SyntagNet: Challenging Supervised Word Sense Disambiguation with Lexical-Semantic Combinations

Marco Maru[1,2], **Federico Scozzafava**[1], **Federico Martelli**[1,3], and **Roberto Navigli**[1]

[1]Department of Computer Science, Sapienza University of Rome
[2]Department of Literature and Modern Cultures, Sapienza University of Rome
[3]Babelscape srl, Italy
`firstname.lastname@uniroma1.it, martelli@babelscape.com`

Abstract

Current research in knowledge-based Word Sense Disambiguation (WSD) indicates that performances depend heavily on the Lexical Knowledge Base (LKB) employed. This paper introduces SyntagNet, a novel resource consisting of manually disambiguated lexical-semantic combinations. By capturing sense distinctions evoked by syntagmatic relations, SyntagNet enables knowledge-based WSD systems to establish a new state of the art which challenges the hitherto unrivaled performances attained by supervised approaches. To the best of our knowledge, SyntagNet is the first large-scale manually-curated resource of this kind made available to the community (at `http://syntagnet.org`).

1 Introduction

Word Sense Disambiguation (WSD) is one of the most challenging Natural Language Processing (NLP) tasks. It deals with lexical ambiguity, and it is core to achieving the much-sought-for goal of Natural Language Understanding (Navigli, 2018). In a broad sense, two major approaches can be adopted for performing WSD: the supervised and the knowledge-based ones.

Supervised methods, which learn a classifier from training data, have improved from 65% (Snyder and Palmer, 2004) to over 71% accuracy, thus proving to perform best for WSD purposes (Yuan et al., 2016; Melacci et al., 2018; Uslu et al., 2018). However, supervision depends heavily on large quantities of reliable sense-annotated data which – especially for languages other than English – are poorly available.

As a major alternative to supervised WSD, knowledge-based approaches drop the requirement for large amounts of training data by drawing on rich Lexical Knowledge Bases (LKB) such as WordNet (Fellbaum, 1998), and allow scaling to multiple languages thanks to multilingual resources such as BabelNet (Navigli and Ponzetto, 2012). The structure of an LKB plays a key role in increasing the overall disambiguation performance. For the purposes of WSD, a critical consideration concerns the nature of the relations connecting concepts: while LKBs tend to focus on the paradigmatic dimension of language, such resources fall short in respect of syntagmatic relations[1], which are also crucial for sense disambiguation due to interconnecting co-occurring words (Navigli and Lapata, 2010).

In this paper we address this deficiency and present, for the first time, a manually-curated large-scale lexical-semantic combination[2] database which associates pairs of concepts with pairs of co-occurring words. Importantly, we prove the effectiveness of our resource by achieving the state of the art in multilingual knowledge-based WSD and by matching supervised WSD performances when integrated into an LKB made up of WordNet and the Princeton WordNet Gloss Corpus[3].

2 Related work

Several studies on knowledge-based algorithms have indicated that the LKB structure is of vital importance in determining the accuracy of sense disambiguation. In particular, it has been demonstrated that WSD performance improves

---

[1]A syntagmatic relation exists between two words which co-occur in spoken or written language more frequently than would be expected by chance and which have different grammatical roles in the sentences in which they occur (Saussure, 1916).

[2]Here, we will use the term "lexical combination" to refer to both lexical collocations and free word associations (without considering idiomatic expressions) and the term "lexical-semantic combination" to refer to sense-annotated lexical combinations.

[3]`http://wordnetcode.princeton.edu/glosstag.shtml`

dramatically when employing an LKB with a larger number of high-quality lexical-semantic relations, i.e., more connections between concepts (Boyd-Graber et al., 2006; Lemnitzer et al., 2008; Ponzetto and Navigli, 2010). During the last two decades, a certain amount of work has been carried out aimed at enriching LKBs with new lexical-semantic relations. To this end, knowledge has been (semi-)automatically extracted from large collections of data and integrated into lexical resources such as WordNet.

As far as semi-automatic approaches are concerned, Mihalcea and Moldovan (2001) conceived eXtended WordNet, a resource providing disambiguated glosses by means of a classification ensemble combined with human supervision. A set of manually disambiguated glosses, called the Princeton WordNet Gloss Corpus (PWNG), which inherently included syntagmatic content, was subsequently also made available in 2008.

The rationale behind the creation of such resources was substantiated in a knowledge-based WSD study conducted by Navigli and Lapata (2010), who hypothesized an improvement in performance by several points when enriching a semantic network with tens of lexical-semantic relations for each target word sense. To achieve this demanding goal, endeavors in the literature focused on the fully-automatic production of semantic combinations, such as those obtained by disambiguating topic signatures (Cuadros and Rigau, 2008; Cuadros et al., 2012, KnowNet and deep-KnowNet) or by disentangling the concepts in ConceptNet (Chen and Liu, 2011).

More recently, Espinosa-Anke et al. (2016) aimed at automatically enriching WordNet with collocational information by leveraging the relations between sense-level embedding spaces (Col-WordNet), while Simov et al. (2016) addressed the enhancement of LKBs by exploiting relations over semantically-annotated corpora as contextual information. To the same end, Simov et al. (2018) employed grammatical role embeddings to gather new syntagmatic relations.

The lack of syntagmatic information in semantic networks was also tackled by the extension of a lexical database by means of *phrasets*, i.e., sets of free combinations of words recurrently used to express a concept (Bentivogli and Pianta, 2004).

Unfortunately, due to their (semi-)automatic nature, the aforementioned resources could not inherently offer wide coverage and high precision at the same time. Compared to other resources geared towards knowledge-based WSD, the novel resource we contribute in this work features: (i) wide coverage with a broad spectrum of possible lexical combinations, and (ii) high precision thanks to being entirely manually curated.

3 SyntagNet: a wide-coverage lexical-semantic combination resource

In this Section, we present SyntagNet, a knowledge resource created starting from lexical combinations extracted from the English Wikipedia[4] and the British National Corpus (Leech, 1992, BNC), and manually disambiguated according to the WordNet 3.0 sense inventory.

3.1 Methodology

Lexical combination extraction First of all, we employed the Stanford CoreNLP pipeline (Manning et al., 2014) to extract the dependency trees[5] for all the sentences in both Wikipedia and the BNC. Then, in order to identify relevant combinations, we determined the strength of correlation between pairs of POS-tagged, lemmatized content words[6] w_1, w_2, co-occurring within a sliding window of 3 words. Each candidate pair (w_1, w_2) was weighted using Dice's coefficient multiplied by a logarithmic factor of the co-occurrence frequency:

$$score(w_1, w_2) = log_2(1 + n_{w_1 w_2}) \frac{2n_{w_1 w_2}}{n_{w_1} + n_{w_2}} \quad (1)$$

where n_{w_i} ($i \in \{1, 2\}$) is the frequency of w_i and $n_{w_1 w_2}$ is the frequency of the two words co-occurring within a window.

Three filters were then applied in order to slim down the list of pairs: (i) we filtered out English stopwords according to the Natural Language Toolkit (Loper and Bird, 2002, NLTK 3.4); (ii) we discarded combinations between verbs and verbs; (iii) we discarded combinations not linked by any of the five most frequent dependencies in our list, namely: `compound`, `dobj` (direct object), `iobj` (indirect object), `nsubj` (nominal subject) and `nmod` (nominal modifier).

Finally, we ranked the resulting lexical combination list according to the geometric mean

---

[4]November 2018 English Wikipedia dump.

[5]According to the Universal Dependencies v2 (https://universaldependencies.org/u/dep/all.html).

[6]Restricted to nouns and verbs in the WordNet dictionary.

word 1	word 2	score	sense 1	sense 2
run_v	program_n	18.07	run_v^{19} (carry out a process or program)	program_n^7 (a sequence of instructions)
run_v	race_n	11.55	run_v^{37} (compete in a race)	race_n^2 (a contest of speed)
run_v	farm_n	3.50	run_v^4 (direct or control)	farm_n^1 (workplace with farm buildings)

Table 1: Examples of high-ranking lexical (left) and semantic (right) combinations, where each lemma's subscript and superscript indicate its part of speech and sense number, respectively, in WordNet.

between i) the logarithmic Dice scores and ii) the frequency count of a pair in a given POS tag/dependency combination. We show some examples with $w_1 = \text{run}_v$, together with their final correlation score in Table 1 (left).

We then repeated the whole process described above, with the following changes: i) we set a sliding window of 6 words; ii) we removed the constraint on the dependency selection; iii) we filtered out all pairs already occurring within the first list; iv) we selected only items attested in multiple English monolingual and collocation dictionaries.

Manual disambiguation We asked eight annotators to manually disambiguate the top-ranking $20,000$ lexical combinations from the first list and $58,000$ lexical combinations from the second list, i.e., to associate each word in a pair (w_1, w_2) with its most appropriate senses in WordNet (in Table 1 (right) we show the senses chosen by the annotators for the corresponding lexical combinations).

The eight annotators shared a background in linguistics (Master's Degree with a minimum C1 English proficiency level) and were well acquainted with WordNet. In order to facilitate the annotation process, we provided each annotator with a unique batch of lexical combinations in a simple interface; for each pair, the annotators visualized all the synsets for each word of the combination (along with WordNet definitions and examples), and a context of up to 25 random sentences in which the combination was extracted. The annotators were asked to input the sense numbers associated with their chosen synsets for both the words in a given pair. Since the combinations can carry different meanings depending on the context, the annotators were allowed to assign multiple senses to the same word in a given combination (e.g., judge in the "public official" sense vs. the "evaluator" sense in the $(\text{judge}_n, \text{decide}_v)$ lexical combination).

As a further measure to ensure quality, the annotators were also asked to skip the annotation of lexical combinations (i) carrying mistakes due to the automatic parsing process, (ii) for which none of the available senses in WordNet would fit

the context, (iii) reflecting idiomatic expressions, (iv) which were multi-word Named Entities.

Overall, the annotators covered $78,000$ lexical combinations ($10,218$ unique nouns and $3,786$ unique verbs for $52,432$ noun-verb relations and $25,568$ noun-noun relations), and obtained $88,019$ semantic combinations ($61,249$ noun-verb and $26,770$ noun-noun semantic relations) linking $20,626$ WordNet 3.0 nodes, i.e., unique synsets ($14,204$ noun synsets and $6,422$ verb synsets), with a relation edge.

We periodically timed the annotators by considering the number of annotations produced on a daily basis, obtaining an average value of 42 disambiguated combinations per hour (1 minute and 26 seconds per word pair). Overall, the annotation process took a period of 9 months.

To determine the reliability of the annotations, we calculated the minimum inter-annotator agreement between pairs of annotators on a random sample of 500 combinations. For each of the 500 lexical combinations used to compute the inter-annotator agreement, the annotators were exceptionally asked to disambiguate the two target words in all of the 25 sentences provided, thus leading to a figure of 25,000 single instances disambiguated per annotator, resulting in a substantial agreement ($\kappa = 0.71$). Moreover, we found that most of the disagreement instances arose out of valid alternative tags, rather than factual errors, due to the fine granularity of the WordNet sense inventory.

4 Experimental setup

We now present the setup of our evaluation, carried out to assess the effectiveness of SyntagNet when employed for knowledge-based WSD.

Disambiguation algorithm We performed our experiments employing UKB[7] (Agirre et al., 2014), a state-of-the-art system for knowledge-based WSD, which applies the Personalized Page Rank (PPR) algorithm (Haveliwala, 2002) to an

---

		English						Multilingual						
								SemEval-13				SemEval-15		
resource	#relations	Sens2	Sens3	Sem07	Sem13	Sem15	All	IT	ES	DE	FR	IT	ES	All
WNG (WordNet+PWNG)	671,779	69.2	65.9	54.9	66.8	70.7	67.1	71.4	71.2	68.0	69.6	62.2	58.1	67.2
WNG+KnowNet20	520,682	67.2	65.8	53.8	67.3	71.5	66.6	71.6	73.1	68.3	70.4	61.4	59.9	67.9
WNG+deepKnowNet95d	522,880	66.9	64.9	53.6	66.9	71.6	66.2	71.4	71.9	67.7	70.5	62.4	58.7	67.5
WNG+BabelNet 4.0	9,447,341	67.5	64.1	53.0	67.6	66.9	65.6	73.8	71.6	**69.9**	67.1	62.4	57.8	67.6
WNG+eXtended WordNet	551,551	67.7	65.7	52.3	67.6	71.0	66.7	72.4	71.8	68.5	69.3	62.4	58.9	67.7
WNG+ColWordNet	8,424	69.2	65.9	54.1	66.7	70.7	67.1	71.4	71.0	68.0	69.3	61.9	57.8	67.0
WNG+SyntagNet	88,019	**71.2**	**71.6**	**59.6**	**72.4**	**75.6**	**71.5**	**74.2**	**73.4**	66.9	**72.7**	**65.0**	**61.2**	**69.3**

Table 2: F1 scores (%) for English all-words fine-grained WSD (left) and for multilingual all-words fine-grained WSD (right). Each row displays results scored by a specific resource combined with the WNG (WordNet+PWNG) baseline. Statistically-significant differences, according to a χ^2 test ($p < 0.01$), compared to the baseline (first row), are underlined. The second column reports the number of relations of the added resource.

input LKB. We used its *PPRw2w* single-sentence context disambiguation method, which initializes the PPR vector using the context of the target word in a given sentence, while excluding the contribution of the target word itself.

Evaluation benchmarks and measures We used five test sets standardized with WordNet 3.0 (Raganato et al., 2017a) including the English all-words tasks from Senseval-2 (Edmonds and Cotton, 2001), Senseval-3 (Snyder and Palmer, 2004), SemEval-2007 (Pradhan et al., 2007), SemEval-2013 (Navigli et al., 2013) and SemEval-2015 (Moro and Navigli, 2015). To run experiments on multilingual WSD, we used the last two of the foregoing datasets, which also include German, Spanish, French and Italian, employing, as sense inventory, the synset lexicalizations provided in BabelNet 4.0[8]. As customary, we computed precision, recall and F1, which in our case coincided, due to UKB always outputting a sense for each target word.

LKBs For the purposes of our evaluation we measured the performance obtained with UKB when combined with different LKBs. As our baseline we used WordNet + PWNG, which is the best configuration of UKB according to its authors. We also evaluated the following LKBs when integrated separately on top of our baseline: (i) the best configurations of KnowNet[9] and deepKnowNet[10], (ii) the subgraph of BabelNet 4.0 induced by WordNet 3.0, (iii) eXtended WordNet[11], (iv) ColWordNet[12], (v) SyntagNet (cf. Section 2).

All of the aforementioned LKBs (whose size in terms of lexical-semantic relations is shown in Table 2 (second column)) are available for download.

5 Experimental results

English WSD As shown in Table 2 (left), SyntagNet enabled UKB to achieve the best results in the English all-words disambiguation tasks, attaining 4.4 overall points above the WNG baseline, which is the only statistically-significant improvement across LKBs. Furthermore, results for the individual datasets exhibit statistically-significant improvements over the baseline on two out of five datasets. We attribute this result to the fully manual nature of SyntagNet, in contrast to the noisy character of the other LKBs. A further justification of our results comes from an analysis we performed on relation samples from the various LKBs: we collected 500 random relations for each LKB we experimented with, and manually tagged each of them as syntagmatic or paradigmatic, revealing that their syntagmatic contribution ranges from 39% (deepKnowNet) to 54% (eXtended WordNet). The fully syntagmatic nature of SyntagNet, instead, effectively blends in with the complementary information available in the baseline (63% of the relations in WNG are paradigmatic).

Table 3 compares UKB + SyntagNet against the best supervised English WSD systems (Yuan et al., 2016; Melacci et al., 2018; Uslu et al., 2018): none of the differences across datasets between the best performing supervised system and SyntagNet is statistically significant according to a χ^2 test ($p < 0.01$), meaning that SyntagNet enables knowledge-based WSD to rival current supervised approaches.

system	Sens2	Sens3	Sem07	Sem13	Sem15	All
LSTMLP●	**73.8**	71.8	**63.5**	69.5	72.6	**71.5**
IMSC2V$_{+PR}$∞	**73.8**	71.9	63.3	<u>68.2</u>	72.8	71.2
fastSense△	73.5	**73.5**	62.4	<u>66.2</u>	73.2	71.1
UKB+SyntagNet	71.2	71.6	59.6	**72.4**	75.6	71.5

Table 3: F1 scores (%) of UKB+SyntagNet against the best supervised systems for English all-words WSD. Reported systems: ● Yuan et al. (2016), ∞ Melacci et al. (2018), △ Uslu et al. (2018). Statistically-significant differences against our results are underlined according to a χ^2 test, $p < 0.01$.

	SemEval-13				SemEval-15		
system	IT	ES	DE	FR	IT	ES	All
BLSTM†	<u>62.0</u>	66.4	**69.2**	<u>55.5</u>	-	-	-
UMCC-DLSI⋆	<u>65.8</u>	71.0	62.1	<u>60.5</u>	-	-	-
T-O-M○	<u>68.2</u>	66.9	63.2	<u>60.5</u>	-	-	-
SUDOKU-RUN1◇	-	-	-	-	59.9	56.0	-
SUDOKU-RUN2◇	-	-	-	-	56.9	57.1	-
Best system‡	<u>68.2</u>	71.0	**69.2**	<u>60.5</u>	59.9	57.1	<u>64.7</u>
UKB+SyntagNet	**74.2**	**73.4**	66.9	**72.7**	**65.0**	61.2	**69.3**

Table 4: F1 scores (%) of UKB+SyntagNet against the best systems for multilingual all-words WSD. Reported systems: † Raganato et al. (2017b), ⋆ Gutiérrez Vázquez et al. (2010), ○ Pasini and Navigli (2017), ◇ Manion (2015), ‡ result obtained by aggregating the outputs of the best systems for each dataset. Statistically-significant differences against our results are underlined according to a χ^2 test, $p < 0.01$.

Multilingual WSD As regards our multilingual evaluation, SyntagNet enabled UKB to attain the best overall result (see Table 2 (right)), which is a statistically-significant improvement of 2.1 points over the baseline. With respect to the comparison against the best systems (Table 4), SyntagNet provides a statistically-relevant boost of 4.6 points in relation to the aggregate score of the compared systems (second to last row), attaining state-of-the-art results on five out of the six datasets taken into account.

5.1 Impact of LKB size

Finally, we graphed the increase in WSD performance obtained when progressively enriching the baseline UKB graph with random samples of $10,000$ SyntagNet relations at each step. As illustrated in Figure 1, the improvements in the English and multilingual settings, respectively, present a growing trend according to a linear regression analysis of the data. This demonstrates that our relations are high-quality and effective for WSD, while leaving room for further improvement as more relations are added in the future.
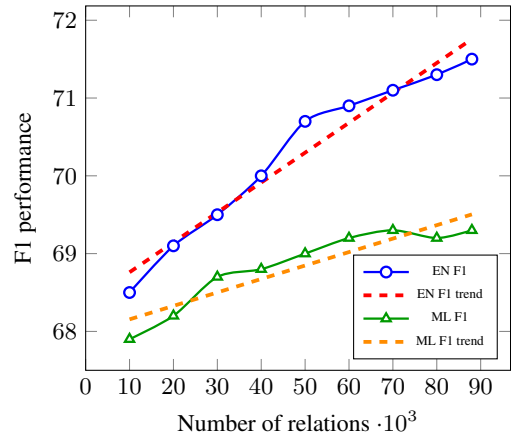

Figure 1: Impact of growing samples of SyntagNet over the baseline on overall F1 for English (EN F1 line) and multilingual (ML F1 line) evaluation datasets.

6 Conclusions

In this paper we put forward two main contributions: 1) we presented SyntagNet (http://syntagnet.org), a new wide-coverage, manually-curated resource of lexical-semantic combinations; 2) we showed that SyntagNet enables state-of-the-art knowledge-based WSD, rivaling the best supervised system on English and surpassing the overall performance of the best multilingual systems by 4.6 points. As future work, we plan to: i) enrich SyntagNet with more combinations, so as to surpass supervised English WSD; ii) include information from adjectives; iii) establish a common evaluation framework to compare the contribution of lexical-semantic combination resources; iv) employ and assess SyntagNet in other NLP tasks, such as word and sense similarity (Navigli and Martelli, 2019) or Semantic Role Labeling (where a newly released WordNet-linked resource, VerbAtlas (Di Fabio et al., 2019), would greatly benefit from collocational information).

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.

Luisa Bentivogli and Emanuele Pianta. 2004. Extending WordNet with Syntagmatic Information. In *Proceedings of second global WordNet conference*, pages 47–53, Brno, Czech Republic.

Jordan Boyd-Graber, Christiane Fellbaum, Daniel Osherson, and Robert Schapire. 2006. Adding Dense, Weighted Connections to WordNet. In *Proceedings of the third international WordNet conference*, pages 29–36, South Jeju Island, Korea.

Junpeng Chen and Juan Liu. 2011. Combining ConceptNet and WordNet for Word Sense Disambiguation. In *Proceedings of 5th International Joint Conference on Natural Language Processing*, pages 686–694, Chiang Mai, Thailand.

Montse Cuadros, Lluís Padró, and German Rigau. 2012. Highlighting relevant concepts from topic signatures. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*, pages 3841–3848, Istanbul, Turkey.

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-Volume 1*, pages 161–168, Manchester, United Kingdom.

Andrea Di Fabio, Simone Conia, and Roberto Navigli. 2019. VerbAtlas: a Novel Large-Scale Verbal Semantic Resource and Its Application to Semantic Role Labeling. In *Proc. of EMNLP-IJCNLP*.

Philip Edmonds and Scott Cotton. 2001. SENSEVAL-2: Overview. In *Proceedings of SENSEVAL-2 Second International Workshop on Evaluating Word Sense Disambiguation Systems*, pages 1–5, Toulouse, France.

Luis Espinosa-Anke, Jose Camacho-Collados, Sara Rodríguez Fernández, Horacio Saggion, and Leo Wanner. 2016. Extending WordNet with Fine-Grained Collocational Information via Supervised Distributional Learning. In *Proceedings of COLING 2016: Technical Papers. The 26th International Conference on Computational Linguistics*, pages 900–910, Osaka, Japan.

Christiane Fellbaum, editor. 1998. *WordNet: an electronic lexical database*. MIT Press, Cambridge, MA, USA.

Yoan Gutiérrez Vázquez, Antonio Fernandez Orquín, Andrés Montoyo Guijarro, and Sonia Vázquez Pérez. 2010. UMCC-DLSI: Integrative Resource for Disambiguation Task. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 427–432, Uppsala, Sweden.

Taher H. Haveliwala. 2002. Topic-Sensitive PageRank. In *Proceedings of the 11th international conference on World Wide Web*, pages 517–526, Honolulu, HI, USA.

Geoffrey Neil Leech. 1992. 100 million words of English: the British National Corpus (BNC). *Language Research*.

Lothar Lemnitzer, Holger Wunsch, and Piklu Gupta. 2008. Enriching GermaNet with verb-noun relations - a case study of lexical acquisition. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation, LREC 2008, May 28-30, 2018*, pages 156–160, Marrakech, Morocco.

Edward Loper and Steven Bird. 2002. NLTK: The Natural Language Toolkit. In *Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics - Volume 1*, ETMTNLP '02, pages 63–70, Stroudsburg, PA, USA.

Steve L. Manion. 2015. SUDOKU: Treating Word Sense Disambiguation & Entity Linking as a Deterministic Problem-via an Unsupervised & Iterative Approach. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval-2015)*, pages 365–369, Denver, CO, USA.

Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60, Baltimore, MD, USA.

Stefano Melacci, Achille Globo, and Leonardo Rigutini. 2018. Enhancing Modern Supervised Word Sense Disambiguation Models by Semantic Lexical Resources. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation, LREC 2018, May 7-12, 2018*, Miyazaki, Japan.

Rada Mihalcea and Dan Moldovan. 2001. eXtended WordNet: Progress Report. In *Proc. of the NAACL Workshop on WordNet and Other Lexical Resources*, pages 95–100, Pittsburgh, PA, USA.

Andrea Moro and Roberto Navigli. 2015. SemEval-2015 task 13: Multilingual all-words sense disambiguation and entity linking. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 288–297, Denver, CO, USA.

Roberto Navigli. 2018. Natural Language Understanding: Instructions for (present and future) use. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI 2018)*, pages 5697–5702, Stockholm, Sweden.

Roberto Navigli, David Jurgens, and Daniele Vannella. 2013. SemEval-2013 task 12: Multilingual word sense disambiguation. In *Second Joint Conference*

on *Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 222–231, Atlanta, GA, USA.

Roberto Navigli and Mirella Lapata. 2010. An Experimental Study of Graph Connectivity for Unsupervised Word Sense Disambiguation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(4):678–692.

Roberto Navigli and Federico Martelli. 2019. An overview of word and sense similarity. *Natural Language Engineering*, 25(6):1–22.

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

Tommaso Pasini and Roberto Navigli. 2017. Train-O-Matic: Large-Scale Supervised Word Sense Disambiguation in Multiple Languages without Manual Training Data. In *Proc. of EMNLP: System Demonstrations*, pages 78–88, Copenhagen, Denmark.

Simone P. Ponzetto and Roberto Navigli. 2010. Knowledge-Rich Word Sense Disambiguation Rivaling Supervised Systems. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1522–1531, Uppsala, Sweden.

Sameer S. Pradhan, Edward Loper, Dmitriy Dligach, and Martha Palmer. 2007. SemEval-2007 Task 17: English Lexical Sample, SRL and All Words. In *Proc. of SemEval-2007*, pages 87–92, Stroudsburg, PA, USA.

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. 2017b. Neural Sequence Learning Models for Word Sense Disambiguation. In *Proc. of EMNLP*, pages 1167–1178, Copenhagen, Denmark.

Ferdinand de Saussure. 1916. *Cours de linguistique générale*. Payot, Paris.

Kiril Simov, Petya Osenova, and Alexander Popov. 2016. Using Context Information for Knowledge-Based Word Sense Disambiguation. In *International Conference on Artificial Intelligence: Methodology, Systems, and Applications*, pages 130–139, Varna, Bulgaria.

Kiril Simov, Alexander Popov, Iliana Simova, and Petya Osenova. 2018. Grammatical Role Embeddings for Enhancements of Relation Density in the Princeton WordNet. In *Proceedings of the 9th Global WordNet Conference*, pages 287–295, Singapore.

Benjamin Snyder and Martha Palmer. 2004. The English all-words task. In *In Proceedings of the 3rd International Workshop on the Evaluation of Systems for the Semantic Analysis of Text (SENSEVAL-3)*, pages 41–43, Barcelona, Spain.

Tolga Uslu, Alexander Mehler, Daniel Baumartz, Alexander Henlein, and Wahed Hemati. 2018. Fast-Sense: An Efficient Word Sense Disambiguation Classifier. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation, LREC 2018, May 7-12, 2018*, Miyazaki, Japan.

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