Distinguishing Paradigmatic Semantic Relations: Human Ratings and Distributional Similarity

Sabine Schulte im Walde
Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart
schulte@ims.uni-stuttgart.de

1 Introduction

Paradigmatic semantic relations such as *synonymy, antonymy, hypernymy* and *(co-)hyponymy* are central to the organisation of the mental lexicon (Miller & Fellbaum, 1991; Murphy, 2003; among others), by providing a structure for the lexical concepts that words express. The relational structure differs across word classes: According to Miller & Fellbaum (1991), *"no single set of semantic relations […] is adequate for the entire lexicon: nouns, adjectives, and verbs each have their own semantic relations and their own organisation determined by the role they must play in the construction of linguistic messages."* For example, while hypernymy is a natural relation for organising the noun lexicon, it is of minor importance for verbs, and unnatural for adjectives. In contrast, antonymy is considered the central relation for organising the adjective lexicon, and antonymy also plays a role in the mental lexicon for verbs, next to hypernymy, synonymy and entailment.

From a computational point of view, distinguishing between paradigmatic semantic relations is important for any application in Natural Language Processing (NLP) such as machine translation and textual entailment, which go beyond a general notion of semantic relatedness and require to identify specific semantic relations. Distributional vector space models offer a means to determine the meaning and the semantic „similarity" of target words within a geometric setting (Turney & Pantel, 2010). Such distributional models rely on the distributional hypothesis and exploit corpus co-occurrences in vector space models to describe and compare the meanings of linguistic units such as words, phrases and sentences (Harris, 1954). Paradigmatic relations are notoriously difficult to distinguish by standard distributional models, however, because the first-order co-occurrence distributions of the related words tend to be very similar across the relations. For example, in *"The boy/girl/person loves/hates the cat."*, the nominal co-hyponyms *boy* and *girl* and their hypernym *person* as well as the verbal antonyms *love* and *hate* occur in identical contexts, respectively.

Our research brings together perspectives from cognitive semantics and distributional semantics, and explores and compares the distinction of paradigmatic semantic relations across the three word classes of nouns, verbs and adjectives. We expect that differences in how natural relations are across word classes are reflected in (a) how humans perceive and distinguish semantic relatedness, and in (b) to what extent corpus-based approaches are successful.

For the cognitive perspective (Section 2), we rely on an existing dataset of paradigmatic semantic relation pairs rated for their relation strength, and demonstrate differences in relation distinction across word classes by humans. For the computational perspective (Section 3), we rely on distributional similarity scores from a standard word space model, as obtained from a large web corpus, and demonstrate (i) that unprocessed distributional similarity is indeed a difficult starting point for distinguishing paradigmatic relations, but (ii) that there are also differences across word classes, and (iii) that even simple classification models are able to exploit the differences successfully. Across the perspectives, we discuss the *naturalness* of the individual relations regarding the specific word classes.

2 Human Ratings on Paradigmatic Relations

For our cognitive explorations, we used a dataset of target-response paradigmatic semantic relation pairs from Scheible & Schulte im Walde (2014). The targets were randomly selected nouns, verbs and adjectives from GermaNet (Hamp & Feldweg, 1997), balanced according to semantic category, polysemy, and type frequency. A first experiment hosted by Amazon Mechanical Turk[1] (AMT) collected synonyms, antonyms and hypernyms for these targets. For example, for the target verb *befehlen* ('to command'), participants proposed antonyms such as *gehorchen* ('to obey'), synonyms such as *anordnen* ('to order'), and hypernyms such as *sagen* ('to say'). Table 1 shows some examples of generated word pairs across word classes and relations, and how many out of 10 participants provided this target-response pair. In total, we collected 8,910 pair tokens across 5,745 pair types.

	ANT		SYN		HYP	
NOUN	*Bein/Arm* (leg/arm)	10	*Killer/Mörder* (killer)	8	*Ekel/Gefühl* (disgust/feeling)	7
	Zeit/Raum (time/space)	3	*Gerät/Apparat* (device)	3	*Arzt/Beruf* (doctor/profession)	5
VERB	*verbieten/erlauben* (forbid/allow)	10	*üben/trainieren* (practise)	6	*trampeln/gehen* (lumber/walk)	6
	setzen/stehen (sit/stand)	4	*setzen/platzieren* (place)	3	*wehen/bewegen* (wave/move)	3
ADJ	*dunkel/hell* (dark/light)	10	*mild/sanft* (smooth)	9	*grün/farbig* (green/colourful)	5
	heiter/trist (cheerful/sad)	2	*bekannt/vertraut* (familiar)	4	*heiter/hell* (bright/light)	1

Table 1. Examples of generated target-response pairs.

In a second experiment participants were asked to rate the strength of a given semantic relation with respect to a word pair on a 6-point scale: [0,5]. For example, workers were presented with the generated pair *befehlen-gehorchen* ('command-obey') and asked to rate the strength of synonymy, antonymy, and hypernymy between the words. For this experiment, we only used a sub-set of the generated target-response pairs. All word pairs were assessed with respect to all three relation types, in separate experiment runs. For example, the pair *befehlen-gehorchen* above, which was generated as an antonymy pair by six participants in the first experiment, received an average rating score of 4.4 regarding antonymy, an average score of 0.3 regarding hypernymy, and an average score of 0.1 regarding synonymy. The average score was the mean over 10 participants' judgements. In total, we obtained 1,684 word pairs with all three ratings from the second experiment.

1 https://www.mturk.com

In order to assess how well the experiment participants could distinguish between the paradigmatic relations, we calculated the differences in mean ratings for a specific relation pair. Taking the example above, with a mean rating of 4.4 for the antonym pair *befehlen-gehorchen* regarding antonymy, and a mean rating of 0.3 regarding hypernymy, the difference in ratings was 4.1. In contrast, the difference in mean ratings for the antonym pair *bedürfen-verzichten* ('require-abstain') was only 2.1, demonstrating that this antonym pair was less clear.

Figures 1 and 2 present the mean differences between the mean ratings across relation pairs, for each word class. Figure 1 does not distinguish between the direction of the original relation and the rated relation, but Figure 2 does. The figures illustrate that the experiment participants found it easier to distinguish between antonyms and synonyms as well as between antonyms and hypernyms, in comparison to distinguishing between hypernyms and synonyms. This finding holds across word classes. We can also see that the easier distinctions are stronger for adjectives and verbs in comparison to nouns. Distinguishing between hypernyms and synonyms is almost impossible for adjectives and verbs, for which hypernymy represents an extremely unnatural semantic relation.

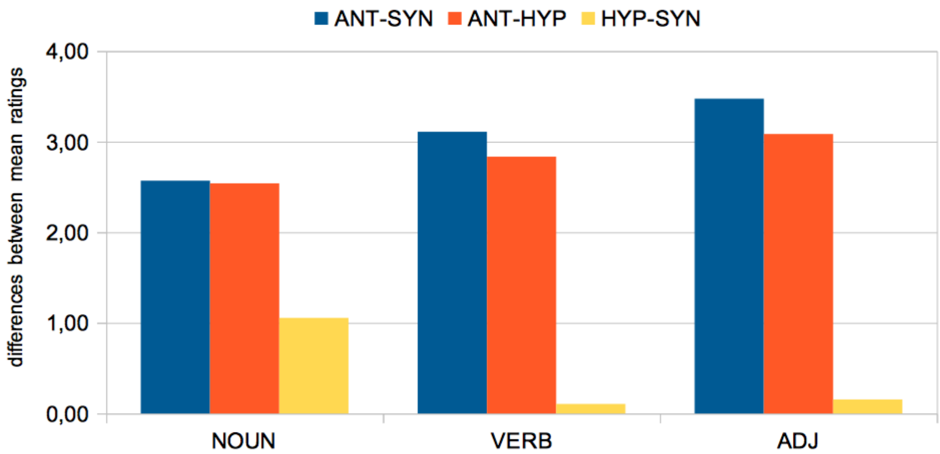

Figure 1. *Human distinction of paradigmatic relation pairs (coarse).*

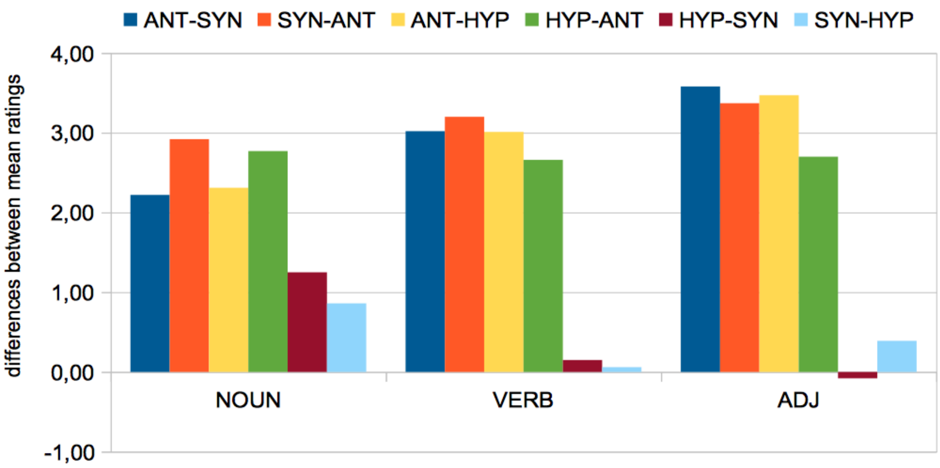

Figure 2. *Human distinction of paradigmatic relation pairs (fine).*

3 Distributional Similarity of Paradigmatic Relations

Regarding distributional information and distributional similarity, we explored various levels of processing co-occurrence information. We started out with unprocessed cosine similarities between the two words in the target-response pairs, in order to illustrate the difficult basis of a distributional model for paradigmatic relations (Section 3.1). In a series of classification experiments, we then automatically categorised the target-response pairs into semantic relations (Section 3.2).

3.1 Cosine Similarity

A vector space model uses corpus co-occurrences to activate and quantify dimensions in word vectors. The geometric distance between two word vectors determines the similarity between the two words. The closer two vectors are in vector space, the more semantically related we expect the underlying words to be. Regarding paradigmatic semantic relations, all word pairs are expected to be close to each other in word space to some extent. In the following, we demonstrate that this is indeed the case.

Our computational model uses a standard vector space to initiate all targets and responses of our relation pairs: We relied on a sentence-internal 20-word co-occurrence window (i.e., looking 20 words to the left and 20 words to the right of a word in the corpus to determine the co-occurring words), exploiting one of the currently largest German web corpora, the DECOW14AX with approx. 12 billion words (Schäfer & Bildhauer, 2012). We weighted the co-occurrence frequencies by applying *local mutual information*. Based on these co-occurrence vectors, we calculated the cosine scores of the target-response pairs, for the same subset of rated pairs underlying Figures 1 and 2. Figure 3 presents the resulting boxplots of cosine scores across word classes and semantic relations. The plots illustrate that the cosine values across the relations are indeed very similar for a specific word class. For synonymy, they tend to be higher than for the other two relations.

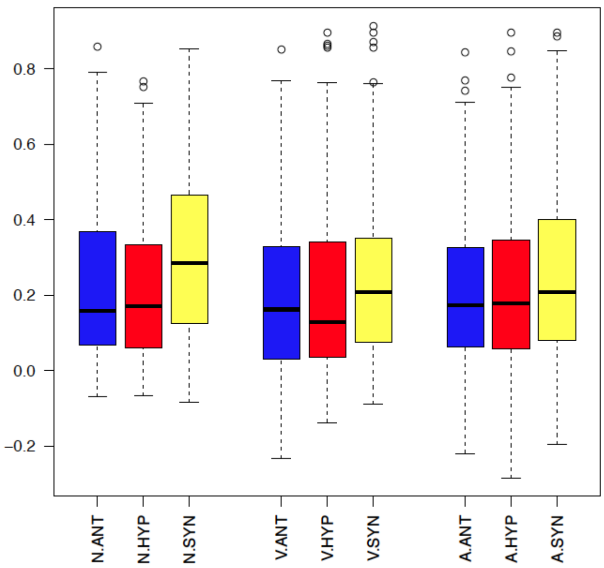

Figure 3. *Boxplots of cosine scores across classes and relations.*

3.2 Distributional Classification Models

In a series of classification experiments, we explored whether automatic methods relying on the distributional word spaces are able to categorise word pairs according to their paradigmatic semantic relations, even though the word pairs are all very close in space. In the following, we present the classification results of a simple nearest-centroid classifier (also known as *Rocchio classifier*) that compared window-based co-occurrence features as used in the previous section against pattern-based co-occurrence features (Schulte im Walde & Köper, 2013; David, 2014). As corpus resource, we relied on the *SdeWaC* (Faaß & Eckart, 2013). The pattern-based approach uses a vector space model generated in the tradition of lexico-syntactic patterns, i.e., linear word sequences between two co-occurring words in the corpus.

The classification was performed as follows. For each word class, we calculated three mean vectors, one for each lexical semantic relation (antonymy, synonymy, hypernymy), as based on a set of training pairs. We then predicted the semantic relation for a set of test pairs in each word class, by choosing for each test pair the most similar mean vector, as determined by the cosine between the vectors. Regarding window-based co-occurrences, we used three variants to initiate the mean values of the semantic classes: *Window-COS:* The class means were computed as the mean cosine scores of the training relation word pairs (which represents a somewhat unrealistic scenario, as it assumes a specific cosine cut-off value that distinguishes between the semantic relations, cf. Figure 3). *Window-DIFF:* For each target-response word pair, we calculated the difference vector, and the class means were computed as the means of the difference vectors. *Window-PROD:* For each target-response word pair, we calculated the product vector, and the class means were computed as the means of the product vectors. Across experiments, we used 5-fold cross-validation for training and testing.

Figure 4 presents the results of the nearest-centroid classifier, across word classes, relations, and co-occurrence information. As expected, the pattern information (which represents the standard co-occurrence information to distinguish between semantic relations) in most cases outperforms the window information. We can also see, however, that the most salient features depend on the word class: regarding verbs, the difference between window and pattern approaches vanishes.

Looking at the most successful classifiers to distinguish relations within word classes, we find that hypernymy is a clear relation for nouns (participating in the two best relation pairs ANT-HYP and HYP-SYN), and that antonymy is a clear relation for adjectives (participating in the two best relation pairs ANT-SYN and ANT-HYP). Interestingly, while there is no direct mapping of these results to the human relation distinction in Figures 1 and 2, these relations are the most natural semantic relation types for the two word classes, respectively. Comparing the results to the cosine boxplots in Figure 3, there is also no direct mapping: Based on Figure 3, we would expect synonymy to be identified best across word classes, but the classifier obviously is able to abstract over the raw cosine scores.

Overall, we were able to show various insights into relation distinction (from the cognitive and the computational perspective), and while we did not find close correspondences between the two perspectives, the automatic classification was best for the most natural relations for nouns and adjectives.

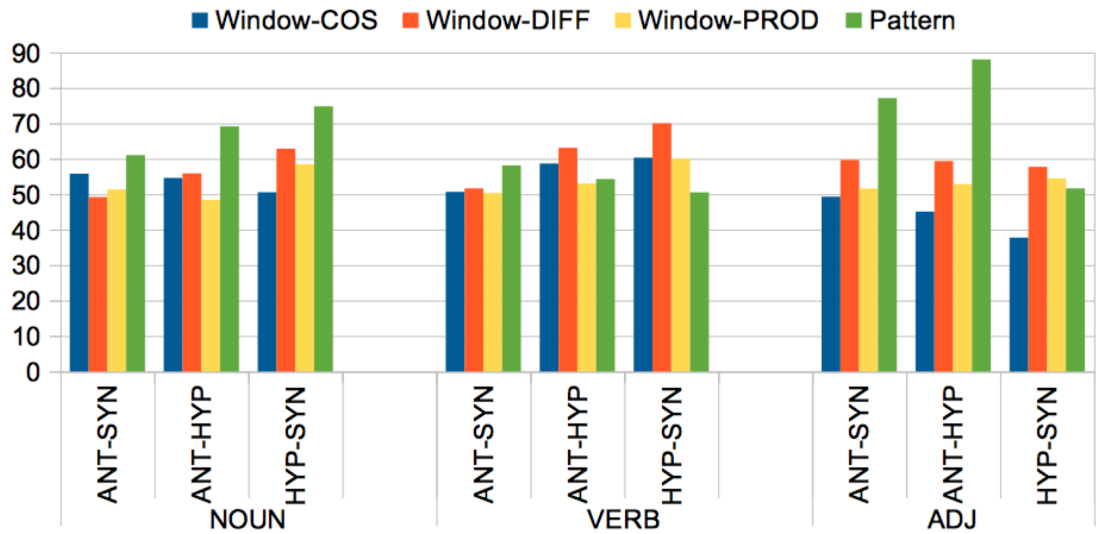

Figure 4. Results of automatic classification, using window and pattern features.

References

David, Benjamin (2014). *Comparison and Combination of Feature Types for the Automatic Classification of Semantic Relation Pairs.* Diplomarbeit, Stuttgart.

Gertrud Faaß and Kerstin Eckart (2013). SdeWaC, a Corpus of Parsable Sentences from the Web. In: *Proceedings of the 25st GSCL Conference,* pp. 184-198.

Birgit Hamp & Helmut Feldweg (1997). GermaNet – A Lexical-Semantic Net for German. In: *Proceedings of the Workshop on Automatic Information Extraction and Building Lexical Semantic Resources for NLP Applications,* pp. 9-15.

George A. Miller & Christiane Fellbaum (1991). Semantic Networks of English. *Cognition* 41:197-229.

M. Lynne Murphy (2003). *Semantic Relations and the Lexicon.* CUP.

Roland Schäfer & Felix Bildhauer (2012). Building Large Corpora from the Web using a New Efficient Tool Chain. In: *Proceedings of the 8th International Conference on Language Resources and Evaluation*, pp. 486-493.

Silke Scheible & Sabine Schulte im Walde (2014). A Database of Paradigmatic Semantic Relation Pairs for German Nouns, Verbs and Adjectives. In: *Proceedings of the Workshop on Lexical and Grammatical Resources for Language Processing,* pp. 111-119.

Sabine Schulte im Walde & Maximilian Köper (2013). Pattern-based Distinction of Paradigmatic Relations for German Nouns, Verbs, Adjectives. In: *Proceedings of the 25st GSCL Conference,* pp. 184-198.

Peter D. Turney & Patrick Pantel (2010). From Frequency to Meaning: Vector Space Models of Semantics. *Journal of Artificial Intelligence Research* 37:141-188.