelNet.[1] Finally, we provide a general overview and statistics of these sense-annotated resources, providing relevant details across resources and languages.

2. Sense-Annotated Corpora

In this Section we describe the main efforts compiling sense-annotated corpora. We present currently available corpora for three resources: WordNet (Section 2.1), Wikipedia (Section 2.2) and BabelNet (Section 2.3). Figure 1 provides an overview of these resources and their underlying corpora.

2.1. WordNet

WordNet (Fellbaum, 1998) has been one of the most widely used knowledge resource in lexical semantics. In fact, it is the *de-facto* standard sense inventory for Word Sense Disambiguation since many years. The core unit in WordNet is the synset. A synset represents a concept or a meaning which is represented by its various lexicalizations (i.e. senses). For example, the synset defined as *motor vehicle with four wheels* can be expressed by its synonym senses *auto, automobile, machine* and *motorcar*. In what follows we list the main WordNet sense-annotated corpora, using WordNet 3.0 as reference sense inventory.

SemCor. The first and most prominent example of sense-annotated corpora is SemCor (Miller et al., 1993b). SemCor was manually annotated and consists of 352 documents from the Brown Corpus (Kucera and Francis, 1979) and 226,040 sense annotations. SemCor is the largest manually-annotated corpus and the most used in the literature to train WSD supervised systems (Agirre et al., 2009;

[1]For a more specific survey on corpora annotated with language-specific versions of WordNet, please refer to Petrolito and Bond (2014).

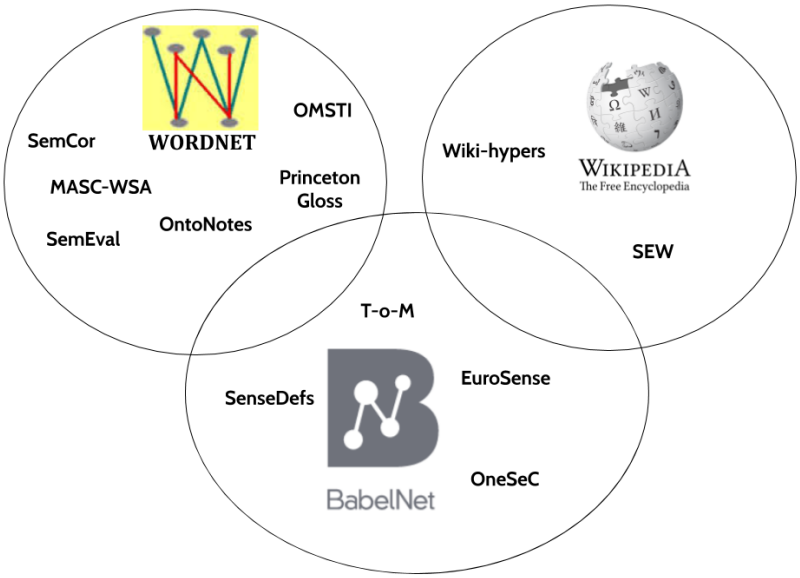

Figure 1: Overview of sense inventories with their corresponding sense-annotated corpora.

Zhong and Ng, 2010; Raganato et al., 2017b; Luo et al., 2018; Loureiro and Jorge, 2019; Huang et al., 2019).

SemEval. SemEval datasets provide reliable benchmarks for testing WSD systems. The main datasets from Senseval and SemEval competitions have been compiled and unified by Raganato et al. (2017a). In particular, the datasets from Senseval-2 (Edmonds and Cotton, 2001), Senseval-3 task 1 (Snyder and Palmer, 2004), SemEval-2007 task 17 (Pradhan et al., 2007), SemEval-2013 task 12 (Navigli et al., 2013), and SemEval-2015 task 13 (Moro and Navigli, 2015). These datasets, which have been mainly used as evaluation benchmarks for WSD systems, contain a total of 7,253 sense annotations.

MASC-WSA. The MASC Word Sense Annotation (MASC-WSA) corpus (Ide et al., 2010)is an excerpt of the Manually Annotated Sub-Corpus of American English (Ide et al., 2008, MASC) and the Open American National Corpus (Ide et al., 2002, ANC) containing annotations for 45 distinct lexemes, i.e., lemma-pos pairs, for a total of 441 distinct WordNet word senses.[2] Each word occurrence has been manually annotated on Amazon Mechanichal Turk by roughly 25 persons for a total of 1M annotations.

Princeton WordNet Gloss. The Princeton WordNet Gloss Corpus[3] is a sense-annotated corpus of textual definitions (glosses) from WordNet synsets. The corpus was tagged semi-automatically: 330,499 instances were annotated manually while the remaining annotations (i.e. 118,856) were obtained automatically. This corpus of disambiguated glosses has already proved to be useful in tasks such as semantic similarity (Pilehvar et al., 2013), domain labeling (González et al., 2012) and Word Sense Disambiguation (Baldwin et al., 2008; Agirre and Soroa, 2009; Camacho-Collados et al., 2015).

OntoNotes. OntoNotes (Weischedel et al., 2013) is a corpus from the Linguistic Data Consortium which comprises different kinds of explicitly-tagged syntactic and semantic information, including annotations at the sense level. The OntoNotes corpus consists of documents from diverse genres such as news, weblogs and telephone conversation. Its 5.0 released version contains 264,622 sense annotations.

OMSTI. The task of gathering sense annotations has proved expensive and not easily scalable. That is the reason why more recent approaches have attempted to exploit semi-automatic or automatic techniques. OMSTI[4] (Taghipour and Ng, 2015a, One Million Sense-Tagged Instances), which is a semi-automatically constructed corpus annotated with WordNet senses, is a prominent example. It was built by exploiting the alignment-based WSD approach of Chan and Ng (2005) on a large English-Chinese parallel corpus (Eisele and Chen, 2010, MultiUN corpus). OMSTI, coupled with SemCor, has already been successfully leveraged as training data for training supervised systems (Taghipour and Ng, 2015a; Iacobacci et al., 2016; Raganato et al., 2017a).

2.2. Wikipedia

Wikipedia is a collaboratively-constructed encyclopedic resource representing concepts and entities with a so-called Wikipedia article. In addition to a large coverage of concepts and entities, Wikipedia provides multilinguality, as it covers over 250 languages and these languages are connected via interlingual links. In this Section we describe

[2]This corpus has also been annotated with other resources, such as FrameNet (Baker et al., 1998). A comparison between the sense annotations of WordNet and lexical units of FrameNet is provided in De Melo et al. (2012).

[3]http://wordnet.princeton.edu/glosstag.shtml

[4]http://www.comp.nus.edu.sg/~nlp/corpora.html

two datasets providing disambiguations in the form of Wikipedia pages.[5] For these two datasets we have used the same version of Wikipedia for a more accurate comparison[6].

Wikipedia hyperlinks. This corpus contains the full Wikipedia multilingual corpus with hyperlinks as sense-annotated instances. Hyperlinks are highlighted mentions within a Wikipedia article that directly links to another Wikipedia page.

SEW. The Semantically Enriched Wikipedia [7] (Raganato et al., 2016, SEW) is a Wikipedia-sense annotated corpus which was constructed by exploiting Wikipedia hyperlinks, propagating them across Wikipedia pages. Its English version comprises over 160M sense annotations with an estimated precision over 90%.

2.3. BabelNet

BabelNet (Navigli and Ponzetto, 2012) is a wide-coverage multilingual semantic network obtained from the integration of various encyclopedias and dictionaries (WordNet and Wikipedia, inter alia). Being a superset of all these resources, BabelNet brings together lexicographic and encyclopedic knowledge, thus containing both named entities and concepts, and, unlike Wikipedia covering only noun instances, instances have diverse Part-Of-Speech (PoS) tags: nouns, verbs, adjectives and adverbs. Given its multilingual nature (i.e. BabelNet covers over 250 languages), BabelNet has been used as a sense inventory for annotating text in languages other than English.

SenseDefs. SenseDefs[8] (Camacho-Collados et al., 2019) extends the effort from the Princeton WordNet Gloss Corpus project (see Section 2.1) by automatically disambiguating textual definitions from various heterogeneous sources in 263 languages. The underlying idea lies on leveraging the cross-complementarities of definitions of identical concepts from different languages and resources. The approach couples a graph-based disambiguation method (Moro et al., 2014) with a refinement based on distributional similarity (Camacho-Collados et al., 2016). The proposed method was evaluated on four European languages (English, Spanish, French and Italian) with an estimated precision of over 80%.

EuroSense. The construction of EuroSense[9] (Delli Bovi et al., 2017) follows a similar approach to SenseDefs. In this case, parallel corpora is exploited for a single multilingual disambiguation. The output is a sense-annotated corpus for 21 languages for the Europarl parallel corpus (Koehn, 2005). The estimated precision for four languages isover 80% on average, with a peak of almost 90% for German.

Train-o-Matic. Similarly to the previous approach, Train-o-Matic[10] (Pasini and Navigli, 2017, T-o-M) aims at automatically annotating words from a raw corpus with senses. The main difference with respect to EuroSense and OMSTI lies in the fact that T-o-M does not need parallel data in order to annotate the input corpus. While being language independent and fully automatic, it proved to lead supervised systems to high performance, close or even better than those achieved when a manually annotated corpus (e.g. SemCor) is used for training. Moreover, it has also proved effective in languages other than English (Pasini et al., 2018): Italian, Spanish, French, German and Chinese.

OneSeC. OneSeC[11] (Scarlini et al., 2019) is the most recent work among those aiming at automatically producing semantically-annotated data. Instead of the well-known "one sense per discourse" assumption made by Gale et al. (1992), this work makes a more relaxed hypotesis, i.e., "one sense per Wikipedia Category". That is, a noun is used always with the same meaning within a Wikipedia Category. By leveraging this conjecture, OneSeC exploits the texts contained within the pages of a given Wikipedia Category to annotate each noun occurrence therein with its most suitable meaning. The corpora for English showed to be of high-quality, leading a supervised English WSD model, i.e., It Makes Sense (Zhong and Ng, 2010, IMS), to achieve results that are higher than those attained by IMS trained on other automatically generated corpora. Furthermore, OneSeC has been used to generate annotated data for four other European languages, namely: Italian, Spanish, German and French.

3. Statistics

In order to have a global overview of all sense-annotated corpora, their main features are displayed in Table 1. For each corpus we include its underlying resource, number of languages covered and total number of sense annotations. In general the datasets are quite heterogeneous in nature, coming from three different resources and constructed via four different strategies: manual, semi-automatic, automatic and collaborative. The number of sense annotations also varies depending on the resource, with Wikipedia- and BabelNet-based corpora contributing with the highest number of annotations. This is correlated with the coverage of these resources: Wikipedia and BabelNet are two orders of magnitude higher than WordNet.

In addition to these global statistics, Table 1 shows local statistics (i.e. number of tokens, number of sense annotations, ambiguity level and entropy) for English, which is the only language covered by all corpora.[12] The ambiguity level of each dataset is computed as the average number of candidate senses per instance (i.e., senses with the same surface form of a target word). The average entropy score

[5] Note that more Wikipedia sense-annotated datasets extracted from the Wikilinks project exist (Singh et al., 2012; Eshel et al., 2017). However, due to privacy and license issues, these datasets cannot be shared directly. Please also refer to Usbeck et al. (2015) for an overview and unification of datasets focused on Entity Linking.

[6] We considered the Wikipedia dumps of November 2014.

[7] http://lcl.uniroma1.it/sew

[8] http://lcl.uniroma1.it/sensedefs

[9] http://lcl.uniroma1.it/eurosense

[10] http://trainomatic.org

[11] http://trainomatic.org/onesec

[12] Due to license restrictions we could not access OntoNotes' full corpus for computing its ambiguity/entropy.

	Resource	Type	#Langs	#Annotations	English			
					#Tokens	#Annotations	Amb	Entropy
SemCor	WordNet	M	1	226,036	802,443	226,036	6.8	0.27
MASC-WSA	WordNet	M	1	1,084,551	1,309,838	1,084,551	8.8	2.09
SemEval-ALL	WordNet	M	1	7,253	25,503	7,253	5.8	0.18
OntoNotes	WordNet	M	1	264,622	1,445,000	264,622	-	-
Princeton Gloss	WordNet	SA	1	449,355	1,621,129	449,355	3.8	0.45
OMSTI	WordNet	SA	1	911,134	30,441,386	911,134	8.9	0.94
Wiki-hypers	Wikipedia	C	271	321,718,966	1,357,105,761	71,457,658	2.6	0.44
SEW	Wikipedia	SA	1	162,614,753	1,357,105,761	162,614,753	7.9	0.40
SenseDefs	BabelNet	A	263	163,029,131	71,109,002	37,941,345	4.6	0.04
EuroSense	BabelNet	A	21	122,963,111	48,274,313	15,502,847	6.5	0.21
T-o-M	BabelNet	A	6	17,987,488	291,550,966	12,722,530	3.6	0.48
OneSeC	BabelNet	A	5	1,222,090	25,017,839	888,417	3.6	0.51

Table 1: Statistics of the sense-annotated corpora across languages and resources. Type "M" stands for Manual, "SA" stands Semi-automatic, "C" for Collaborative and "A" for Automatic.

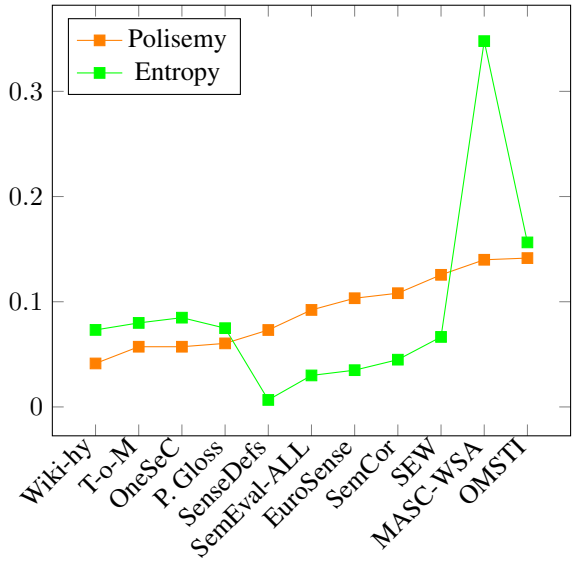

Figure 2: Normalized entropy and ambiguity levels across datasets.

is first computed separately for each word sense distribution of frequencies, and then averaged by the number of unique lemmas in the dataset (Pasini and Navigli, 2018). These two measures are complementary: while the average ambiguity is a precise indicator of the number of senses associated to a word within the corpus, the average entropy represent the skewness of word senses distributions, with higher entropy scores meaning that the sense distributions are closer to the uniform distribution.

4. Analysis

To gain insights on the features of each sense-annotated corpora, we provide a small analysis on the entropy and ambiguity levels. Figure 2 shows the average ambiguity of a dataset (x axis) with respect to the average entropy (y axis) of the lemma annotations therein[13]. As can be observed, datasets with higher degree of ambiguity tend to be also more entropic with MASC-WSA being the corpus with the least skewed sense distributions.[14] On the other hand, datasets with lower levels of ambiguity tend to have more unbalanced distributions and consequently a lower degree of entropy. For instance, EuroSense, which was automatically-constructed, have the most similar entropy to that of SemCor and SemEval datasets, which were manually-curated. We note that SenseDefs is the dataset with the lowest entropy. Going more in-depth we observed that, due to its nature, SenseDefs contains a large number of unambiguous named entities, i.e., containing a single sense in its underlying sense inventory BabelNet.

5. Conclusion

In this paper we have presented an overview of available sense-annotated datasets for WordNet, Wikipedia and BabelNet, and for various languages. These datasets correspond to a wide variety of approaches, from manual construction to automatic or semi-automatic methods. By listing and providing statistics for all these datasets we are pursuing two main goals: (1) motivating and providing information about sense-annotated corpora to be used for research purposes, and (2) highlighting the main properties of the various sense-annotated corpora across resources. Moreover, this paper represents a first step for obtaining a fully-integrated repository of sense-annotated corpora which can be easily leveraged for research and evaluation purposes. As future work it would be interesting to integrate these sense-annotated resources into a unified multilingual repository, following the lines of Raganato et al.

[13] Average ambiguity and entropy values shown in Table 1 are normalized by the sum of all ambiguity or entropy values.

[14] The high entropy of MASC-WSA may also be attributed to the specific selection of the vocabulary, which was decided by a committee from WordNet and FrameNet management teams.

(2017a) and Vial et al. (2018) for WordNet sense-annotated corpora in English.

Acknowledgments

The authors would like to thank Claudio Delli Bovi and Alessandro Raganato for interesting discussions and their input with some of the datasets.

Tommaso Pasini gratefully acknowledges the support of the ERC Consolidator Grant MOUSSE No. 726487 under the European Union's Horizon 2020 research and innovation programme.

6. Bibliographical References

Agirre, E. and Soroa, A. (2009). Personalizing PageRank for Word Sense Disambiguation. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, Athens, Greece, 30 March–3 April 2009*, pages 33–41.

Agirre, E., De Lacalle, O. L., Fellbaum, C., Marchetti, A., Toral, A., and Vossen, P. (2009). Semeval-2010 task 17: All-words word sense disambiguation on a specific domain. In *Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions*, pages 123–128. Association for Computational Linguistics.

Agirre, E., de Lacalle, O. L., and Soroa, A. (2014). Random walks for knowledge-based word sense disambiguation. *Computational Linguistics*, 40(1):57–84.

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 (COLING-ACL '98)*, pages 86–90, Montreal, Canada.

Baldwin, T., Su, N. K., Bond, F., Fujita, S., Martinez, D., and Tanaka, T. (2008). Mrd-based word sense disambiguation: Further extending lesk. In *Proceedings of International Joint Conference on Natural Language Processing*, pages 775–780.

Butnaru, A., Ionescu, R. T., and Hristea, F. (2017). Shotgunwsd: An unsupervised algorithm for global word sense disambiguation inspired by dna sequencing. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 916–926, Valencia, Spain. Association for Computational Linguistics.

Camacho-Collados, J., Pilehvar, M. T., and Navigli, R. (2015). A unified multilingual semantic representation of concepts. In *Proceedings of ACL*, pages 741–751, Beijing, China, July.

Camacho-Collados, J., Pilehvar, M. T., and Navigli, R. (2016). Nasari: Integrating explicit knowledge and corpus statistics for a multilingual representation of concepts and entities. *Artificial Intelligence*, 240:36–64.

Camacho-Collados, J., Bovi, C. D., Raganato, A., and Navigli, R. (2019). Sensedefs: a multilingual corpus of semantically annotated textual definitions. *Language Resources and Evaluation*, 53(2):251–278.

Chan, Y. S. and Ng, H. T. (2005). Scaling up word sense disambiguation via parallel texts. In *AAAI*, volume 5, pages 1037–1042.

Chaplot, D. S. and Salakhutdinov, R. (2018). Knowledge-based word sense disambiguation using topic models. In *Proceedings of AAAI*.

De Melo, G., Baker, C. F., Ide, N., Passonneau, R. J., and Fellbaum, C. (2012). Empirical comparisons of masc word sense annotations. In *LREC*, pages 3036–3043.

Delli Bovi, C., Camacho-Collados, J., Raganato, A., and Navigli, R. (2017). EuroSense: Automatic harvesting of multilingual sense annotations from parallel text. In *Proc.of ACL*, volume 2, pages 594–600.

Edmonds, P. and Cotton, S. (2001). Senseval-2: overview. In *Proc. of SensEval 2*, pages 1–5. ACL.

Eisele, A. and Chen, Y. (2010). MultiUN: A Multilingual Corpus from United Nation Documents. In *Proceedings of the Seventh conference on International Language Resources and Evaluation*, pages 2868–2872.

Eshel, Y., Cohen, N., Radinsky, K., Markovitch, S., Yamada, I., and Levy, O. (2017). Named entity disambiguation for noisy text. In *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, pages 58–68. Association for Computational Linguistics.

Fellbaum, C. (1998). *WordNet: An Electronic Database*. MIT Press, Cambridge, MA.

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

Gale, W. A., Church, K. W., and Yarowsky, D. (1992). One sense per discourse. In *Proceedings of the workshop on Speech and Natural Language*, pages 233–237. Association for Computational Linguistics.

González, A., Rigau, G., and Castillo, M. (2012). A graph-based method to improve Wordnet domains. In *Proceedings of 13th International Conference on Intelligent Text Processing and Computational Linguistics (CICLING)*, pages 17–28, New Delhi, India.

Henrich, V., Hinrichs, E., and Vodolazova, T. (2012). Webcage: a web-harvested corpus annotated with germanet senses. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 387–396. Association for Computational Linguistics.

Huang, L., Sun, C., Qiu, X., and Huang, X.-J. (2019). Glossbert: Bert for word sense disambiguation with gloss knowledge. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3500–3505.

Iacobacci, I., Pilehvar, M. T., and Navigli, R. (2015).

SensEmbed: Learning Sense Embeddings for Word and Relational Similarity. In *Proc. of ACL*, pages 95–105.

Iacobacci, I., Pilehvar, M. T., and Navigli, R. (2016). Embeddings for word sense disambiguation: An evaluation study. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pages 897–907.

Ide, N., Erjavec, T., and Tufiş, D. (2002). Sense discrimination with parallel corpora. In *Proceedings of the ACL-02 Workshop on WSD: Recent Successes and Future Directions, Philadelphia, Penn., July 2002*, pages 54–60.

Ide, N., Baker, C., Fellbaum, C., Fillmore, C., and Passonneau, R. (2008). Masc: The manually annotated sub-corpus of american english. In *6th International Conference on Language Resources and Evaluation, LREC 2008*, pages 2455–2460. European Language Resources Association (ELRA).

Ide, N., Baker, C., Fellbaum, C., and Passonneau, R. (2010). The manually annotated sub-corpus: A community resource for and by the people. In *Proceedings of the ACL 2010 Conference Short Papers*, pages 68–73, Uppsala, Sweden, July. Association for Computational Linguistics.

Kågebäck, M. and Salomonsson, H. (2016). Word Sense Disambiguation using a Bidirectional LSTM. In *Proc. of CogALex*, pages 51–56.

Koehn, P. (2005). Europarl: A parallel corpus for statistical machine translation. In *Proceedings of Machine Translation Summit X*.

Kucera, H. and Francis, W. (1979). A standard corpus of present-day edited american english, for use with digital computers (revised and amplified from 1967 version).

Loureiro, D. and Jorge, A. (2019). Language modelling makes sense: Propagating representations through WordNet for full-coverage word sense disambiguation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5682–5691, Florence, Italy, July. Association for Computational Linguistics.

Luo, F., Liu, T., Xia, Q., Chang, B., and Sui, Z. (2018). Incorporating glosses into neural word sense disambiguation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 2473–2482.

Mancini, M., Camacho-Collados, J., Iacobacci, I., and Navigli, R. (2017). Embedding words and senses together via joint knowledge-enhanced training. In *Proceedings of CoNLL*, pages 100–111, Vancouver, Canada.

Melamud, O., Goldberger, J., and Dagan, I. (2016). context2vec: Learning generic context embedding with bidirectional lstm. In *Proceedings of CONLL*, pages 51–61.

Miller, G. A., Leacock, C., Tengi, R., and Bunker, R. (1993a). A semantic concordance. In *Proceedings of the 3rd DARPA Workshop on Human Language Technology*, pages 303–308, Plainsboro, N.J.

Miller, G. A., Leacock, C., Tengi, R., and Bunker, R. (1993b). A semantic concordance. In *Proceedings of the 3rd DARPA Workshop on Human Language Technology*, pages 303–308, Plainsboro, N.J.

Moro, A. and Navigli, R. (2015). Semeval-2015 task 13: Multilingual all-words sense disambiguation and entity linking. In *Proc. of SemEval-2015*.

Moro, A., Raganato, A., and Navigli, R. (2014). Entity Linking meets Word Sense Disambiguation: a Unified Approach. *Transaction of ACL (TACL)*, 2:231–244.

Navigli, R. and Ponzetto, S. P. (2012). BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250.

Navigli, R., Jurgens, D., and Vannella, D. (2013). Semeval-2013 task 12: Multilingual word sense disambiguation. In *Proc. of SemEval 2013*, pages 222–231, Atlanta, USA.

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

Otegi, A., Aranberri, N., Branco, A., Hajic, J., Neale, S., Osenova, P., Pereira, R., Popel, M., Silva, J., Simov, K., and Agirre, E. (2016). QTLeap WSD/NED Corpora: Semantic Annotation of Parallel Corpora in Six Languages. In *Proc. of LREC*, pages 3023–3030.

Pasini, T. and Navigli, R. (2017). Train-o-matic: Large-scale supervised word sense disambiguation inmultiple languages without manual training data. In *Proceedings of Empirical Methods in Natural Language Processing*, Copenhagen, Denmark.

Pasini, T. and Navigli, R. (2018). Two knowledge-based methods for high-performance sense distribution learning. *Proceedings of AAAI, New Orleans, United States*.

Pasini, T., Elia, F. M., and Navigli, R. (2018). Huge Automatically Extracted Training Sets for Multilingual Word Sense Disambiguation. In *Proceedings of LREC*, Miyazaki, Japan.

Petrolito, T. and Bond, F. (2014). A survey of wordnet annotated corpora. In *Proceedings Global WordNet Conference, GWC-2014*, pages 236–245.

Pilehvar, M. T., Jurgens, D., and Navigli, R. (2013). Align, disambiguate and walk: A unified approach for measuring semantic similarity. In *Proc. of ACL*, pages 1341–1351.

Pradhan, S. S., Loper, E., Dligach, D., and Palmer, M. (2007). Semeval-2007 task 17: English lexical sample, srl and all words. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 87–92.

Raganato, A., Delli Bovi, C., and Navigli, R. (2016). Automatic Construction and Evaluation of a Large Semantically Enriched Wikipedia. In *Proceedings of IJCAI*, pages 2894–2900, New York City, NY, USA, July.

Raganato, A., Camacho-Collados, J., and Navigli, R. (2017a). Word Sense Disambiguation: A Unified Evaluation Framework and Empirical Comparison. In *Proc. of EACL*, pages 99–110, Valencia, Spain.

Raganato, A., Delli Bovi, C., and Navigli, R. (2017b). Neural sequence learning models for word sense disambiguation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1167–1178. Association for Computational Linguistics.

Scarlini, B., Pasini, T., and Navigli, R. (2019). Just

"OneSeC" for Producing Multilingual Sense-Annotated Data. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 699–709.

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

Singh, S., Subramanya, A., Pereira, F., and McCallum, A. (2012). Wikilinks: A large-scale cross-document coreference corpus labeled via links to wikipedia. *University of Massachusetts, Amherst, Technical Report UM-CS-2012- 015*.

Snyder, B. and Palmer, M. (2004). The english all-words task. In *Proc. of Senseval-3*, pages 41–43, Barcelona, Spain.

Taghipour, K. and Ng, H. T. (2015a). One million sense-tagged instances for word sense disambiguation and induction. *CoNLL 2015*, pages 338–344.

Taghipour, K. and Ng, H. T. (2015b). Semi-Supervised Word Sense Disambiguation Using Word Embeddings in General and Specific Domains. *Proc. of NAACL-HLT*, pages 314–323.

Usbeck, R., Röder, M., Ngonga Ngomo, A.-C., Baron, C., Both, A., Brümmer, M., Ceccarelli, D., Cornolti, M., Cherix, D., Eickmann, B., et al. (2015). Gerbil: general entity annotator benchmarking framework. In *Proceedings of the 24th International Conference on World Wide Web*, pages 1133–1143. International World Wide Web Conferences Steering Committee.

Vial, L., Lecouteux, B., and Schwab, D. (2018). UFSAC: Unification of Sense Annotated Corpora and Tools. In Nicoletta Calzolari (Conference chair), et al., editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 7-12, 2018. European Language Resources Association (ELRA).

Weischedel, R., Palmer, M., Marcus, M., Hovy, E., Pradhan, S., Ramshaw, L., Xue, N., Taylor, A., Kaufman, J., Franchini, M., et al. (2013). Ontonotes release 5.0. *Linguistic Data Consortium, Philadelphia, PA*.

Yuan, D., Richardson, J., Doherty, R., Evans, C., and Altendorf, E. (2016). Semi-supervised word sense disambiguation with neural models. *Proceedings of COLING*, pages 1374–1385.

Zhong, Z. and Ng, H. T. (2010). It makes sense: A wide-coverage word sense disambiguation system for free text. In *Proc. of of the ACL*, pages 78–83, Uppsala, Sweden. ACL.