

Taking Antonymy Mask off in Vector Space

Enrico Santus

The Hong Kong Polytechnic University
Hong Kong

e.santus@connect.polyu.hk

Qin Lu

The Hong Kong Polytechnic University
Hong Kong

csluqin@comp.polyu.edu.hk

Alessandro Lenci

Università di Pisa
Pisa, Italy

alessandro.lenci@ling.unipi.it

Chu-Ren Huang

The Hong Kong Polytechnic University
Hong Kong

churen.huang@polyu.edu.hk

Abstract

Automatic detection of antonymy is an important task in Natural Language Processing (NLP) for Information Retrieval (IR), Ontology Learning (OL) and many other semantic applications. However, current unsupervised approaches to antonymy detection are still not fully effective because they cannot discriminate antonyms from synonyms. In this paper, we introduce *APAnt*, a new Average-Precision-based measure for the unsupervised discrimination of antonymy from synonymy using Distributional Semantic Models (DSMs). *APAnt* makes use of Average Precision to estimate the extent and salience of the intersection among the most descriptive contexts of two target words. Evaluation shows that the proposed method is able to distinguish antonyms and synonyms with high accuracy across different parts of speech, including nouns, adjectives and verbs. *APAnt* outperforms the *vector cosine* and a baseline model implementing the *co-occurrence hypothesis*.

1 Introduction

Antonymy is one of the fundamental relations shaping the organization of the semantic lexicon and its identification is very challenging for computational models (Mohammad et al., 2008; Deese, 1965; Deese, 1964). Yet, antonymy is essential for many Natural Language Processing (NLP) applications, such as Information Retrieval (IR), Ontology Learning (OL), Machine Translation (MT), Sentiment Analysis (SA) and Dialogue Systems (Roth and Schulte im Walde, 2014; Mohammad et al., 2013). In particular, the automatic identification of semantic opposition is a crucial component for the detection and generation of paraphrases (Marton et al., 2011), the understanding of contradictions (de Marneffe et al., 2008) and the detection of humor (Mihalcea and Strapparava, 2005).

Several existing computational lexicons and thesauri explicitly encode antonymy, together with other semantic relations. Although such resources are often used to support the above mentioned NLP tasks, hand-coded lexicons and thesauri have low coverage and many scholars have shown their limits: Mohammad et al. (2013), for example, have noticed that “more than 90% of the contrasting pairs in GRE closest-to-opposite questions are not listed as opposites in WordNet”.

The automatic identification of semantic relations is a core task in computational semantics. Distributional Semantic Models (DSMs) have often been exploited for their well known ability to identify semantically similar lexemes using corpus-derived co-occurrences encoded as distributional vectors (Santus et al., 2014a; Baroni and Lenci, 2010; Turney and Pantel, 2010; Padó and Lapata, 2007; Sahlgren, 2006). These models are based on the *Distributional Hypothesis* (Harris, 1954) and represent lexical semantic similarity in function of distributional similarity, which can be measured by *vector cosine* (Turney and Pantel, 2010). However, these models are characterized by a major shortcoming. That is, they are not able to discriminate among different kinds of semantic relations linking distributionally similar lexemes. For instance, the nearest neighbors of *castle* in the vector space typically include hypernyms like *building*, co-hyponyms like *house*, meronyms like *brick*, antonyms like *shack*, together with other semantically related words. While impressive results have been achieved in the automatic identification of synonymy (Baroni and Lenci, 2010; Padó and Lapata, 2007), methods for the identification of hypernymy (Santus et al., 2014a; Lenci and Benotto, 2012) and antonymy (Roth and Schulte im Walde, 2014; Mohammad et al. 2013) still need much work to achieve satisfying precision and coverage (Turney, 2008; Mohammad et al., 2008). This is the reason why semi-supervised pattern-based approaches have often been preferred to purely unsupervised DSMs (Pantel and Pennacchiotti, 2006; Hearst, 1992).

In this paper, we introduce *APAnt*, a new Average-Precision-based distributional measures that is able to successfully discriminate antonyms from synonyms, outperforming *vector cosine* and a baseline system based on the *co-occurrence hypothesis*, formulated by Charles and Miller in 1989 and confirmed in other studies, such as those of Justeson and Katz (1991) and Fellbaum (1995).

Our measure is based on a distributional interpretation of the so-called *paradox of simultaneous similarity and difference between the antonyms* (Cruse, 1986). According to this paradox, antonyms are similar to synonyms in every dimension of meaning except one. Our hypothesis is that the different dimension of meaning is a salient one and it can be identified

with DSMs and exploited for discriminating antonyms from synonyms.

The rest of the paper is organized as follows. Section 2 gives the definition and illustrates the various types of antonyms. Section 3 gives a brief overview of related works. Section 4 presents the proposed *APAnt* measure. Section 5 shows the performance evaluation of the proposed measure. Section 6 is the conclusion.

2 Antonymy: definition and types

People do not always agree on classifying word pairs as antonyms (Mohammed et al., 2013), confirming that antonymy identification is indeed a difficult task. This is true even for native speakers. Antonymy is in fact a complex relation and opposites can be of different types, making this class hard to define (Cruse, 1986).

Over the years, many scholars from different disciplines have tried to provide a precise definition of this semantic relation. Though, they are yet to reach any conclusive agreement. Kempson (1977) defines opposites as word pairs with a “binary incompatible relation”, such that the presence of one meaning entails the absence of the other. In this sense, *giant* and *dwarf* are good opposites, while *giant* and *person* are not. Cruse (1986) points out the above-mentioned *paradox of simultaneous similarity and difference* between the antonyms, claiming that opposites are indeed similar in every dimension of meaning except in a specific one (e.g., both *giant* and *dwarf* refer to a person, with a head, two legs and two feet, but with very different size).

Mohammad et al. (2013) have used these two definitions to distinguish between (1) opposites, which are word pairs that are strongly incompatible with each other and/or are saliently different across a dimension of meaning; (2) contrasting word pairs, which have some non-zero degree of binary incompatibility and/or some non-zero difference across a dimension of meaning; (3) antonyms, which are opposites that are also gradable adjectives.

Semantic opposition is so complex that other classifications might be adopted as well (Bejar et al., 1991; Cruse, 1986). Moreover, opposites can also be sub-classified. Even though there is no agreement about the number of sub-types, we briefly mention a simple – but comprehensive –

sub-classification adopted by Mohammad et al. (2013) to exemplify the complexity of the class. In their paper, Mohammad et al. used a simple sub-classification to make their crowdsourcing annotation task easier to perform. This sub-classification, mostly based on Cruse (1986), includes (1) antipodals (e.g. *top-bottom*), pairs whose terms are at the opposite extremes of a specific meaning dimension; (2) complementaries (e.g. *open-shut*), pairs whose terms divide the domain in two mutual exclusive compartments; (3) disjoints (e.g. *hot-cold*), pairs whose words occupy non-overlapping regions in a specific semantic dimension; (4) gradable opposites (e.g. *long-short*), adjective- or adverb-pairs that gradually describe some semantic dimensions, such as length, speed, etc.; (5) reversibles (e.g. *rise-fall*), verb-pairs whose words respectively describe the change from A to B and the change from B to A.

Since our aim is to discriminate antonyms from synonyms, our attention is not focused on distinguishing different types of opposites. In this work, we will adopt a broad definition of antonymy, including all the previously mentioned types of opposites together with paronyms, which are a specific type of co-hyponyms (Huang et al., 2007). In fact, while co-hyponyms are simply coordinates depending from the same hypernym, paronyms are co-hyponyms partitioning a conceptual field in subfields. Different from co-hyponyms, paronyms must be very similar to each other and change only in respect to one dimension of meaning. For instance, *dry season*, *spring*, *summer*, *autumn* and *winter* are co-hyponyms, but only *spring*, *summer*, *autumn* and *winter* are paronyms.

3 Related Works

The foundation of most corpus-based research on antonymy is the *co-occurrence hypothesis*, (Lobanova, 2012). This derives from an observation by Charles and Miller (1989) that antonyms co-occur in the same sentence more often than expected by chance. This claim has found many empirical confirmations, such as by Justeson and Katz (1991) and Fellbaum (1995).

Another large part of related research has been focused on the study of lexical-syntactic constructions that can work as linguistic tests for antonymy definition and classification (Cruse,

1986). Some syntagmatic properties were also identified. Ding and Huang (2014; 2013), for example, have noticed that, unlike co-hyponyms, antonyms generally have a strongly preferred word order when they co-occur in a coordinate context (i.e. A and/or B).

Starting from these observations, computational methods for antonymy identification were implemented. Most of them rely on pattern based approaches (Schulte im Walde and Köper, 2013; Lobanova et al., 2010; Turney, 2008; Pantel and Pennacchiotti, 2006; Lin et al., 2003), which use specific patterns to distinguish antonymy-related pairs from others. Pattern based methods, however, are mostly semi-supervised. Moreover they require a large amount of data and suffer from low recall, because they can be applied only to frequent words, which are the only ones likely to occur with the selected patterns.

Lucerto et al. (2002) used the number of tokens between the target words together with some other clues (e.g. the presence/absence of conjunctions like *but*, *from*, *and*, etc.) in order to identify contrasting words. Unfortunately the method has very limited coverage.

Schwab et al. (2002) used oppositeness vectors, which were created by identifying possible opposites relying on dictionary definitions. The approach was tested only on a few word pairs and it can hardly be regarded as a general solution.

Turney (2008) proposed a supervised algorithm for the identification of several semantic relations, including synonyms and opposites. The algorithm relied on a training set of word pairs with class labels to assign the labels also to a testing set of word pairs. All word pairs were represented as vectors encoding the frequencies of co-occurrence in textual patterns extracted from a large corpus of web pages. The system achieved an accuracy of 75% against a frequency baseline of 65.4%.

Mohammad et al. (2008) proposed a method for determining what they have defined as the “degrees of antonymy”. This concept, which is related to the canonicity (Jones et al., 2007), was aimed to reflect the results of psycholinguistic experiments, which show that some antonyms are perceived as ‘better’ (e.g. *big – small*) than others (e.g. *big – normal*). For each target word pair, they used thesaurus categories to decide whether a pair is an instance of antonymy or not. Their method

then assigned the degree of antonymy using co-occurrence statistics, achieving a good precision.

Mohammad et al. (2013) used an analogical method based on a given set of contrasting words to identify and classify different kinds of opposites by hypothesizing that for every opposing pair of words, A and B, there is at least another opposing pair, C and D, such that A is similar to C and B is similar to D. Their approach outperformed other measures. But, it is not unsupervised and uses a thesaurus as knowledge.

Kim and de Marneffe (2013) exploited word vectors learned by Neural Network Language Models (NNLMs) to extract scalar relationships between adjectives (e.g., okay < good < excellent), outperforming other approaches in their indirect yes/no question answer pairs (IQAP) evaluation (de Marneffe et al., 2010).

Schulte im Walde and Köper (2013) proposed a vector space model relying on lexico-syntactic patterns to distinguish between synonymy, antonymy and hypernymy. Their approach was tested on German nouns, verbs and adjectives, achieving a precision of 59.80%, which was above the majority baselines.

More recently, Roth and Schulte im Walde (2014) proposed that discourse relations can be used as indicators for paradigmatic relations, including antonymy.

4 *APAnt*: an Average-Precision-based measure

In this work we make use of the observation that antonyms are often similar in every semantic dimension except one (Cruse, 1986). In the previous section we have shown the example of *giant* and *dwarf*, which in fact differ only with respect to *size*. This peculiarity of antonymy – called by Cruse (1986) the *paradox of simultaneous similarity and difference* – has an important distributional correlate. Antonyms, in fact, occur in similar contexts as much as synonyms do, making the DSMs models unable to discriminate them. However, according to Cruse's definition, we can expect one dimension of meaning in which the antonyms have different behaviors. That is, they occur with different contexts. We can also expect that this dimension of meaning is a salient one. For example, *size* is a salient dimension of meaning for the words *giant*

and *dwarf*, and we can expect that while *giant* occurs more often with words more related to large size such as *big*, *huge*, *destroy*, etc., *dwarf* is more likely to occur in contexts more related to small size, such as *small*, *hide*, and so on. We hypothesize, therefore, that if we isolate the N most salient contexts for two distributionally similar lexemes and we intersect them, we can predict whether these two lexemes are antonyms or synonyms by looking at the extent and salience of this intersection: the broader and more salient the intersection, the higher the probability that the lexemes are synonyms; *vice versa* the narrower and less salient the intersection, the higher the probability that the lexemes are antonyms.

To verify this hypothesis, we select the N most salient contexts of the two target words ($N=100$ ¹). We define the salience of a context for a specific target word by ranking the contexts through *Local Mutual Information* (LMI; Evert, 2005) and picking the first N , as already done by Santus et al. (2014a). Once the N most salient contexts for the two target words have been identified, we verify the extent and the salience of the contexts shared by both the target words. We predict that synonyms share a significantly higher number of salient contexts than antonyms.

To estimate the extent and the salience of the shared contexts, we adapt the Average Precision measure (AP; Voorhees and Harman, 1999), a common Information Retrieval (IR) evaluation metric already used by Kotlerman et al. (2010) to identify lexical entailment. In IR systems, this measure is used to evaluate the ranked documents returned for a specific query. It assigns higher values to the rankings in which most or all the relevant documents are on the top (recall), while irrelevant documents are either removed or in the bottom (precision). For our purposes, we modify this measure in order to increase the scores as a function of (1) the extent of the intersection between the N most relevant contexts of the two target words and (2) the maximum salience of the common contexts. To do so, we consider the common contexts as relevant documents and their maximum salience as their rank. Consequently,

¹ $N=100$ is the result of an optimization of the model against the dataset. Also the following suboptimal values have been tried: 50 and 150. In all the cases, the model outperformed the baseline.

when a common context is found, the score will be increased by a value that depends on the maximum salience of the context for the two target words. For instance, in the pair *dog-cat*, if *home* is a common context, and it has salience=1 for *dog* and salience= $N-1$ for *cat*, we will consider *home* as a relevant document with rank=1.

The equation (1) below provides the formal definition of *APAnt* measure:

$$APAnt = 1 / \sum_{f \in F_1 \cap F_2} \frac{1}{\min(\text{rank}_1(f_1), \text{rank}_2(f_2))} \quad (1)$$

where F_x is the set of the N most salient features of a term x and $\text{rank}_x(f_x)$ is the rank of the feature f_x in the salience ranked feature list for the term x . It is important to note that *APAnt* is defined as a reciprocal measure, so that higher scores are assigned to antonyms.

5 Experiments and Evaluation

The evaluation includes two parts. The first part is to examine the discrimination ability of our method through box-plot visualizations, which summarize the distributions of scores per relation. In the second part, the Average Precision measure (AP; Kotlerman et al., 2010) is used to compute the ability of our proposed measure to discriminate antonyms from synonyms for nouns, adjectives and verbs. For comparison, we compare our performance with the *vector cosine* scores and with a baseline model using co-occurrence frequency of the target pairs.

5.1 The DSM and the Dataset

In our experiments, we use a standard window-based DSM recording co-occurrences with context window of the nearest 2 content words both to the left and right of each target word. Co-occurrences are extracted from a combination of the freely available ukWaC and WaCkypedia corpora (with 1.915 billion and 820 million words, respectively) and weighted with LMI (Santus et al., 2014a).

To assess *APAnt*, we rely on a subset of English word pairs collected by Alessandro Lenci and Giulia Benotto in 2012/13 using Amazon Mechanical Turk, following the method described by Scheible and Schulte im Walde (2014). Among

the criteria used for the collection, Lenci and Benotto balanced target items across word categories and took in consideration the frequency, the degree of ambiguity and the semantic classes.

Our subset contains 2.232 word pairs², including 1.070 antonym pairs and 1.162 synonym pairs. The antonyms include 434 noun pairs (e.g. *parody-reality*), 262 adjective pairs (e.g. *unknown-famous*) and 374 verb pairs (e.g. *try-procrastinate*). The synonyms include 409 noun pairs (e.g. *completeness-entirety*), 364 adjective pairs (e.g. *determined-focused*) and 389 verb pairs (e.g. *picture-illustrate*).

5.2 Results

5.2.1 *APAnt* Values Distribution

Figure 1 and Figure 2 show the box-plots summarizing the logarithmic distributions of *APAnt* and baseline scores for antonyms and synonyms, respectively. The logarithmic distribution is used to smooth the range of data, which would otherwise be too large and sparse for the box-plot representation. Figure 3 shows the box-plot summarizing the *vector cosine* scores. Since *vector cosine* scores range between 0 and 1, we multiplied them by ten to scale up for comparison with the other two box-plots in Figure 1 and Figure 2.

Box-plots display the median of a distribution as a horizontal line within a box extending from the first to the third quartile, with whiskers covering 1.5 of the interquartile range in each direction from the box, and outliers plotted as circles.

The box-plots in Figure 1, Figure 2 and Figure 3 include test data with all part of speech types (i.e. nouns, adjectives and verbs). The box-plots for individual parts of speech are not reported in the paper because they do not show significant differences.

² The sub-set includes all the pairs for which both the target words exist in the DSM.

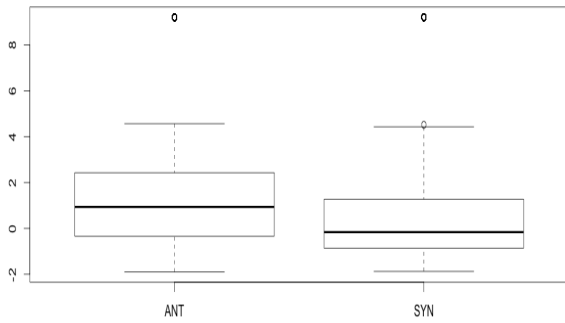


Figure 1: Logarithmic distribution of *APAnt* scores for antonym and synonym pairs ($N=100$) across nouns, adjectives and verbs.

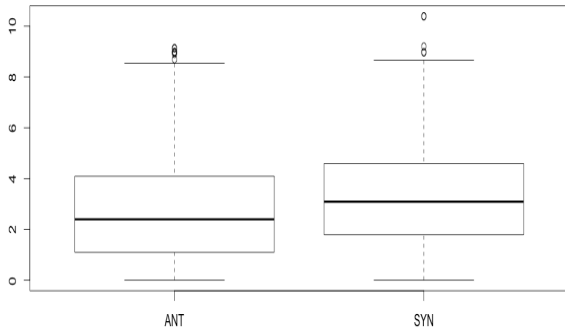


Figure 2: Logarithmic distribution of the baseline scores for antonym and synonym pairs across nouns, adjectives and verbs³.

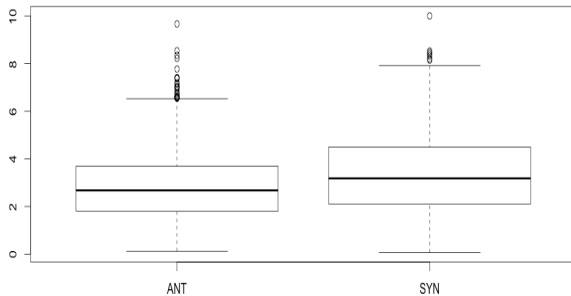


Figure 3: Distribution of the *vector cosine* scores for antonym and synonym pairs across nouns, adjectives and verbs⁴.

³ 410 pairs with co-occurrence equal to zero on a total of 2.232 have been removed to make the box-plot readable, because $\log(0) = -inf$

⁴ Since *vector cosine* scores range between 0 and 1, we multiplied them by ten to scale up for comparison with the other two box-plots in Figure 1 and Figure 2.

The more the boxes in the plot overlap, the less distinctive the measure is. In Figure 2 and Figure 3, we can observe that the baseline and the *vector cosine* tend to promote synonyms on antonyms, and also that there is a large range of overlap among synonyms and antonyms distributions, showing the weakness of these two measures for discriminate antonyms from synonyms. On the other hand, in Figure 1 we can observe that *APAnt* scores are much higher for antonymy-related pairs. In terms of distribution of values, in fact, synonyms have much lower values in *APAnt*. This shows that *APAnt* is clearly more biased towards antonym, differently from the *vector cosine* or the simple co-occurrence. Moreover, results also suggest the partial inaccuracy of the *co-occurrence hypothesis*. The tendency of co-occurring is not a hallmark of antonyms, but it is a property shared by synonyms too.

5.2.2 Average Precision

Table 1 shows the second performance measure we used in our evaluation, the Average Precision (Santus et al., 2014a; Lenci and Benotto, 2012; Kotlerman et al., 2010) computed for *APAnt*, baseline and *vector cosine* scores. As already mentioned above, AP is a measure used in Information Retrieval to combine precision, relevance ranking and overall recall. The best possible score we can obtain is 1 for antonymy and 0 for synonymy, which would correspond to the perfect discrimination between antonyms and synonyms.

<i>ALL PoS</i>	ANT	SYN
<i>APAnt</i> , $N=50$	0.71	0.57
<i>APAnt</i>, $N=100$	0.73	0.55
<i>APAnt</i> , $N=150$	0.72	0.55
Baseline	0.56	0.74
Cosine	0.55	0.75

Table 1: Average Precision (AP) values per relation for *APAnt* ($N=50$, 100 and 150), baseline and *vector cosine* across the parts of speech.

APAnt performs the best, compared to the reference methods, which mostly promote synonyms on antonyms. In fact, *APAnt* ($N=100$) is at the same time able (i) to better identify antonyms (+0.17 in comparison to the baseline and +0.18 over the *vector cosine*) and (ii) to better discriminate them from synonyms (-0.19 with respect to the baseline and -0.20 in comparison to the *vector cosine*). Regardless the value of N (either equal to 50, 100 or 150), *APAnt* clearly outperforms the baseline and the *vector cosine* by an identification improvement ranging from 26.7% ($N=50$ to baseline) to 32.7% ($N=100$ to *vector cosine*). These values confirm the trend shown in the box-plots of Figure 1, Figure 2 and Figure 3, proving that *APAnt* is a very effective measure to distinguish antonymy from synonymy.

Below we also list the AP values for the different parts of speech (i.e. nouns, adjectives and verbs) with the parameter $N=100$. As it can be observed, *APAnt* always outperforms the baseline. However, a slightly lower performance can be noticed in Table 3, where the AP scores for adjectives are 0.65 for both antonyms and synonyms.

<i>NOUNS</i>	ANT-N	SYN-N
<i>APAnt</i> , $N=100$	0.79	0.48
Baseline	0.53	0.77
Cosine	0.54	0.74

Table 2: Average Precision (AP) values per relation for *APAnt*, baseline and *vector cosine* on nouns.

<i>ADJECTIVES</i>	ANT-J	SYN-J
<i>APAnt</i> , $N=100$	0.65	0.65
Baseline	0.57	0.74
Cosine	0.58	0.73

Table 3: Average Precision (AP) values per relation for *APAnt*, baseline and *vector cosine* on adjectives.

<i>VERBS</i>	ANT-V	SYN-V
<i>APAnt</i> , $N=100$	0.74	0.52
Baseline	0.53	0.75
Cosine	0.52	0.77

Table 4: Average Precision (AP) values per relation for *APAnt*, baseline and *vector cosine* on verbs.

A possible explanation of this result might be that the different number of pairs per relation influences the AP values. In our dataset, in fact, we have 364 synonymy-related pairs against 262 antonym pairs for adjectives (+102 synonymy-related pairs, +39%).

To test this hypothesis, we randomly extract 262 synonymy-related pairs from the 364 that are present in our dataset and we re-calculate the AP scores for both the relations. The results can be found in Table 5.

<i>ADJECTIVES</i>	ANT-J	SYN-J
<i>APAnt</i> , $N=100$	0.72	0.60
Baseline	0.66	0.69
Cosine	0.68	0.66

Table 5: Average Precision (AP) values per relation for *APAnt*, baseline and *vector cosine* on adjectives, after extracting 262 pairs per relation.

The results in Table 5 confirm that *APAnt* works properly also for adjectives. It is in fact able to better identify antonyms (+0.06 on the baseline and +0.04 on *vector cosine*) and to better discriminate them from synonyms (-0.09 on the baseline and -0.06 on *vector cosine*). However, this is the lowest result among the three parts of speech used in our experiments.

The different results for the three parts of speech should be interpreted in relation to our hypothesis. It is in fact possible that while opposing nouns (e.g. *giant* – *dwarf*) share very few or none salient contexts, opposing verbs (e.g. *rise* – *fall*) and – even more – opposing adjectives (e.g. *hot* – *cold*) share some salient contexts, making the discrimination task more difficult for these parts of

speech. In any case, the accuracy of our method has strongly outperformed the baseline for all the parts of speech, confirming the robustness of our hypothesis.

6 Conclusions and Ongoing Work

This paper introduces *APAnt*, a new distributional measure for the identification of antonymy based on a distributional interpretation of the *paradox of simultaneous similarity and difference between the antonyms* (preliminary results about *APAnt* were published by Santus et al., 2014b, at CLIC-IT conference).

APAnt is evaluated in a discrimination task in which both antonymy- and synonymy-related pairs are present. The evaluation has been performed on nouns, adjectives and verbs. In the task, *APAnt* has outperformed the *vector cosine* and the baseline implementing the *co-occurrence hypothesis* (Fellbaum, 1995; Justeson and Katz, 1991; Charles and Miller, 1989) for all the parts of speech, achieving good accuracy for all of them. However, its performance is higher for nouns, slightly lower for verbs and significantly lower for adjectives. These differences across parts of speech might be due to the fact that while opposing nouns share very few salient contexts, opposing verbs and – even more – opposing adjectives share some salient contexts, making the discrimination task more difficult. In all the cases, however, *APAnt* performance supports our hypothesis, according to which synonyms share a number of salient contexts that is significantly higher than the one shared by antonyms.

Moreover, following Santus et al. (2014a), we did not work with the full set of contexts of the target words, but only a subset of the N most salient ones. We assume, in fact, that they better describe the relevant distributional behavior of a specific term, while considering the full set would include also much noise. The N most salient contexts were selected after having been ranked through LMI (Evert, 2005). This method can be certainly applied for the study of other semantic relations.

Ongoing research includes the application of *APAnt* to discriminate antonymy also from other semantic relations and to automatically extract antonymy-related pairs for the population of

ontologies and lexical resources. Further work can be conducted to apply *APAnt* to other languages.

Acknowledgments

This work is partially supported by HK PhD Fellowship Scheme under PF12-13656.

References

- Baroni, Marco and Alessandro Lenci. 2010. Distributional Memory: A general framework for corpus-based semantics. *Computational Linguistics*, 36(4):673–721.
- Bejar, Isaac I., Roger Chaffin, and Susan Embretson. 1991. *Cognitive and Psychometric Analysis of Analogical Problem Solving*. Springer-Verlag, New York.
- Charles, Walter G. and George A. Miller. 1989. Contexts of antonymous adjectives. *Applied Psychology*, 10:357–375.
- Cruse, David A. 1986. *Lexical Semantics*. Cambridge University Press, Cambridge.
- Evert, Stefan. 2005. *The Statistics of Word Cooccurrences*. Dissertation, Stuttgart University.
- de Marneffe, Marie-Catherine, Christopher D. Manning, and Christopher Potts. 2010. Was it good? It was provocative. Learning the meaning of scalar adjectives. In *Proceedings of the 48th Meeting of the Association for Computational Linguistics (ACL-10)*, pages 167–176.
- de Marneffe, Marie-Catherine, Anna Rafferty, and Christopher D. Manning. 2008. Finding contradictions in text. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics (ACL-08)*, pages 1039–1047, Columbus, OH.
- Deese, J. 1964. The Associative Structure of Some Common English Adjectives. *Journal of Verbal Learning and Verbal Behavior*, 3:347–57.
- Deese J. 1965. *The Structure of Associations in Language and Thought*. Johns Hopkins University Press, Baltimore.
- Ding, Jing and Chu-Ren Huang. 2014. Word Order in Chinese Opposite Compounds. Presented at the *15th Chinese Lexical Semantics Workshop*. Macau: University of Macau. 10-12 June.
- Ding, Jing and Chu-Ren Huang. 2013. Markedness of Opposite. Pp. 191-195. In Pengyuan Liu and Qi Su Eds. *Chinese Lexical Semantics*. LNAI 8229. Springer, Heidelberg.
- Fellbaum, Christiane. 1995. Co-occurrence and antonymy. *International Journal of Lexicography*, 8:281–303.
- Harris, Zellig. 1954. Distributional structure. *Word*, 10(23):146–162.
- Hearst, Marti. 1992. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the Fourteenth International Conference on Computational Linguistics*, pages 539–546, Nantes.
- Huang, Chu-Ren, I-Li Su, Pei-Yi Hsiao, and Xiu-Ling Ke. 2007. Paronyms, Co-Hyponyms and Antonyms: Representing Semantic Fields with Lexical Semantic Relations. In *Proceedings of Chinese Lexical Semantics Workshop 2007*, Hong Kong Polytechnic University, May 20-23.
- Jones, Steven, Carita Paradis, M. Lynne Murphy, and Caroline Willners. 2007. Googling for ‘opposites’: A web-based study of antonym canonicity. *Corpora*, 2(2):129–154.
- Justeson, John S. and Slava M. Katz. 1991. Co-occurrences of antonymous adjectives and their contexts. *Computational Linguistics*, 17:1–19.
- Kempson, Ruth M. 1977. *Semantic Theory*. Cambridge University Press, Cambridge.
- Kim, Joo-Kyung and Marie-Catherine de Marneffe. 2013. Deriving Adjectival Scales from Continuous Space Word Representations. In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP 2013)*, pages 1625–1630.
- Kotlerman, Lili, Ido Dagan, Idan Szpektor, and Maayan Zhitomirsky-Geffet. 2010. Directional Distributional Similarity for Lexical Inference. *Natural Language Engineering*, 16(4):359–389.
- Lenci, Alessandro and Giulia Benotto. 2012. Identifying hypernyms in distributional semantic spaces. In *SEM 2012 – The First Joint Conference on Lexical and Computational Semantics*, 2:75–79, Montréal, Canada.
- Lin, Dekang, Shaojun Zhao, Lijuan Qin, and Ming Zhou. 2003. Identifying synonyms among distributionally similar words. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI-03)*, pages 1,492–1,493, Acapulco.
- Lobanova, Anna. 2012. *The Anatomy of Antonymy: a Corpus-driven Approach*. Dissertation. University of Groningen.
- Lobanova, Anna, Tom van der Kleij, and Jennifer Spenader. 2010. Defining antonymy: A corpus-based study of opposites by lexico-syntactic patterns. *International Journal of Lexicography*, 23(1):19–53.
- Lucerto, Cupertino, David Pinto, and Héctor Jiménez-Salazar. 2002. An automatic method to identify antonymy. In *Workshop on Lexical Resources and*

- the Web for Word Sense Disambiguation*, pages 105–111, Puebla.
- Marton, Yuval, Ahmed El Kholy, and Nizar Habash. 2011. Filtering antonymous, trend-contrasting, and polarity-dissimilar distributional paraphrases for improving statistical machine translation. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, pages 237–249, Edinburgh.
- Mihalcea, Rada and Carlo Strapparava. 2005. Making computers laugh: Investigations in automatic humor recognition. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 531–538, Vancouver.
- Mohammad, Saif, Bonnie Dorr, Graeme Hirst, and Peter D. Turney. 2013. Computing lexical contrast. *Computational Linguistics*, 39(3):555–590.
- Mohammad, Saif, Bonnie Dorr, and Graeme Hirst. 2008. Computing word-pair antonymy. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2008)*, pages 982–991, Waikiki, HI.
- Padó, Sebastian and Lapata, Mirella. 2007. Dependency-based Construction of Semantic Space Models. *Computational Linguistics*, 33(2):161–199.
- Pantel, Patrick and Marco Pennacchiotti. 2006. Espresso: Leveraging generic patterns for automatically harvesting semantic relations. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 113–120, Sydney, Australia.
- Roth, Michael and Sabine Schulte im Walde. 2014. Combining word patterns and discourse markers for paradigmatic relation classification. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL)*, 2:524–530, Baltimore, Maryland, USA.
- Sahlgren, Magnus. 2006. *The Word-Space Model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces*. Ph.D. dissertation, Department of Linguistics, Stockholm University.
- Santus, Enrico, Alessandro Lenci, Qin Lu, and Sabine Schulte im Walde. 2014a. Chasing Hypernyms in Vector Spaces with Entropy. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, 2:38–42, Gothenburg, Sweden.
- Santus, Enrico, Qin Lu, Alessandro Lenci, and Chu-Ren Huang. 2014b. Unsupervised Antonym-Synonym Discrimination in Vector Space. In *Atti della Conferenza di Linguistica Computazionale Italiana (CLIC-IT)*, Pisa, Italy.
- Scheible, Silke and Sabine Schulte im Walde. 2014. A Database of Paradigmatic Semantic Relation Pairs for German Nouns, Verbs, and Adjectives. In *Proceedings of the COLING Workshop on Lexical and Grammatical Resources for Language Processing*, Dublin, Ireland, August 2014.
- Schulte im Walde, Sabine and Maximilian Köper. 2013. Pattern-based distinction of paradigmatic relations for German nouns, verbs, adjectives. In *Language Processing and Knowledge in the Web*, 184–198. Springer.
- Schwab, Didier, Mathieu Lafourcade, and Violaine Prince. 2002. Antonymy and conceptual vectors. In *Proceedings of the 19th International Conference on Computational Linguistics (COLING-02)*, pages 904–910, Taipei, Taiwan.
- Turney, Peter D. and Patrick Pantel. 2010. From Frequency to Meaning: Vector Space Models of Semantics. *Journal of Artificial Intelligence Research*, 37:141–188.
- Turney, Peter D. 2008. A uniform approach to analogies, synonyms, antonyms, and associations. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING-08)*, pages 905–912, Manchester.
- Voorhees, Ellen M., and Donna K. Harman. 1999. *The Seventh Text REtrieval Conference (TREC-7)*, 7, National Institute of Standards and Technology, NIST.