Using Linked Disambiguated Distributional Networks for Word Sense Disambiguation

Alexander Panchenko[‡], Stefano Faralli[†], Simone Paolo Ponzetto[†], and Chris Biemann[‡]

[‡]Language Technology Group, Computer Science Dept., University of Hamburg, Germany
[†]Web and Data Science Group, Computer Science Dept., University of Mannheim, Germany
{panchenko,biemann}@informatik.uni-hamburg.de
{faralli,simone}@informatik.uni-mannheim.de

Abstract

We introduce a new method for unsupervised knowledge-based word sense disambiguation (WSD) based on a resource that links two types of sense-aware lexical networks: one is induced from a corpus using distributional semantics, the other is manually constructed. The combination of two networks reduces the sparsity of sense representations used for WSD. We evaluate these enriched representations within two lexical sample sense disambiguation benchmarks. Our results indicate that (1) features extracted from the corpus-based resource help to significantly outperform a model based solely on the lexical resource; (2) our method achieves results comparable or better to four state-of-the-art unsupervised knowledge-based WSD systems including three hybrid systems that also rely on text corpora. In contrast to these hybrid methods, our approach does not require access to web search engines, texts mapped to a sense inventory, or machine translation systems.

1 Introduction

The representation of word senses and the disambiguation of lexical items in context is an on-going long-established branch of research (Agirre and Edmonds, 2007; Navigli, 2009). Traditionally, word senses are defined and represented in lexical resources, such as WordNet (Fellbaum, 1998), while more recently, there is an increased interest in approaches that induce word senses from corpora using graph-based distributional approaches (Dorow and Widdows, 2003; Biemann, 2006; Hope and Keller, 2013), word sense embeddings (Neelakantan et al., 2014; Bartunov et al.,

2016) and combination of both (Pelevina et al., 2016). Finally, some hybrid approaches emerged, which aim at building sense representations using information from both corpora and lexical resources, e.g. (Rothe and Schütze, 2015; Camacho-Collados et al., 2015a; Faralli et al., 2016). In this paper, we further explore the last strain of research, investigating the utility of hybrid sense representation for the word sense disambiguation (WSD) task.

In particular, the contribution of this paper is a new unsupervised knowledge-based approach to WSD based on the *hybrid aligned resource* (HAR) introduced by Faralli et al. (2016). The key difference of our approach from prior hybrid methods based on sense embeddings, e.g. (Rothe and Schütze, 2015), is that we rely on sparse lexical representations that make the sense representation readable and allow to straightforwardly use this representation for word sense disambiguation, as will be shown below. In contrast to hybrid approaches based on sparse interpretable representations, e.g. (Camacho-Collados et al., 2015a), our method requires no mapping of texts to a sense inventory and thus can be applied to larger text collections. By linking symbolic distributional sense representations to lexical resources, we are able to improve representations of senses, leading to performance gains in word sense disambiguation.

2 Related Work

Several prior approaches combined distributional information extracted from text (Turney and Pantel, 2010) from text with information available in lexical resources, such as WordNet. Yu and Dredze (2014) proposed a model to learn word embeddings based on lexical relations of words from WordNet and PPDB (Ganitkevitch et al., 2013). The objective function of their model

combines the objective function of the skip-gram model (Mikolov et al., 2013) with a term that takes into account lexical relations of a target word. Faruqui et al. (2015) proposed a related approach that performs a post-processing of word embeddings on the basis of lexical relations from the same resources. Pham et al. (2015) introduced another model that also aim at improving word vector representations by using lexical relations from WordNet. The method makes representations of synonyms closer than representations of antonyms of the given word. While these three models improve the performance on word relatedness evaluations, they do not model word senses. Jauhar et al. (2015) proposed two models that tackle this shortcoming, learning sense embeddings using the word sense inventory of WordNet.

Iacobacci et al. (2015) proposed to learn sense embeddings on the basis of the BabelNet lexical ontology (Navigli and Ponzetto, 2012). Their approach is to train the standard skip-gram model on a pre-disambiguated corpus using the Babelfy WSD system (Moro et al., 2014). NASARI (Camacho-Collados et al., 2015a) relies on Wikipedia and WordNet to produce vector representations of senses. In this approach, a sense is represented in lexical or sense-based feature spaces. The links between WordNet and Wikipedia are retrieved from BabelNet. MUFFIN (Camacho-Collados et al., 2015b) adapts several ideas from NASARI, extending the method to the multi-lingual case by using BabelNet synsets instead of monolingual WordNet synsets.

The approach of Chen et al. (2015) to learning sense embeddings starts from initialization of sense vectors using WordNet glosses. It proceeds by performing a more conventional context clustering, similar what is found to unsupervised methods such as (Neelakantan et al., 2014; Bartunov et al., 2016).

Rothe and Schütze (2015) proposed a method that learns sense embedding using word embeddings and the sense inventory of WordNet. The approach was evaluated on the WSD tasks using features based on the learned sense embeddings.

Goikoetxea et al. (2015) proposed a method for learning word embeddings using random walks on a graph of a lexical resource. Nieto Piña and Johansson (2016) used a similar approach based on random walks on a WordNet to learn sense embeddings.

All these diverse contributions indicate the benefits of hybrid knowledge sources for learning word and sense representations.

3 Unsupervised Knowledge-based WSD using Hybrid Aligned Resource

We rely on the hybrid aligned lexical semantic resource proposed by Faralli et al. (2016) to perform WSD. We start with a short description of this resource and then discuss how it is used for WSD.

3.1 Construction of the Hybrid Aligned Resource (HAR)

The hybrid aligned resource links two lexical semantic networks using the method of Faralli et al. (2016): a corpus-based distributionally-induced network and a manually-constructed network. Sample entries of the HAR are presented in Table 1. The corpus-based part of the resource, called *proto-conceptualization* (PCZ), consists of sense-disambiguated lexical items (PCZ ID), disambiguated related terms and hypernyms, as well as context clues salient to the lexical item. The knowledge-based part of the resource, called *conceptualization*, is represented by synsets of the lexical resource and relations between them (WordNet ID). Each sense in the PCZ network is subsequently linked to a sense of the knowledge-based network based on their similarity calculated on the basis of lexical representations of senses and their neighbors. The construction of the PCZ involves the following steps (Faralli et al., 2016):

Building a Distributional Thesaurus (DT). At this stage, a similarity graph over terms is induced from a corpus, where each entry consists of the most similar 200 terms for a given term using the JoBimText method (Biemann and Riedl, 2013).

Word Sense Induction. In DTs, entries of polysemous terms are mixed, i.e. they contain related terms of several senses. The Chinese Whispers (Biemann, 2006) graph clustering is applied to the ego-network (Everett and Borgatti, 2005) of the each term, as defined by its related terms and connections between then observed in the DT to derive word sense clusters.

Labeling Word Senses with Hypernyms. Hearst (1992) patterns are used to extract hypernyms from the corpus. These hypernyms are assigned to senses by aggregating hypernym

PCZ ID	WordNet ID	Related Terms	Hypernyms	Context Clues
mouse:0	mouse:1	rat:0, rodent:0, monkey:0, ...	animal:0, species:1, ...	rat:conj_and, white-footed:amod, ...
mouse:1	mouse:4	keyboard:1, computer:0, printer:0 ...	device:1, equipment:3, ...	click:-prep_of, click:-nn, ...
keyboard:0	keyboard:1	piano:1, synthesizer:2, organ:0 ...	instrument:2, device:3, ...	play:-dobj, electric:amod, ..
keyboard:1	keyboard:1	keypad:0, mouse:1, screen:1 ...	device:1, technology:0 ...	computer, qwerty:amod ...

Table 1: Sample entries of the hybrid aligned resource (HAR) for the words "mouse" and "keyboard". Trailing numbers indicate sense identifiers. Relatedness and context clue scores are not shown for brevity.

relations over the list of related terms for the given sense into a weighted list of hypernyms.

Disambiguation of Related Terms and Hypernyms. While target words contain sense distinctions (PCZ ID), the related words and hypernyms do not carry sense information. At this step, each hypernym and related term is disambiguated with respect to the *induced* sense inventory (PCZ ID). For instance, the word "keyboard" in the list of related terms for the sense "mouse:1" is linked to its "device" sense represented ("keyboard:1") as "mouse:1" and "keyboard:1" share neighbors from the IT domain.

Retrieval of Context Clues. Salient contexts of senses are retrieved by aggregating salient dependency features of related terms. Context features that have a high weight for many related terms obtain a high weight for the sense.

3.2 HAR Datasets

We experiment with two different corpora for PCZ induction as in (Faralli et al., 2016), namely a 100 million sentence news corpus (*news*) from Gigaword (Parker et al., 2011) and LCC (Richter et al., 2006), and a 35 million sentence Wikipedia corpus (*wiki*).[1] Chinese Whispers sense clustering is performed with the default parameters, producing an average number of 2.3 (news) and 1.8 (wiki) senses per word in a vocabulary of 200 thousand words each, with the usual power-law distribution of sense cluster sizes. On average, each sense is related to about 47 senses and has assigned 5 hypernym labels. These disambiguated distributional networks were linked to WordNet 3.1 using the method of Faralli et al. (2016).

3.3 Using the Hybrid Aligned Resource in Word Sense Disambiguation

We experimented with four different ways of enriching the original WordNet-based sense repre-

---

[1]The used PCZ and HAR resources are available at: https://madata.bib.uni-mannheim.de/171

sentation with contextual information from the HAR on the basis of the mappings listed below:

WordNet. This baseline model relies solely on the WordNet lexical resource. It builds sense representations by collecting synonyms and sense definitions for the given WordNet synset and synsets directly connected to it. We removed stop words and weight words with term frequency.

WordNet + Related (news). This model augments the WordNet-based representation with related terms from the PCZ items (see Table 1). This setting is designed to quantify the added value of lexical knowledge in the related terms of PCZ.

WordNet + Related (news) + Context (news). This model includes all features of the previous models and complements them with context clues obtained by aggregating features of the words from the WordNet + Related (news) (see Table 1).

WordNet + Related (news) + Context (wiki). This model is built in the same way as the previous model, but using context clues derived from Wikipedia (see Section 3.2).

In the last two models, we used up to 5000 most relevant context clues per word sense. This value was set experimentally: performance of the WSD system gradually increased with the number of context clues reaching a plateau at the value of 5000. During aggregation, we excluded stop words and numbers from context clues. Besides, we transformed syntactic context clues presented in Table 1 to terms, stripping the dependency type. so they can be added to other lexical representations. For instance, the context clue "rat:conj_and" of the entry "mouse:0" was reduced to the feature "rat".

Table 2 demonstrates features extracted from WordNet as compared to feature representations enriched with related terms of the PCZ.

Each WordNet word sense is represented with one of the four methods described above. These sense representations are subsequently used to per-

Model	Sense Representation
WordNet	memory, device, floppy, disk, hard, disk, disk, computer, science, computing, diskette, fixed, disk, floppy, magnetic, disc, magnetic, disk, hard, disc, storage, device
WordNet + Related (news)	recorder, disk, floppy, console, diskette, handset, desktop, iPhone, iPod, HDTV, kit, RAM, Discs, Blu-ray, computer, GB, microchip, site, cartridge, printer, tv, VCR, Disc, player, LCD, software, component, camcorder, cellphone, card, monitor, display, burner, Web, stereo, internet, model, iTunes, turntable, chip, cable, camera, iphone, notebook, device, server, surface, wafer, page, drive, laptop, screen, pc, television, hardware, YouTube, dvr, DVD, product, folder, VCR, radio, phone, circuitry, partition, megabyte, peripheral, format, machine, tuner, website, merchandise, equipment, gb, discs, MP3, hard-drive, piece, video, storage device, memory device, microphone, hd, EP, content, soundtrack, webcam, system, blade, graphic, microprocessor, collection, document, programming, battery, keyboard, HD, handheld, CDs, reel, web, material, hard-disk, ep, chart, debut, configuration, recording, album, broadcast, download, fixed disk, planet, pda, microfilm, iPod, videotape, text, cylinder, cpu, canvas, label, sampler, workstation, electrode, magnetic disc, catheter, magnetic disk, Video, mobile, cd, song, modem, mouse, tube, set, ipad, signal, substrate, vinyl, music, clip, pad, audio, compilation, memory, message, reissue, ram, CD, subsystem, hdd, touchscreen, electronics, demo, shell, sensor, file, shelf, processor, cassette, extra, mainframe, motherboard, floppy disk, lp, tape, version, kilobyte, pacemaker, browser, Playstation, pager, module, cache, DVD, movie, Windows, cd-rom, e-book, valve, directory, harddrive, smartphone, audiotape, technology, hard disk, show, computing, computer science, Blu-Ray, blu-ray, HDD, HD-DVD, scanner, hard disc, gadget, booklet, copier, playback, TiVo, controller, filter, DVDs, gigabyte, paper, mp3, CPU, dvd-r, pipe, cd-r, playlist, slot, VHS, film, videocassette, interface, adapter, database, manual, book, channel, changer, storage

Table 2: Original and enriched representations of the third sense of the word "disk" in the WordNet sense inventory. Our sense representation is enriched with related words from the hybrid aligned resource.

form WSD in context. For each test instance consisting of a target word and its context, we select the sense whose corresponding sense representation has the highest cosine similarity with the target word's context.

4 Evaluation

We perform an extrinsic evaluation and show the impact of the hybrid aligned resource on word sense disambiguation performance. While there exist many datasets for WSD (Mihalcea et al., 2004; Pradhan et al., 2007; Manandhar et al., 2010, *inter alia*), we follow Navigli and Ponzetto (2012) and use the SemEval-2007 Task 16 on the "Evaluation of wide-coverage knowledge resources" (Cuadros and Rigau, 2007). This task is specifically designed for evaluating the impact of lexical resources on WSD performance. The SemEval-2007 Task 16 is, in turn, based on two "lexical sample" datasets, from the Senseval-3 (Mihalcea et al., 2004) and SemEval-2007 Task 17 (Pradhan et al., 2007) evaluation campaigns. The first dataset has coarse- and fine-grained annotations, while the second contains only fine-grained sense annotations. In all experiments, we use the official task's evaluator to compute standard metrics of recall, precision, and F-score.

5 Results

Impact of the corpus-based features. Figure 1 compares various sense representations in terms of F-score. The results show that, expanding WordNet-based sense representations with distributional information gives a clear advantage over the original representation on both Senseval-3 and SemEval-2007 datasets. Using related words specific to a given WordNet sense provides dramatic

improvements in the results. Further expansion of the sense representation with context clues (cf. Table 1) provide a modest further improvement on the SemEval-2007 dataset and yield no further improvement on the case of the Senseval-3 dataset.

Comparison to the state-of-the-art. We compare our approach to four state-of-the-art systems: KnowNet (Cuadros and Rigau, 2008), BabelNet, WN+XWN (Cuadros and Rigau, 2007), and NASARI. KnowNet builds sense representations based on snippets retrieved with a web search engine. We use the best configuration reported in the original paper (KnowNet-20), which extends each sense with 20 keywords. BabelNet in its core relies on a mapping of WordNet synsets and Wikipedia articles to obtain enriched sense representations. The WN+XWN system is the top-ranked unsupervised knowledge-based system of Senseval-3 and SemEval-2007 datasets from the original competition (Cuadros and Rigau, 2007). It alleviates sparsity by combining WordNet with the eXtended WordNet (Mihalcea and Moldovan, 2001). The latter resource relies on parsing of WordNet glosses.

For KnowNet, BabelNet, and WN+XWN we use the scores reported in the respective original publications. However, as NASARI was not evaluated on the datasets used in our study, we used the following procedure to obtain NASARI-based sense representations: Each WordNet-based sense representation was extended with all features from the lexical vectors of NASARI.[2]

Thus, we compare our method to three hybrid systems that induce sense representations on the

---

[2]We used the version of lexical vectors (July 2016) featuring 4.4 million of BabelNet synsets, yet covering only 72% of word senses of the two datasets used in our experiments.
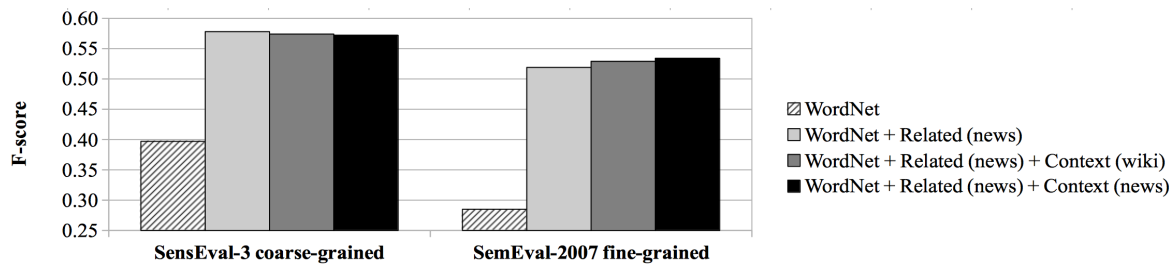
Figure 1: Performance of different word sense representation strategies.

Model	Senseval-3 fine-grained			SemEval-2007 fine-grained		
	Precision	Recall	F-score	Precision	Recall	F-score
Random	19.1	19.1	19.1	27.4	27.4	27.4
WordNet	29.7	29.7	29.7	44.3	21.0	28.5
WordNet + Related (news)	**47.5**	**47.5**	**47.5**	54.0	50.0	51.9
WordNet + Related (news) + Context (news)	47.2	47.2	47.2	54.8	51.2	52.9
WordNet + Related (news) + Context (wiki)	46.9	46.9	46.9	**55.2**	**51.6**	**53.4**
BabelNet	**44.3**	**44.3**	**44.3**	**56.9**	**53.1**	**54.9**
KnowNet	44.1	44.1	44.1	49.5	46.1	47.7
NASARI (lexical vectors)	32.3	32.2	32.2	49.3	45.8	47.5
WN+XWN	38.5	38.0	38.3	54.9	51.1	52.9

Table 3: Comparison of our approach to the state of the art unsupervised knowledge-based methods on the SemEval-2007 Task 16 (weighted setting). The best results overall are underlined.

basis of WordNet and texts (KnowNet, BabelNet, NASARI) and one purely knowledge-based system (WN+XWN). Note that we do not include the supervised TSSEM system in this comparison, as in contrast to all other considered methods, it relies on a large sense-labeled corpus.

Table 3 presents results of the evaluation. On the Senseval-3 dataset, our hybrid models show better performance than all unsupervised knowledge-based approaches considered in our experiment. On the SemEval-2007 dataset, the only resource which exceeds the performance of our hybrid model is BabelNet. The extra performance of BabelNet on the SemEval dataset can be explained by its multilingual approach: additional features are obtained using semantic relations across synsets in different languages. Besides, machine translation is used to further enrich coverage of the resource (Navigli and Ponzetto, 2012).

These results indicate on the high quality of the sense representations obtained using the hybrid aligned resource. Using related words of induced senses improves WSD performance by a large margin as compared to purely WordNet-based model on both datasets. Adding extra contextual features further improves slightly results on one dataset. Thus, we recommend enriching sense representations with related words and op-

tionally with context clues. Finally, note that, while our method shows competitive results compared to other state-of-the-art hybrid systems, it does not require access to web search engines (KnowNet), texts mapped to a sense inventory (BabelNet, NASARI), or machine translation systems (BabelNet).

6 Conclusion

The hybrid aligned resource (Faralli et al., 2016) successfully enriches sense representations of a manually-constructed lexical network with features derived from a distributional disambiguated lexical network. Our WSD experiments on two datasets show that this additional information extracted from corpora let us substantially outperform the model based solely on the lexical resource. Furthermore, a comparison of our sense representation method with existing hybrid approaches leveraging corpus-based features demonstrate its state-of-the-art performance.

Acknowledgments

References

Eneko Agirre and Philip Edmonds. 2007. *Word sense disambiguation: Algorithms and applications*, volume 33. Springer Science & Business Media.

Sergey Bartunov, Dmitry Kondrashkin, Anton Osokin, and Dmitry Vetrov. 2016. Breaking sticks and ambiguities with adaptive skip-gram. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics (AISTATS'2016)*, pages 130–138, Cadiz, Spain. JMLR: W&CP volume 51.

Chris Biemann and Martin Riedl. 2013. Text: Now in 2D! A Framework for Lexical Expansion with Contextual Similarity. *Journal of Language Modelling*, 1(1):55–95.

Chris Biemann. 2006. Chinese whispers - an efficient graph clustering algorithm and its application to natural language processing problems. In *Proceedings of TextGraphs: the First Workshop on Graph Based Methods for Natural Language Processing*, pages 73–80, New York City. Association for Computational Linguistics.

José Camacho-Collados, Mohammad Taher Pilehvar, and Roberto Navigli. 2015a. Nasari: a novel approach to a semantically-aware representation of items. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 567–577, Denver, Colorado. Association for Computational Linguistics.

José Camacho-Collados, Mohammad Taher Pilehvar, and Roberto Navigli. 2015b. A unified multilingual semantic representation of concepts. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 741–751, Beijing, China. Association for Computational Linguistics.

Tao Chen, Ruifeng Xu, Yulan He, and Xuan Wang. 2015. Improving distributed representation of word sense via wordnet gloss composition and context clustering. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 15–20, Beijing, China. Association for Computational Linguistics.

Montse Cuadros and German Rigau. 2007. Semeval-2007 task 16: Evaluation of wide coverage knowledge resources. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 81–86, Prague, Czech Republic. Association for Computational Linguistics.

Montse Cuadros and German Rigau. 2008. KnowNet: Building a large net of knowledge from the web. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pages 161–168, Manchester, UK, August. Coling 2008 Organizing Committee.

Beate Dorow and Dominic Widdows. 2003. Discovering Corpus-Specific Word Senses. In *Proceedings of the Tenth Conference on European Chapter of the Association for Computational Linguistics - Volume 2*, EACL '03, pages 79–82, Budapest, Hungary. Association for Computational Linguistics.

Martin Everett and Stephen P. Borgatti. 2005. Ego network betweenness. *Social Networks*, 27(1):31–38.

Stefano Faralli, Alexander Panchenko, Chris Biemann, and Simone P. Ponzetto. 2016. Linked disambiguated distributional semantic networks. In *International Semantic Web Conference (ISWC'2016)*, pages 56–64, Kobe, Japan. Springer.

Manaal Faruqui, Jesse Dodge, Sujay Kumar Jauhar, Chris Dyer, Eduard Hovy, and Noah A. Smith. 2015. Retrofitting word vectors to semantic lexicons. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1606–1615, Denver, Colorado. Association for Computational Linguistics.

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

Juri Ganitkevitch, Benjamin Van Durme, and Chris Callison-Burch. 2013. Ppdb: The paraphrase database. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 758–764, Atlanta, Georgia. Association for Computational Linguistics.

Josu Goikoetxea, Aitor Soroa, and Eneko Agirre. 2015. Random walks and neural network language models on knowledge bases. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1434–1439, Denver, Colorado. Association for Computational Linguistics.

Marti A. Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 6th International Conference on Computational Linguistics (COLING'1992)*, pages 539–545, Nantes, France.

David Hope and Bill Keller. 2013. MaxMax: A Graph-Based Soft Clustering Algorithm Applied to Word Sense Induction. In *Computational Linguistics and Intelligent Text Processing: 14th International Conference, CICLing 2013, Samos, Greece, March 24-30, 2013, Proceedings, Part I*, pages 368–381. Springer Berlin Heidelberg, Berlin, Heidelberg.

Ignacio Iacobacci, Mohammad Taher Pilehvar, and Roberto Navigli. 2015. Sensembed: Learning sense embeddings for word and relational similarity.

In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 95–105, Beijing, China. Association for Computational Linguistics.

Sujay Kumar Jauhar, Chris Dyer, and Eduard Hovy. 2015. Ontologically grounded multi-sense representation learning for semantic vector space models. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 683–693, Denver, Colorado. Association for Computational Linguistics.

Suresh Manandhar, Ioannis P. Klapaftis, Dmitriy Dligach, and Sameer S. Pradhan. 2010. SemEval-2010 task 14: Word sense induction & disambiguation. In *Proceedings of the 5th International Workshop on Semantic Evaluation (ACL'2010)*, pages 63–68, Uppsala, Sweden. Association for Computational Linguistics.

Rada Mihalcea and Dan Moldovan. 2001. extended wordnet: Progress report. In *In Proceedings of NAACL Workshop on WordNet and Other Lexical Resources*, pages 1–5, Pittsburgh, PA, USA. Association for Computational Linguistics.

Rada Mihalcea, Timothy Chklovski, and Adam Kilgarriff. 2004. The SENSEVAL-3 English lexical sample task. In *SENSEVAL-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text*, pages 25–28, Barcelona, Spain. Association for Computational Linguistics.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Proceedings of Advances in Neural Information Processing Systems 26 (NIPS'2013)*, pages 3111–3119, Harrahs and Harveys, CA, USA.

Andrea Moro, Alessandro Raganato, and Roberto Navigli. 2014. Entity linking meets word sense disambiguation: a unified approach. *Transactions of the Association for Computational Linguistics*, 2:231–244.

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.

Roberto Navigli. 2009. Word sense disambiguation: A survey. *ACM CSUR*, 41(2):1–69.

Arvind Neelakantan, Jeevan Shankar, Alexandre Passos, and Andrew McCallum. 2014. Efficient non-parametric estimation of multiple embeddings per word in vector space. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1059–1069,

Doha, Qatar. Association for Computational Linguistics.

Luis Nieto Piña and Richard Johansson. 2016. Embedding senses for efficient graph-based word sense disambiguation. In *Proceedings of TextGraphs-10: the Workshop on Graph-based Methods for Natural Language Processing*, pages 1–5, San Diego, CA, USA. Association for Computational Linguistics.

Robert Parker, David Graff, Junbo Kong, Ke Chen, and Kazuaki Maeda. 2011. *English Gigaword Fifth Edition*. Linguistic Data Consortium, Philadelphia.

Maria Pelevina, Nikolay Arefiev, Chris Biemann, and Alexander Panchenko. 2016. Making sense of word embeddings. In *Proceedings of the 1st Workshop on Representation Learning for NLP*, pages 174–183, Berlin, Germany. Association for Computational Linguistics.

Nghia The Pham, Angeliki Lazaridou, and Marco Baroni. 2015. A multitask objective to inject lexical contrast into distributional semantics. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 21–26, Beijing, China. Association for Computational Linguistics.

Sameer Pradhan, Edward Loper, Dmitriy Dligach, and Martha Palmer. 2007. Semeval-2007 task-17: English lexical sample, srl and all words. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 87–92, Prague, Czech Republic. Association for Computational Linguistics.

Matthias Richter, Uwe Quasthoff, Erla Hallsteinsdóttir, and Chris Biemann. 2006. Exploiting the Leipzig Corpora Collection. In *Proceedings of the Fifth Slovenian and First International Language Technologies Conference (IS-LTC)*, Ljubljana, Slovenia.

Sascha Rothe and Hinrich Schütze. 2015. Autoextend: Extending word embeddings to embeddings for synsets and lexemes. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1793–1803, Beijing, China, July. Association for Computational Linguistics.

Peter D. Turney and Patrick Pantel. 2010. From frequency to meaning: Vector space models of semantics. *JAIR*, 37:141–188.

Mo Yu and Mark Dredze. 2014. Improving lexical embeddings with semantic knowledge. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 545–550, Baltimore, Maryland. Association for Computational Linguistics.