A Short Survey on Sense-Annnotated Corpora

Tommaso Pasini  
Department of Informatics  
Sapienza University of Rome, IT  
pasini@di.uniroma1.it

Jose Camacho-Collados  
School of Computer Science  
and Informatics  
Cardiff University, UK  
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 currently available sense-annotated corpora, both manually and (semi)automatically constructed, for diverse languages and lexical resources (i.e. WordNet, Wikipedia, BabelNet). General statistics and specific features of each sense-annotated dataset are also provided.

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. 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 Salomonsen, 2016) have been more effective in terms of performance when sense-annotated corpora are available (Raganato et al., 2017a). Unfortunately, obtaining such data is heavily time-consuming and expensive (Schubert, 2006), and reasonable amounts of sense-annotated corpora tend to be available for English only. This produces the so-called knowledge-acquisition bottleneck (Gale et al., 1992).

The first main approach towards building sense-annotated corpora was SemCor (Miller et al., 1993), providing annotations for the WordNet sense inventory (Fellbaum, 1998). Since then, several semi-automatic and automatic approaches have also been proposed. 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 andNavigli, 2017), as well as to learn effective embedded representations for senses (Iacobacci et al., 2015; Flekova and Gurevych, 2016).

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. Finally, we provide a general overview and statistics of these sense-annotated resources, providing relevant details across resources and languages.

2 Sense-Annnotated 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 gives an overview of these resources and their underlying corpora.

2.1 WordNet
WordNet (Fellbaum, 1998) has been one of the most widely used knowledge resources in lexi-
In fact, it has been the de-facto sense inventory for Word Sense Disambiguation for 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., 1993). SemCor was manually annotated and consists of 352 documents from the Brown Corpus (Kucera and Francis, 1979) and 226,040 sense annotations. SemCor has been the largest manually-annotated corpus for many years, and is the main corpus used in the literature to train supervised WSD systems (Agirre et al., 2009; Zhong and Ng, 2010; Raganato et al., 2017b; Luo et al., 2018).

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 in the main been used as evaluation benchmarks for WSD systems, contain a total of 7,253 sense annotations.

Princeton WordNet Gloss. The Princeton WordNet Gloss Corpus\(^2\) is a sense-annotated corpus of textual definitions (glosses) from WordNet synsets. The corpus was tagged semi-automatically: 330,499 manually sense instances were annotated manually while the remaining annotations (i.e. 118,856) were obtained automatically. This corpus of disambiguated glosses has already been proved 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\(^3\) (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 consisting of concepts and entities and their corresponding pages. 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 two datasets providing disambiguations in the form of

---

\(^2\)http://wordnet.princeton.edu/glosstag.shtml

\(^3\)http://www.comp.nus.edu.sg/~nlp/corpora.html
Wikipedia pages. For these two datasets we have used the same version of Wikipedia for a more accurate comparison.

**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 (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 coming from 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 (Camacho-Collados et al., 2016a) 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., 2016b). 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 (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 with figures over 80% on average, with a peak of almost 90% for German.

**Train-o-Matic.** Similarly to the previous approach, Train-o-Matic (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. Being language independent and fully automatic, it has been 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.

### 3 Statistics

In order to have a global overview of all sense-annotated corpora, the main features of each sense-annotated corpus 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. The ambiguity level of each

---

4Note 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.

5We used the Wikipedia dumps of November 2014.

6http://lcl.uniroma1.it/sew

7http://lcl.uniroma1.it/sensedefs

8http://lcl.uniroma1.it/eurosense

9http://trainomatic.org

10Due to license restrictions we could not access
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.

| Resource     | Type    | #Langs | #Annotations | #Tokens | #Annotations | Amb | Entropy |
|--------------|---------|--------|-------------|---------|-------------|-----|---------|
| SemCor       | WordNet | M      | 1           | 226,036 | 802,443      | 6.8 | 0.27    |
| SemEval-ALL  | WordNet | M      | 1           | 7,253   | 25,503       | 5.8 | 0.18    |
| OntoNotes    | WordNet | M      | 1           | 264,622 | 1,445,000    | -   | -       |
| Princeton Gloss | WordNet | SA     | 1           | 449,355 | 1,621,129   | 3.8 | 0.45    |
| OMSTI        | WordNet | SA     | 1           | 911,134 | 30,441,386  | 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    |

As can be seen in Figure 2, where we renormalized the average polysemy (i.e. ambiguity level) and the average entropy scores, datasets with lower or higher degree of polysemy tend to be also more entropic. On the other hand, datasets with lower levels of ambiguity tend to have more skewed distributions and consequently a lower entropy. For instance, EuroSense, which was automatically-constructed, have the most similar entropy to that of SemCor and SemEval datasets, which were manually-curated. On the other hand, OMSTI is the corpus with higher entropy. This finding seems reasonable inasmuch the approach is limited to adding examples to words that are already covered by SemCor without any control over the number of new sentences added to each word sense. In the other extreme, we note that SenseDefs is the dataset with the lowest entropy. Going more in-depth we observed that SenseDefs contains many unambiguous named entities, i.e., containing a single sense in its underlying sense inventory BabelNet.

4 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. (2017a) for WordNet sense-annotated corpora in English.

Acknowledgments

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

References

Eneko Agirre, Oier Lopez De Lacalle, Christiane Fellbaum, Andrea Marchetti, Antonio Toral, and Piek Vossen. 2009. Senseval-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.

Eneko Agirre, Oier Lopez de Lacalle, and Aitor Soroa. 2014. Random walks for knowledge-based word sense disambiguation. Computational Linguistics, 40(1):57–84.

Eneko Agirre and Aitor Soroa. 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.

Timothy Baldwin, Nam Kim Su, Francis Bond, Sanae Fujita, David Martinez, and Takaaki Tanaka. 2008. Mrd-based word sense disambiguation: Further extending leks. In Proceedings of International Joint Conference on Natural Language Processing, pages 775–780.

Andrei Butnaru, Radu Tudor Ionescu, and Florentina Hriatea. 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.

José Camacho-Collados, Claudio Delli Bovi, Alessandro Raganato, and RobertoNavigli. 2016a. A Large-Scale Multilingual Disambiguation of Glosses. In Proceedings of LREC, pages 1701–1708, Portoroz, Slovenia.

José Camacho-Collados, Mohammad Taher Pilehvar, and RobertoNavigli. 2015. A unified multilingual semantic representation of concepts. In Proceedings of ACL, pages 741–751, Beijing, China.

José Camacho-Collados, Mohammad Taher Pilehvar, and RobertoNavigli. 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 AAAI, volume 5, pages 1037–1042.

Devendra Singh Chaplot and Ruslan Salakhutdinov. 2018. Knowledge-based word sense disambiguation using topic models. In Proceedings of AAAI.

Claudio Delli Bovi, Jose Camacho-Collados, Alessandro Raganato, and RobertoNavigli. 2017. EuroSense: Automatic harvesting of multilingual sense annotations from parallel text. In Proc.of ACL, volume 2, pages 594–600.

Philip Edmonds and Scott Cotton. 2001. Senseval-2: overview. In Proc. of SensEval 2, pages 1–5. ACL.

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

Yotam Eshel, Noam Cohen, Kira Radinsky, Shaul Markovitch, Ikuya Yamada, and Omer Levy. 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.

Christiane Fellbaum. 1998. WordNet: An Electronic Database. MIT Press, Cambridge, MA.

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

William A. Gale, Kenneth Church, and David Yarowsky. 1992. A method for disambiguating word senses in a corpus. Computers and the Humanities, 26:415–439.

Aitor González, German Rigau, and Mauro Castillo. 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.

Verena Henrich, Erhard Hinrichs, and Tatiana Vodolaza. 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.

Ignacio Iacobucci, Mohammad Taher Pilehvar, and RobertoNavigli. 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 Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 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 Proceedings of Machine Translation Summit X.

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

Fuli Luo, Tianyu Liu, Qiaolin Xia, Baobao Chang, and Zhifang Sui. 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.

Oren Melamud, Jacob Goldberger, and Ido Dagan. 2016. context2vec: Learning generic context embedding with bidirectional lstm. In Proceedings of CONLL, pages 51–61.

George A. Miller, Claudia Leacock, Randee Tengi, and Ross Bunker. 1993. A semantic concordance. In Proceedings of the 3rd DARPA Workshop on Human Language Technology, pages 303–308, Plainsboro, N.J.

Andrea Moro and Roberto Navigli. 2015. Semeval-2015 task 13: Multilingual all-words sense disambiguation and entity linking. In Proc. of SemEval-2015.

Andrea Moro, Alessandro Raganato, and Roberto Navigli. 2014. Entity Linking meets Word Sense Disambiguation: a Unified Approach. Transaction of ACL (TAACL), 2:231–244.

Roberto Navigli, David Jurgens, and Daniele Vannella. 2013. Semeval-2013 task 12: Multilingual word sense disambiguation. In Proc. of SemEval 2013, pages 222–231, Atlanta, USA.

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.

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

Tommaso Pasini, Francesco Maria Elia, and Roberto Navigli. 2018. Huge Automatically Extracted Training Sets for Multilingual Word Sense Disambiguation. In Proceedings of LREC, Miyazaki, Japan.

Tommaso Pasini and Roberto Navigli. 2017. Train-o-matic: Large-scale supervised word sense disambiguation in multiple languages without manual training data. In Proceedings of Empirical Methods in Natural Language Processing, Copenhagen, Denmark.

Tommaso Pasini and Roberto Navigli. 2018. Two knowledge-based methods for high-performance sense distribution learning. Proceedings of AAAI, New Orleans, United States.

Tommaso Petrolito and Francis Bond. 2014. A survey of wordnet annotated corpora. In Proceedings Global WordNet Conference, GWC-2014, pages 236–245.

Mohammad Taher Pilehvar, David Jurgens, and Roberto Navigli. 2013. Align, disambiguate and walk: A unified approach for measuring semantic similarity. In Proc. of ACL, pages 1341–1351.

Sameer S Pradhan, Edward Loper, Dmitriy Dligach, and Martha Palmer. 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.

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

Alessandro Raganato, Claudio Delli Bovi, and Roberto Navigli. 2016. Automatic Construction and Evaluation of a Large Semantically Enriched Wikipedia. In Proceedings of IJCAI, pages 2894–2900, New York City, NY, USA.

Alessandro Raganato, Claudio Delli Bovi, and Roberto Navigli. 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.

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

Sameer Singh, Amarnag Subramanyam, Fernando Pereira, and Andrew McCallum. 2012. Wikilinks: A large-scale cross-document coreference corpus labeled via links to wikipedia. University of Massachusetts, Amherst, Technical Report UM-CS-2012-015.

Benjamin Snyder and Martha Palmer. 2004. The english all-words task. In Proc. of Senseval-3, pages 41–43, Barcelona, Spain.
Kaveh Taghipour and Hwee Tou Ng. 2015a. One million sense-tagged instances for word sense disambiguation and induction. *CoNLL 2015*, 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.

Ricardo Usbeck, Michael Röder, Axel-Cyrille Ngonga Ngomo, Ciro Baron, Andreas Both, Martin Brümmer, Diego Ceccarelli, Marco Cornolti, Didier Cherix, Bernd Eickmann, 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.

Ralph Weischedel, Martha Palmer, Mitchell Marcus, Eduard Hovy, Sameer Pradhan, Lance Ramshaw, Nianwen Xue, Ann Taylor, Jeff Kaufman, Michelle Franchini, et al. 2013. Ontonotes release 5.0. *Linguistic Data Consortium, Philadelphia, PA*.

Dayu Yuan, Julian Richardson, Ryan Doherty, Colin Evans, and Eric Altendorf. 2016. Semi-supervised word sense disambiguation with neural models. *Proceedings 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 the ACL*, pages 78–83, Uppsala, Sweden. ACL.