Symmetric Patterns *and* Coordinations: Fast *and* Enhanced Representations of Verbs *and* Adjectives

Roy Schwartz,[1] **Roi Reichart,**[2] **Ari Rappoport**[1]
[1]Institute of Computer Science, The Hebrew University
[2]Faculty of Industrial Engineering and Management, Technion, IIT
{roys02|arir}@cs.huji.ac.il roiri@ie.technion.ac.il

Abstract

State-of-the-art word embeddings, which are often trained on bag-of-words (*BOW*) contexts, provide a high quality representation of aspects of the semantics of nouns. However, their quality decreases substantially for the task of verb similarity prediction. In this paper we show that using symmetric pattern contexts (*SPs*, e.g., "X and Y") improves word2vec verb similarity performance by up to 15% and is also instrumental in adjective similarity prediction. The unsupervised *SP* contexts are even superior to a variety of dependency contexts extracted using a supervised dependency parser. Moreover, we observe that *SPs* and dependency coordination contexts (*Coor*) capture a similar type of information, and demonstrate that *Coor* contexts are superior to other dependency contexts including the set of all dependency contexts, although they are still inferior to *SPs*. Finally, there are substantially fewer *SP* contexts compared to alternative representations, leading to a massive reduction in training time. On an 8G words corpus and a 32 core machine, the *SP* model trains in 11 minutes, compared to 5 and 11 hours with *BOW* and all dependency contexts, respectively.

1 Introduction

In recent years, vector space models (VSMs) have become prominent in NLP. VSMs are often evaluated by measuring their ability to predict human judgments of lexical semantic relations between pairs of words, mostly association or similarity. While many datasets for these tasks are limited to pairs of nouns, the recent SimLex999 word similarity dataset (Hill et al., 2014) also consists of similarity scores for *verb* and *adjective* pairs. State-of-the-art VSMs such as word2vec skip-gram (*w2v-SG*, (Mikolov et al., 2013a)) and GloVe (Pennington et al., 2014) excel at noun-related tasks. However, their performance substantially decreases on *verb* similarity prediction in SimLex999, and their adjective representations have rarely been evaluated (Section 2).

In this paper we show that a key factor in the reduced performance of the *w2v-SG* model on verb representation is its reliance on bag-of-words (*BOW*) contexts: contexts of the represented words that consist of words in their physical proximity. We investigate a number of alternative contexts for this model, including various dependency contexts, and show that simple, automatically acquired symmetric patterns (*SPs*, e.g., "X or Y", (Hearst, 1992; Davidov and Rappoport, 2006)) are the most useful contexts for the representation of verbs and also adjectives. Moreover, the *SP*-based model is much more compact than the alternatives, making its training an order of magnitude faster.

In particular, we train several versions of the *w2v-SG* model, each with a different context type, and evaluate the resulting word embeddings on the task of predicting the similarity scores of the verb and adjective portions of SimLex999. Our results show that *SP* contexts (SG-*SP*) obtain the best results on both tasks: Spearman's ρ scores of 0.459 on verbs and 0.651 on adjectives. These results are 15.2% and 4.7% better than *BOW* contexts and 7.3% and 6.5% better than all dependency contexts (*DepAll*). Moreover, the number of *SP* contexts is substantially smaller than the alternatives, making it extremely fast to train: 11 minutes only on an 8G word corpus using a 32 CPU core machine, compared to 5 and 11 hours for *BOW* and *DepAll*, respectively.

Recently, Schwartz et al. (2015) presented a count-based VSM that utilizes *SP* contexts (SRR15). This model excels on verb similarity, outperforming VSMs that use other contexts (e.g., *BOW* and *DepAll*) by more than 20%. In this paper we show that apart from its *SP* contexts, the success of SRR15 is attributed in large to its explicit representation of antonyms (*live/die*); turning this feature off reduces its performance to be on par with SG-*SP*. As opposed to Schwartz et al. (2015), we keep our VSM fixed across experiments (*w2v-SG*), changing only the context type. This allows us to attribute our improved results to one factor: *SP* contexts.

We further observe that *SP* contexts are tightly connected to syntactic *coordination* contexts (*Coor*, Section 3). Following this observation, we compare the *w2v-SG* model with three dependency-based context types: (a) *Coor* contexts; (b) all dependency links (*DepAll*); and (c) all dependency links excluding *Coor* links ($Coor^C$).[1] Our results show that training with *Coor* contexts is superior to training with the other context types, leading to improved similarity prediction of 2.7-4.1% and 4.3-6.9% on verbs and adjectives respectively.

These results demonstrate the prominence of *Coor* contexts in verb and adjective representation: these contexts are even better than their combination with the rest of the dependency-based contexts (the *DepAll* contexts). Nonetheless, although *Coor* contexts are extracted using a supervised dependency parser, they are still inferior to *SP* contexts, extracted automatically from plain text (Section 3), by 4.6% and 2.2% for verb and adjective pairs.

2 Background

Word Embeddings for Verbs and Adjectives. A number of evaluation sets consisting of word pairs scored by humans for semantic relations (mostly association and similarity) are in use for VSM evaluation. These include: RG-65 (Rubenstein and Goodenough, 1965), MC-30 (Miller and Charles, 1991), WordSim353 (Finkelstein et al., 2001), MEN (Bruni et al., 2014) and SimLex999 (Hill et al., 2014).[2]

Nouns are dominant in almost all of these datasets. For example, RG-65, MC-30 and WordSim353 consist of noun pairs almost exclusively. A few datasets contain pairs of verbs (Yang and Powers, 2006; Baker et al., 2014). The MEN dataset, although dominated by nouns, also contains verbs and adjectives. Nonetheless, the human judgment scores in these datasets reflect *relatedness* between words. In contrast, the recent SimLex999 dataset (Hill et al., 2014) contains word *similarity* scores for nouns (666 pairs), verbs (222 pairs) and adjectives (111 pairs). We use this dataset to study the effect of context type on VSM performance in a verb and adjective similarity prediction task.

Context Type in Word Embeddings. Most VSMs (e.g., (Collobert et al., 2011; Mikolov et al., 2013b; Pennington et al., 2014)) define the context of a target word to be the words in its physical proximity (bag-of-words contexts). Dependency contexts, consisting of the words connected to the target word by dependency links (Grefenstette, 1994; Padó and Lapata, 2007; Levy and Goldberg, 2014), are another well researched alternative. These works did not recognize the importance of syntactic coordination contexts (*Coor*).

Patterns have also been suggested as VSM contexts, but mostly for representing *pairs* of words (Turney, 2006; Turney, 2008). While this approach has been successful for extracting various types of word relations, using patterns to represent *single* words is useful for downstream applications. Recently, Schwartz et al. (2015) explored the value of *symmetric* pattern contexts for word representation, an idea this paper develops further.

A recently published approach (Melamud et al., 2016) also explored the effect of the type of context on the performance of word embedding models. Nonetheless, while they also explored bag-of-words and dependency contexts, they did not experiment with *SPs* or coordination contexts, which we find to be most useful for predicting word similarity.

Limitations of Word Embeddings. Recently, a few papers examined the limitations of word embedding models in representing different types of se-

[1] $Coor \cup Coor^C = DepAll$, $Coor \cap Coor^C = \emptyset$

[2] For a comprehensive list see: `wordvectors.org/`

mantic information. Levy et al. (2015) showed that word embeddings do not capture semantic relations such as hyponymy and entailment. Rubinstein et al. (2015) showed that while state-of-the-art embeddings are successful at capturing taxonomic information (e.g., *cow* is an animal), they are much less successful in capturing attributive properties (*bananas* are yellow). In (Schwartz et al., 2015), we showed that word embeddings are unable to distinguish between pairs of words with opposite meanings (antonyms, e.g., good/bad). In this paper we study the difficulties of bag-of-words based word embeddings in representing verb similarity.

3 Symmetric Patterns (*SPs*)

Lexico-syntactic patterns are templates of text that contain both words and wildcards (Hearst, 1992), e.g., "X *and* Y" and "X *for a* Y". Pattern *instances* are sequences of words that match a given pattern, such that concrete words replace each of the wildcards. For example, "**John** *and* **Mary**" is an instance of the pattern "X *and* Y". Patterns have been shown useful for a range of tasks, including word relation extraction (Lin et al., 2003; Davidov et al., 2007), knowledge extraction (Etzioni et al., 2005), sentiment analysis (Davidov et al., 2010) and authorship attribution (Schwartz et al., 2013).

Symmetric patterns (*SPs*) are lexico-syntactic patterns that comply to two constraints: (a) Each pattern has exactly two wildcards (e.g., **X** *or* **Y**); and (b) When two words (*X*,*Y*) co-occur in an *SP*, they are also likely to co-occur in this pattern in opposite positions, given a large enough corpus (e.g., "**X** *or* **Y**" and "**Y** *or* **X**"). For example, the pattern "**X** *and* **Y**" is symmetric as for a large number of word pairs (e.g., (*eat,drink*)) both members are likely to occur in both of its wildcard positions (e.g., "eat *and* drink", "drink *and* eat").

SPs have shown useful for tasks such as word clustering (Widdows and Dorow, 2002; Davidov and Rappoport, 2006), semantic class learning (Kozareva et al., 2008) and word classification (Schwartz et al., 2014). In this paper we demonstrate the value of *SP*-based contexts in vector representations of verbs and adjectives. The rationale behind this context type is that two words that co-occur in an *SP* tend to take the same semantic role in the sentence, and are thus likely to be similar in meaning (e.g., "(John and Mary) sang").

***SP* Extraction.** Many works that applied *SPs* in NLP tasks employed a hand-crafted list of patterns (Widdows and Dorow, 2002; Dorow et al., 2005; Feng et al., 2013). Following Schwartz et al. (2015) we employ the DR06 algorithm (Davidov and Rappoport, 2006), an unsupervised algorithm that extracts *SPs* from plain text. We apply this algorithm to our corpus (Section 4) and extract 11 *SPs*: "X *and* Y", "X *or* Y", "X *and the* Y", "X *or the* Y", "X *or a* Y", "X *nor* Y", "X *and one* Y", "*either* X *or* Y", "X *rather than* Y", "X *as well as* Y", "*from* X *to* Y". A description of the DR06 algorithm is beyond the scope of this paper; the interested reader is referred to (Davidov and Rappoport, 2006).

***SP* Contexts.** We generate *SP* contexts by taking the co-occurrence counts of pairs of words in *SPs*. For example, in the *SP* token "*boys and girls*", the term *girls* is taken as an *SP* context of the word *boys*, and *boys* is taken as an *SP* context of *girls*.

We do not make a distinction between the different *SPs*. E.g., "*boys* **and** *girls*" and "*boys* **or** *girls*" are treated the same. However, we distinguish between left and right contexts. For example, we generate different contexts for the word *girls*, one for left-hand contexts ("**girls** *and boys*") and another for right-hand contexts ("*boys and* **girls**").

***SPs* and Coordinations.** *SPs* and syntactic coordinations (*Coors*) are intimately related. For example, of the 11 *SPs* extracted in this paper by the DR06 algorithm (listed above), the first eight represent coordination structures. Moreover, these *SPs* account for more than 98% of the *SP* instances in our corpus. Indeed, due to the significant overlap between *SPs* and *Coors*, the former have been proposed as a simple model of the latter (Nakov and Hearst, 2005).[3]

Despite their tight connection, *SPs* sometimes fail to properly identify the components of *Coors*. For example, while *SPs* are instrumental in capturing shallow *Coors*, they fail in capturing coordination between phrases. Consider the sentence *John*

[3]Note though that the exact syntactic annotation of coordination is debatable both in the linguistic community (Tesnière, 1959; Hudson, 1980; Mel'čuk, 1988) and also in the NLP community (Nilsson et al., 2006; Schwartz et al., 2011; Schwartz et al., 2012).

walked and Mary ran: the *SP* "X *and* Y" captures the phrase *walked and Mary*, while the *Coor* links the heads of the connected phrases ("*walked*" and "*ran*"). *SPs*, on the other hand, can go beyond *Coors* and capture other types of symmetric structures like "*from* X *to* Y" and "X *rather than* Y".

Our experiments reveal that both *SPs* and *Coors* are highly useful contexts for verb and adjective representation, at least with respect to word similarity. Interestingly, *Coor* contexts, extracted using a supervised dependency parser, are less effective than *SP* contexts, which are extracted from plain text.

4 Experiments

Model. We keep the VSM fixed throughout our experiments, changing only the context type. This methodology allows us to evaluate the impact of different contexts on the VSM performance, as context choice is the only modeling decision that changes across experimental conditions.

Our VSM is the word2vec skip-gram model (*w2v-SG*, Mikolov et al. (2013a)), which obtains state-of-the-art results on a variety of NLP tasks (Baroni et al., 2014). We employ the word2vec toolkit.[4] For all context types other than *BOW* we use the word2vec package of (Levy and Goldberg, 2014),[5] which augments the standard word2vec toolkit with code that allows arbitrary context definition.

Experimental Setup. We experiment with the verb pair (222 pairs) and adjective pair (111 pairs) portions of SimLex999 (Hill et al., 2014). We report the Spearman ρ correlation between the ranks derived from the scores of the evaluated models and the human scores provided in SimLex999.[6]

We train the *w2v-SG* model with five different context types: (a) *BOW* contexts (SG-*BOW*); (b) all dependency links (SG-*DepAll*) (c) dependency-based coordination contexts (i.e., those labeled with *conj*, SG-*Coor*); (d) all dependency links except for coordinations (SG-*Coor*^{C}); and (e) *SP* contexts. Our training corpus is the 8G words corpus generated by the word2vec script.[7]

Model	Verb	Adj.	Noun	Time	#Cont.
SG-*BOW*	0.307	0.604	**0.501**	320	13G
SG-*DepAll*	0.386	0.586	0.499	551	14.5G
SG-*Coor*	0.413	0.629	0.428	23	550M
SG-*Coor*^{C}	0.372	0.56	0.494	677	14G
SG-*SP*	**0.459**	**0.651**	0.415	11	270M
SRR15	0.578	0.663	0.497	—	270M
SRR15^{-}	0.441	0.68	0.421	—	270M

Table 1: Spearman's ρ scores on the different portions of SimLex999. The top part presents results for the word2vec skip-gram model (*w2v-SG*) with various context types (see text). The bottom lines present the results of the count *SP*-based model of Schwartz et al. (2015), with (SRR15) and without (SRR15^{-}) its antonym detection method. The two rightmost columns present the run time of the *w2v-SG* models in minutes (Time) and the number of context instances used by the model (#Cont.).[10] For each SimLex999 portion, the score of the best *w2v-SG* model across context types is highlighted in bold font.

Models (b)-(d) require the dependency parse trees of the corpus as input. To generate these trees, we employ the Stanford POS Tagger (Toutanova et al., 2003)[8] and the stack version of the MALT parser (Nivre et al., 2009).[9] The *SP* contexts are generated using the *SPs* extracted by the DR06 algorithm from our training corpus (see Section 3).

For *BOW* contexts, we experiment with three window sizes (2, 5 and 10) and report the best results (window size of 2 across conditions). For dependency based contexts we follow the standard convention in the literature: we consider the immediate heads and modifiers of the represented word. All models are trained with 500 dimensions, the default value of the word2vec script. Other hyper-parameters were also set to the default values of the code packages.

Results. Table 1 presents our results. The SG-*SP* model provides the most useful verb and adjective representations among the *w2v-SG* models. Compared to *BOW* (SG-*BOW*), the most commonly used

[4]https://code.google.com/p/word2vec/

[5]https://bitbucket.org/yoavgo/word2vecf

[6]Model scores are computed in the standard way: applying the cosine similarity metric to the vectors learned for the words participating in the pair.

[7]code.google.com/p/word2vec/source/browse/trunk/demo-train-big-model-v1.sh

[8]nlp.stanford.edu/software/tagger.shtml

[9]http://www.maltparser.org/index.html

context type, SG-*SP* results are 15.2% and 4.7% higher on verbs and adjectives respectively. Compared to dependency links (SG-*DepAll*), the improvements are 7.3% and 6.5%. For completeness, we compare the models on the noun pairs portion, observing that SG-*BOW* and SG-*DepAll* are ∼8.5% better than SG-*SP*. This indicates that different word classes require different representations.

The results for SG-*Coor*, which is trained with syntactic coordination (*Coor*) contexts, show that these contexts are superior to all the other dependency links (SG-$Coor^C$) by 4.1% and 6.9% on verbs and adjectives. Importantly, comparing the SG-*Coor* model to the SG-*DepAll* model, which augments the *Coor* contexts with the other syntactic dependency contexts, reveals that SG-*DepAll* is actually inferior by 2.7% and 4.3% in Spearman ρ on verbs and adjectives respectively. Interestingly, *Coor* contexts, which are extracted using a supervised parser, are still inferior by 4.6% and 2.2% to *SPs*, which capture similar contexts but are extracted from plain text.

Table 1 also shows the training times of the various *w2v-SG* models on a 32G memory, 32 CPU core machine. SG-*SP* and SG-*Coor*, which take 11 minutes and 23 minutes respectively to train, are substantially faster than the other *w2v-SG* models. For example, they are more than an order of magnitude faster than SG-*BOW* (320 minutes) and SG-$Coor^C$ (677 minutes). This is not surprising, as there are far fewer *SP* contexts (270M) and *Coor* contexts (550M) than *BOW* contexts (13G) and $Coor^C$ contexts (14G) (#Cont. column).

Finally, the performance of the SG-*SP* model is still substantially inferior to the SRR15 *SP*-based model (Schwartz et al., 2015). As both models use the same *SP* contexts, this result indicates that other modeling decisions in SRR15 lead to its superior performance. We show that this difference is mostly attributed to one feature of SRR15: its method for detecting antonym pairs (*good/bad*). Indeed, the SRR15 model without its antonym detection method ($SRR15^-$) obtains a Spearman ρ of 0.441, compared to 0.459 of SG-*SP* on verb pairs. For adjectives, however, $SRR15^-$ is 1.7% better than SRR15, increasing the difference from SG-*SP* to 2.9%.[11]

[10] We compare the *w2v-SG* models training time only. SRR15 and $SRR15^-$ are count-based models and have no training step.

5 Conclusions

We demonstrated the effectiveness of symmetric pattern contexts in word embedding induction. Experiments with the word2vec model showed that these contexts are superior to various alternatives for verb and adjective representation. We further pointed at the connection between symmetric patterns and syntactic coordinations. We showed that coordinations are superior to other syntactic contexts, but are still inferior to symmetric patterns, although the extraction of symmetric patterns requires less supervision.

Future work includes developing a model that successfully combines the various context types explored in this paper. We are also interested in the representation of other word classes such as adverbs for which no evaluation set currently exists. Finally, the code for generating the SG-*SP* embeddings, as well as the vectors experimented with in this paper, are released and can be downloaded from `http://www.cs.huji.ac.il/~roys02/papers/sp_sg/sp_sg.html`

Acknowledgments

This research was funded (in part) by the Intel Collaborative Research Institute for Computational Intelligence (ICRI-CI), the Israel Ministry of Science and Technology Center of Knowledge in Machine Learning and Artificial Intelligence (Grant number 3-9243). The second author was partially funded by the Microsoft/Technion research center for electronic commerce and the Google faculty research award.

References

Simon Baker, Roi Reichart, and Anna Korhonen. 2014. An unsupervised model for instance level subcategorization acquisition. In *Proc. of EMNLP*.

Marco Baroni, Georgiana Dinu, and Germán Kruszewski. 2014. Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors. In *Proc. of ACL*.

[11] We report results for our reimplementation of SRR15 and $SRR15^-$.

Elia Bruni, Nam-Khanh Tran, and Marco Baroni. 2014. Multimodal distributional semantics. *JAIR*.

Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. Natural language processing (almost) from scratch. *JMLR*, 12:2493–2537.

Dmitry Davidov and Ari Rappoport. 2006. Efficient unsupervised discovery of word categories using symmetric patterns and high frequency words. In *Proc. of ACL-COLING*.

Dmitry Davidov, Ari Rappoport, and Moshe Koppel. 2007. Fully unsupervised discovery of concept-specific relationships by web mining. In *Proc. of ACL*.

Dmitry Davidov, Oren Tsur, and Ari Rappoport. 2010. Enhanced sentiment learning using twitter hashtags and smileys. In *Proc. of COLING*.

Beate Dorow, Dominic Widdows, Katarina Ling, Jean-Pierre Eckmann, Danilo Sergi, and Elisha Moses. 2005. Using Curvature and Markov Clustering in Graphs for Lexical Acquisition and Word Sense Discrimination.

Oren Etzioni, Michael Cafarella, Doug Downey, Ana-Maria Popescu, Tal Shaked, Stephen Soderland, Daniel S Weld, and Alexander Yates. 2005. Unsupervised named-entity extraction from the web: An experimental study. *Artificial intelligence*, 165(1):91–134.

Song Feng, Jun Seok Kang, Polina Kuznetsova, and Yejin Choi. 2013. Connotation lexicon: A dash of sentiment beneath the surface meaning. In *Proc. of ACL*.

Lev Finkelstein, Evgeniy Gabrilovich, Yossi Matias, Ehud Rivlin, Zach Solan, Gadi Wolfman, and Eytan Ruppin. 2001. Placing search in context: The concept revisited. In *Proc. of WWW*.

Gregory Grefenstette. 1994. *Explorations in automatic thesaurus discovery*. Boston: Kluwer.

Marti A. Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *Proc. of COLING – Volume 2*.

Felix Hill, Roi Reichart, and Anna Korhonen. 2014. Simlex-999: Evaluating semantic models with (genuine) similarity estimation. *arXiv:1408.3456 [cs.CL]*.

Richard A. Hudson. 1980. *Arguments for a Non-transformational Grammar*. Chicago: University of Chicago Press.

Zornitsa Kozareva, Ellen Riloff, and Eduard Hovy. 2008. Semantic class learning from the web with hyponym pattern linkage graphs. In *Proc. of ACL-HLT*.

Omer Levy and Yoav Goldberg. 2014. Dependency-based word embeddings. In *Proc. of ACL*.

Omer Levy, Steffen Remus, Chris Biemann, and Ido Dagan. 2015. Do supervised distributional methods really learn lexical inference relations? In *Proc. of NAACL*.

Dekang Lin, Shaojun Zhao, Lijuan Qin, and Ming Zhou. 2003. Identifying synonyms among distributionally similar words. In *Proc. of IJCAI*.

Oren Melamud, David McClosky, Siddharth Patwardhan, and Mohit Bansal. 2016. The role of context types and dimensionality in learning word embeddings. In *Proc. of NAACL*.

Igor Mel'čuk. 1988. *Dependency Syntax: Theory and Practice*. State University of New York Press.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013a. Distributed representations of words and phrases and their compositionality. In *Proc. of NIPS*.

Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. 2013b. Linguistic regularities in continuous space word representations. In *Proc. of NAACL-HLT*.

George A Miller and Walter G Charles. 1991. Contextual correlates of semantic similarity. *Language and cognitive processes*.

Preslav Nakov and Marti Hearst. 2005. Using the web as an implicit training set: application to structural ambiguity resolution. In *Proc. of HLT-EMNLP*.

Jens Nilsson, Joakim Nivre, and Johan Hall. 2006. Graph transformations in data-driven dependency parsing. In *Proc. of ACL-COLING*.

Joakim Nivre, Marco Kuhlmann, and Johan Hall. 2009. An improved oracle for dependency parsing with online reordering. In *Proc. of IWPT*.

Sebastian Padó and Mirella Lapata. 2007. Dependency-based construction of semantic space models. *Computational Linguistics*.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *Proc. of EMNLP*.

Herbert Rubenstein and John B. Goodenough. 1965. Contextual correlates of synonymy. *Communications of the ACM*.

Dana Rubinstein, Effi Levi, Roy Schwartz, and Ari Rappoport. 2015. How well do distributional models capture different types of semantic knowledge? In *Proc. of ACL*.

Roy Schwartz, Omri Abend, Roi Reichart, and Ari Rappoport. 2011. Neutralizing linguistically problematic annotations in unsupervised dependency parsing evaluation. In *Proc. of ACL-HLT*.

Roy Schwartz, Omri Abend, and Ari Rappoport. 2012. Learnability-based syntactic annotation design. In *Proc. of COLING*.

Roy Schwartz, Oren Tsur, Ari Rappoport, and Moshe Koppel. 2013. Authorship attribution of micro-messages. In *Proc. of EMNLP*.

Roy Schwartz, Roi Reichart, and Ari Rappoport. 2014. Minimally supervised classification to semantic categories using automatically acquired symmetric patterns. In *Proc. of COLING*.

Roy Schwartz, Roi Reichart, and Ari Rappoport. 2015. Symmetric pattern based word embeddings for improved word similarity prediction. In *Proc. of CoNLL*.

Lucien Tesnière. 1959. *Éléments de syntaxe structurale*. Paris: K1incksieck.

Kristina Toutanova, Dan Klein, Christopher D Manning, and Yoram Singer. 2003. Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proc. of NAACL*.

Peter D. Turney. 2006. Similarity of semantic relations. *Computational Linguistics*.

Peter D. Turney. 2008. A uniform approach to analogies, synonyms, antonyms, and associations. In *Proc. of COLING*.

Dominic Widdows and Beate Dorow. 2002. A graph model for unsupervised lexical acquisition. In *Proc. of COLING*.

Dongqiang Yang and David M. W. Powers. 2006. Verb similarity on the taxonomy of wordnet. In *Proc. of GWC*.