German Perception Verbs: Automatic Classification of Prototypical and Multiple Non-literal Meanings

Benjamin David, Sylvia Springorum, Sabine Schulte im Walde
Institut für Maschinelle Sprachverarbeitung
Universität Stuttgart
{benjamin.david,sylvia.springorum,schulte}@ims.uni-stuttgart.de

Abstract

This paper presents a token-based automatic classification of German perception verbs into literal vs. multiple non-literal senses. Based on a corpus-based dataset of German perception verbs and their systematic meaning shifts, we identify one verb of each of the four perception classes *optical, acoustic, olfactory, haptic*, and use Decision Trees relying on syntactic and semantic corpus-based features to classify the verb uses into 3-4 senses each. Our classifier reaches accuracies between 45.5% and 69.4%, in comparison to baselines between 27.5% and 39.0%. In three out of four cases analyzed our classifier's accuracy is significantly higher than the according baseline.

1 Introduction

In contrast to Word Sense Disambiguation in general (cf. Agirre and Edmonds (2006); Navigli (2009)), most *computational approaches to modelling literal vs. non-literal meaning* are still restricted to a binary distinction between two sense categories (literal vs. non-literal), rather than between multiple literal and non-literal senses. For example,[1] Bannard (2007), and Fazly et al. (2009) identified light verb constructions as non-literal verb uses; Birke and Sarkar (2006), Birke and Sarkar (2007), Sporleder and Li (2009), and Li and Sporleder (2009) distinguished literal vs. idiomatic meaning. Concerning metonymic language, most approaches address various senses, which are however very restricted to two domains, locations and organizations (Markert and

Nissim, 2002; Nastase and Strube, 2009; Nastase et al., 2012). One of the few studies going beyond a binary classification is represented by Shutova et al. (2013) who classified literal vs. metaphorical verb senses on a large scale and for multiple non-literal meanings. Cook and Stevenson (2006) also took multiple sense distinctions into account, focusing on English 'up' particle verbs.

In this paper,[2] we address the automatic classification of German perception verbs into literal vs. non-literal meanings. Our research goes beyond a binary classification and distinguishes between multiple non-literal senses. Taking the PhD thesis by Ibarretxe-Antunano (1999) as a starting point, a preparatory step places German perception verbs into four classes: optical, acoustic, olfactory and haptic. In the main part, we then choose one perception verb from each class ('betrachten', 'hören', 'wittern', 'spüren')[3] which each have multiple literal/non-literal senses, and rely on syntactic and semantic corpus-based features and a Decision Tree classifier to perform a token-based assignment to senses. We address both a binary (literal vs. non-literal) and a multiple sense discrimination.

The paper describes related work in Section 2, specifies the perception verbs and their features in Sections 3 and 4, and performs automatic token-based word sense classification in Section 5.

---

[1] See Section 2 for details on related work.

[3] Since the verbs have multiple meanings, we do not translate them here but in Section 3.

2 Related Work

Computational work on non-literal meaning comprises research from various sub-fields. Approaches to *light verb constructions* (Bannard, 2007; Fazly et al., 2007; Fazly et al., 2009) relied on measures of syntactic variation of phrases, in combination with standard association measures, to perform a type-based classification. Approaches to *literal vs. non-literal/figurative/idiomatic meaning* performed binary classifications (Birke and Sarkar, 2006; Birke and Sarkar, 2007; Sporleder and Li, 2009; Li and Sporleder, 2009), relying on various contextual indicators: Birke and Sarkar exploited seed sets of literal vs. non-literal sentences, and used distributional similarity to classify English verbs. Li and Sporleder defined two models of text cohesion to classify V+NP and V+PP combinations. All four approaches were token-based.

Approaches to metaphoric language predominantly focus on binary classification. The most prominent research has been carried out by Shutova, best summarized in Shutova et al. (2013). Shutova performed both metaphor identification and interpretation, focusing on English verbs. She relied on a seed set of annotated metaphors and standard verb and noun clustering, to classify literal vs. metaphorical verb senses. Gedigian et al. (2006) also predicted metaphorical meanings of English verb tokens, relying on manual rather than unsupervised data, and a maximum entropy classifier. Turney et al. (2011) assume that metaphorical word usage is correlated with the abstractness of a word's context, and classified word senses in a given context as either literal or metaphorical. Their targets were adjective-noun combinations and verbs.

Approaches to metonymic language represent a considerable development regarding features and classification approaches since 2002: Markert and Hahn (2002) proposed a rule-based ranking system exploring the contribution of selectional preferences vs. discourse and anaphoric information; Markert and Nissim (2002) presented the first supervised classifier for location names and compared window co-occurrences, collocations and grammatical features; Nissim and Markert (2005) extended the framework towards organization names and focused on grammatical features; Nastase and Strube (2009) enriched the feature set from Markert and Nastase by lexical features from the Sketch Engine (Kilgarriff et al., 2004) fed into WordNet supersenses, and by encyclopedic knowledge from Wikipedia relations; Nastase et al. (2012) used an unsupervised classifier and global context relying on the Wikipedia concept network.

3 Dataset of German Perception Verbs

In this section, we describe the creation of our dataset of German perception verbs in three steps: (i) sampling the perception verbs (Section 3.1), (ii) identification of literal and non-prototypical meanings (Section 3.2), and (iii) corpus annotation with perception senses (Section 3.3).

3.1 Sampling of Perception Verbs

As there is no available resource providing a complete list of German perception verbs, we combined the information of several dictionaries and thesauri to create such a dataset. As a starting point, we defined a base verb for each type of perception: *sehen* 'see' for optical verbs, *hören* 'hear' for acoustic verbs, *riechen* 'smell' for olfactory verbs, *tasten* 'touch' for haptic verbs and *schmecken* 'taste' for gustatory verbs. Using these verbs as starting points, all their synonyms or closely related words were determined, relying on Ballmer and Brennenstuhl (1986) and Schumacher (1986). Using the enlarged set of verbs, we again added all their synonyms and closely related words. We repeated this cycle and at the same time made sure that each additional verb belongs exclusively to the desired perception class, until no further changes occurred. The sampling process determined 54 optical, 15 acoustic, 9 olfactory, 12 haptic and one gustatory verbs.

For the classification experiments, we selected one verb from each perception class, disregarding the sole gustatory verb. The selected olfactory and haptic verbs only undergo passive perception meanings, the optical verb only undergoes active perception meanings, and the acoustic verb holds both active and passive perception meanings.[4]

---

[4] Active perception is controlled perception (as in 'listens to the music'); passive perception is non-controlled perception (as in 'hears faint barking').

3.2 Non-Prototypical Meanings

Analyzing the senses of the perception verbs in our dataset was carried out in accordance with *Polysemy and Metaphor in Perception Verbs: A Cross-Linguistic Study* (Ibarretxe-Antunano, 1999), which systematically determined non-prototypical meanings of perception verbs cross-linguistically for English, Spanish and Basque. For example, Ibarretxe-Antunano (1999) identified three major groups of shifted meanings for vision verbs, (i) the *Intellection group* including *to understand, to foresee, to visualize, to consider, to revise*; (ii) the *Social group* including *to meet, to visit, to receive, to go out with, to get on badly*; (iii) the *Assurance group* including *to ascertain, to make sure, to take care*. We applied her cross-lingual meaning shifts to all German perception verbs in our dataset, if possible, to identify the meanings of the perception verbs. As in Ibarretxe-Antunano (1999), the applicability was determined by corpus evidence (see below).

The following lists specify the main senses of the perception verbs that were selected for the classification experiments, with the first category in each list referring to the literal meaning.

Optical: *'betrachten'*

- to look at (lit.)
- to define/name/interpret
- to analyze objectively
- to analyze subjectively

Acoustic: *'hören'*

- to hear (lit.)
- to (dis-)like/ignore
- to obey
- to be informed

Olfactory: *'wittern'*

- to sense (by smell, lit.)
- to advance towards a goal/event
- to predict

Haptic: *'spüren'*

- to feel (lit.)
- to realize
- to feel (emotions)
- to suspect

Taking the acoustic verb 'hören' as an example, we illustrate the corpus uses of the verb by one sentence for each sense.

- to hear (lit.):
  'Er *hörte* die Wölfe heulen.'
  He heard (lit.) the wolves howl.
- to (dis-)like/ignore:
  'Sie können es nicht mehr *hören*.'
  They don't want to hear about it anymore.
- to obey:
  'Wenn er nicht *hört*, gibt's kein Futter.'
  If he doesn't obey/listen, he doesn't get food.
- to be informed:
  'Davon habe ich noch nie *gehört*.'
  I never heard/read/etc. about that.

3.3 Annotation of Verb Senses

Based on the sense definitions, we performed a manual annotation to create a gold standard for our classification experiments: A random selection of 200 sentences for each of the four selected perception verbs was carried out, gathering 50 sentences for each meaning. As an exception, 'wittern' (olfactory) only has three prominent meanings, resulting in 150 annotated sentences. The random selection was based on a subcategorization database (Scheible et al., 2013) extracted from a parsed version (Bohnet, 2010) of the *SdeWaC* corpus (Faaß and Eckart, 2013), a web corpus containing 880 million words.

These randomly selected sentences were annotated by two native speakers of German with a linguistic background (doctoral candidates in computational linguistics). The annotators were asked to label each sentence with one of the specified meanings of the respective verb. In cases where the annotators disagreed, the first author of this paper took the final decision. Agreement and majority class baselines are shown in Table 1.

Verb	Perception	Baseline	Agreement
betrachten	optical	33.5%	63.0%
hören	acoustic	35.5%	64.5%
spüren	haptic	27.5%	75.0%
wittern	olfactory	39.0%	69.4%

Table 1: Baseline and inter-annotator agreement.

4 Syntax-Semantic Verb Features

The feature vector used to classify verb instances is split into three subsets of features: syntactic, verb-modifying and semantic features. The subsets are described in the following subsections.

4.1 Syntactic and Verb-Modifying Features

The syntactic and the verb-modifying features rely on the subcategorization database by Scheible et al. (2013). This resource is a compact but linguistically detailed database for German verb subcategorization, containing verbs extracted from the SdeWaC along with the following information:

(1) *verb information*: dependency relation of the target verb according to the TIGER annotation scheme (Brants et al., 2004; Seeker and Kuhn, 2012); verb position in the sentence; part-of-speech tag and lemma of the verb;

(2) *subcategorization information*: list of all verb complements;

(3) *applied linguistic rule* that was used to extract the verb and subcategorization information from the dependency parses;

(4) *whole sentence*.

Based on the database information, we defined the following features:

Syntactic features:

- *Sentence Rule*: Rule to extract the verb and subcategorization information; relies on the verb form and the dependency constellation of the verb.

- *Sentence Form*: Dependency relations of the verb complex according to TIGER.

- *Adjective*: Presence of an adjective represented by a Boolean value.

- *Accusative Object*: Presence of an accusative object represented by a Boolean value.

- *Subjunction*: Either "none" or the lemma of the subjunction if available.

- *Modal Verb*: Either "none" or the lemma of the modal verb if available.

- *Negation*: Presence of a negation represented by a Boolean value.

Verb-modifying features:

- *Verb Form*: Part-of-speech tag.

- *Adverb*: Presence of an adverb represented by a Boolean value.

- *Adverbial or Prepositional Object*: A Boolean value for each preposition introducing a prepositional object.

4.2 Semantic Features

The semantic features rely on two different resources, GermaNet and German Polarity Clues.
(1) Information on hypernymy is extracted from *GermaNet*, which has been modelled along the lines of the Princeton WordNet for English (Miller et al., 1990; Fellbaum, 1998) and shares its general design principles (Hamp and Feldweg, 1997; Kunze and Wagner, 1999). Lexical units denoting the same concept are grouped into synonym sets ('synsets'), which are interlinked via conceptual-semantic relations (such as hypernymy) and lexical relations (such as antonymy).

GermaNet provides up to 20 hypernym levels. We used the most common concepts from the 3rd level (counted down from the unique top level):

- Texture
- Situation
- Quality
- Cognitive Object
- Common Object
- Pronouns (added to the original net)
- None available (added to the original net)

(2) Information on adverb and adjective sentiment is extracted from the *German Polarity Clues* (Waltinger, 2010), which labels adjectives and adverbs as "positive", "negative" or "neutral".
We extracted the following semantic features:

Semantic features:

- *Subject Hypernym*: Hypernym of the subject.

- *Object Hypernym*: Hypernym of the direct accusative object.

- *Adverb/Adjective Sentiment*: Either "none" if no adverbs or adjectives are available; or the adverb/adjective sentiment label.

5 Classification

Our classification experiments were performed with WEKA. The classifier algorithm used is *J48*, a Java reimplementation of the *C4.5* algorithm (Quinlan, 1993). For training and testing, ten-fold cross-validation was applied.

The classification experiments were done separately for each perception type, i.e. for each verb. Table 2 lists the classification results for the verb *hören*,[5] distinguishing between the subsets of syntactic, verb-modifying and semantic features as well as the results for the combined vector. *Instances* refers to the number of sentences for the respective meaning. *Fraction* shows the proportion of instances of one meaning in relation to all classified instances for the respective verb. *Classifier accuracy* shows the proportion of instances which have been correctly classified by our classifier; significance according to chi-square is marked, if applicable. *Annotator agreement* is the proportion of instances in which the two annotators chose the same meaning.

6 Discussion

In the following, we provide qualitative analyses and discussions regarding our classifications.

6.1 Features

For the optical perception verb *sehen* and the acoustic perception verb *hören*, the verb-modifying and the semantic subset of features, as well as the combined set of all features, significantly beat the baselines (33.5% and 35.5%, respectively). The two subsets are equally successful at classifying optical and acoustic verb instances, reaching between 52.5% and 55.5%.

For the haptic perception verb *spüren*, each of the subset vectors and the overall feature vector provide results significantly better than the baseline (27.5%). The best subset vector for this verb is the syntactic one with an accuracy of 43.0%.

The olfactory perception verb *wittern* is not classified significantly better than the baseline (39.0%) by any subset or the combined set. The

---

[5]The results for the three other verbs are omitted for space reasons. We nevertheless include them into our discussion below. See David (2013) for further results.

best subset vector for classification is the syntactic one with 43.9% accuracy.

The semantic subset vector turns out to be the overall best with an average of 47.2%. For all but the olfactory verb classification any subset of features returns higher accuracy than the baseline.

6.2 Ambiguity

The classification results and confusion matrices (see an example in Table 3) show that ambiguity is the biggest source of misclassifications. In the confusion matrices one can observe that often meaning "A" is misinterpreted as meaning "B", which is in turn often misinterpreted as meaning "A". Interestingly, meanings confused by the classification algorithm are very similar to those confused by human annotators.

	0	1	2
0: Prototypical	28	0	26
1: Adv. towards Goal	15	1	30
2: Predict	21	0	43

Table 3: Confusion matrix for olfactory/syntactic.

6.3 Lack of Detailed Semantic Data

The hypernym data covers a very high level of abstraction. This data distinguishes between, for example, *texture* and *objects*, but it does not distinguish between, for example, *animals* and *plants*, which might have been more desirable. High levels of abstraction had to be chosen for this research project as the lower levels of abstraction would have resulted in several hundred feature values and thus most probably have run into severe sparse data problems. Future work will nevertheless address an improved identification of semantic levels of abstraction.

6.4 Literal Meaning as Residual Class

The varying results by feature subsets for a verb's prototypical instances suggest to have a closer look at their classification. The correctly classified instances increase and decrease in proportion to the correctly classified instances of all other meanings. Looking into the decision trees which result in classification as "prototypical" instances, it turns out that the prototypical meaning shows residual class characteristics, cf. Table 4: It

	Instances	Fraction	Accuracy	Agreement
syntactic	200	100.0%	46.0%	68.5%
prototypical	71	35.5%	46.5%	59.2%
to (dis)like	11	5.5%	36.4%	81.8%
to obey	47	23.6%	70.2%	95.7%
to be informed	71	35.5%	31.0%	57.7%
verb-modifying	200	100.0%	***53.0%	68.5%
prototypical	71	35.5%	50.7%	59.2%
to (dis)like	11	5.5%	0.0%	81.8%
to obey	47	23.6%	51.1%	95.7%
to be informed	71	35.5%	64.8%	57.7%
semantic	200	100.0%	***55.5%	68.5%
prototypical	71	35.5%	40.8%	59.2%
to (dis)like	11	5.5%	0.0%	81.8%
to obey	47	23.6%	70.2%	95.7%
to be informed	71	35.5%	70.4%	57.7%
overall	200	100.0%	***57.0%	68.5%
prototypical	71	35.5%	39.4%	59.2%
to (dis)like	11	5.5%	18.2%	81.8%
to obey	47	23.6%	72.3%	95.7%
to be informed	71	35.5%	70.4%	57.7%

Table 2: Classification results for acoustic perception verb *hören* (baseline: 35.5%).

	optic	acoust	olfac	haptic	*avg*
baseline	19.0	35.5	32.9	25.0	28.1
annotation	84.2	59.2	92.4	92.0	82.0
syntactic	5.3	46.5	51.8	54.0	50.8
verb-mod	2.6	50.7	37.0	56.0	47.9
semantic	2.6	40.8	72.2	20.0	44.3
overall	42.1	39.4	44.4	46.0	43.0

Table 4: Prototypical meaning by subsets.

seems that only the inability to determine a non-prototypical meaning through the use of distinct features results in a classification as prototypical.

6.5 Choice of Non-literal Meanings

The classification results also depend on whether fine-grained or coarse-grained senses are used. A fine-grained sense definition would lead to less variation within a sense class but to a higher number of meanings. This in turn would require more manually annotated data to cover all meanings with enough corpus examples, therefore we decided to only use the reduced and coarse-grained sense selection. However, it is not clear where to draw the line, as there are cases where a verb can have two meanings at once in one context.

7 Conclusion

This paper presented a token-based automatic classification of German perception verbs into literal vs. multiple non-literal senses. Based on a corpus-based dataset of German perception verbs and their systematic meaning shifts, following Ibarretxe-Antunano (1999), we identified one verb of each of the four perception classes optical, acoustic, olfactory and haptic, and used Decision Trees relying on syntactic and semantic corpus-based features to classify the verb uses into 3 to 4 senses each. Our classifier reached accuracies between 45.5% and 69.4%, in comparison to baselines between 27.5% and 39.0%. In three out of four cases analysed our classifier's accuracy was significantly higher than the according baseline.

References

Eneko Agirre and Philip Edmonds, editors. 2006. *Word Sense Disambiguation: Algorithms and Applications*. Springer-Verlag. URL: http://www.wsdbook.org/.

Thomas T. Ballmer and Waltraud Brennenstuhl. 1986. *Deutsche Verben*. Gunter Narr Verlag, Tübingen.

Colin Bannard. 2007. A Measure of Syntactic Flexibility for Automatically Identifying Multiword Expressions in Corpora. In *Proceedings of the ACL*

Workshop on A Broader Perspective on Multiword Expressions, pages 1–8, Prague, Czech Republic.

Julia Birke and Anoop Sarkar. 2006. A Clustering Approach for the Nearly Unsupervised Recognition of Nonliteral Language. In *Proceedings of the 11th Conference of the European Chapter of the ACL*, pages 329–336, Trento, Italy.

Julia Birke and Anoop Sarkar. 2007. Active Learning for the Identification of Nonliteral Language. In *Proceedings of the Workshop on Computational Approaches to Figurative Language*, pages 21–28, Rochester, NY.

Bernd Bohnet. 2010. Top Accuracy and Fast Dependency Parsing is not a Contradiction. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 89–97, Beijing, China.

Sabine Brants, Stefanie Dipper, Peter Eisenberg, Silvia Hansen-Schirra, Esther König, Wolfgang Lezius, Christian Rohrer, George Smith, and Hans Uszkoreit. 2004. TIGER: Linguistic Interpretation of a German Corpus. *Research on Language and Computation*, 2(4):597–620.

Paul Cook and Suzanne Stevenson. 2006. Classifying Particle Semantics in English Verb-Particle Constructions. In *Proceedings of the ACL/COLING Workshop on Multiword Expressions: Identifying and Exploiting Underlying Properties*, pages 45–53, Sydney, Australia.

Benjamin David. 2013. Deutsche Wahrnehmungsverben: Bedeutungsverschiebungen und deren manuelle und automatische Klassifikation. Studienarbeit. Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.

Gertrud Faaß and Kerstin Eckart. 2013. SdeWaC – a Corpus of Parsable Sentences from the Web. In *Proceedings of the International Conference of the German Society for Computational Linguistics and Language Technology*, pages 61–68, Darmstadt, Germany.

Afsaneh Fazly, Suzanne Stevenson, and Ryan North. 2007. Automatically Learning Semantic Knowledge about Multiword Predicates. *Language Resources and Evaluation*, 41:61–89.

Afsaneh Fazly, Paul Cook, and Suzanne Stevenson. 2009. Unsupervised Type and Token Identification of Idiomatic Expressions. *Computational Linguistics*, 35(1):61–103.

Christiane Fellbaum, editor. 1998. *WordNet – An Electronic Lexical Database*. Language, Speech, and Communication. MIT Press, Cambridge, MA.

Matt Gedigian, John Bryant, Srini Narayanan, and Branimir Ciric. 2006. Catching Metaphors. In *Proceedings of the 3rd Workshop on Scalable Natural Language Understanding*, pages 41–48, New York City, NY.

Birgit Hamp and Helmut Feldweg. 1997. GermaNet – a Lexical-Semantic Net for German. In *Proceedings of the ACL Workshop on Automatic Information Extraction and Building Lexical Semantic Resources for NLP Applications*, pages 9–15, Madrid, Spain.

B. Iraide Ibarretxe-Antunano. 1999. *Polysemy and Metaphor in Perception Verbs*. Ph.D. thesis, University of Edinburgh.

Adam Kilgarriff, Pavel Rychlý, Pavel Smrz, and David Tugwell. 2004. The Sketch Engine. In *Proceedings of the 11th EURALEX International Congress*, pages 105–111, Lorient, France.

Claudia Kunze and Andreas Wagner. 1999. Integrating GermaNet into EuroWordNet, a Multilingual Lexical-Semantic Database. *Sprache und Datenverarbeitung*, 23(2):5–19.

Linlin Li and Caroline Sporleder. 2009. Classifier Combination for Contextual Idiom Detection Without Labelled Data. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 315–323, Singapore.

Katja Markert and Udo Hahn. 2002. Understanding Metonymies in Discourse. *Artificial Intelligence*, 136:145–198.

Katja Markert and Malvina Nissim. 2002. Metonymy Resolution as a Classification Task. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 204–213.

George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine J. Miller. 1990. Introduction to Wordnet: An On-line Lexical Database. *International Journal of Lexicography*, 3(4):235–244.

Vivi Nastase and Michael Strube. 2009. Combining Collocations, Lexical and Encyclopedic Knowledge for Metonymy Resolution. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 910–918, Singapore.

Vivi Nastase, Alex Judea, Katja Markert, and Michael Strube. 2012. Local and Global Context for Supervised and Unsupervised Metonymy Resolution. In *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 183–193, Jeju Island, Korea.

Roberto Navigli. 2009. Word Sense Disambiguation: A Survey. *ACM Computing Surveys*, 41(2).

Malvina Nissim and Katja Markert. 2005. Learning to buy a Renault and talk to BMW: A Supervised Approach to Conventional Metonymy. In *Proceedings of the 6th International Workshop on Computational Semantics*, pages 225–235, Tilburg, The Netherlands.

Ross Quinlan. 1993. *C4.5: Programs for Machine Learning*. Morgan Kaufmann.

Silke Scheible, Sabine Schulte im Walde, Marion Weller, and Max Kisselew. 2013. A Compact but Linguistically Detailed Database for German Verb Subcategorisation relying on Dependency Parses from a Web Corpus: Tool, Guidelines and Resource. In *Proceedings of the 8th Web as Corpus Workshop*, pages 63–72, Lancaster, UK.

Helmut Schumacher. 1986. *Verben in Feldern*. de Gruyter, Berlin.

Wolfgang Seeker and Jonas Kuhn. 2012. Making Ellipses Explicit in Dependency Conversion for a German Treebank. In *Proceedings of the 8th International Conference on Language Resources and Evaluation*, pages 3132–3139, Istanbul, Turkey.

Ekaterina Shutova, Simone Teufel, and Anna Korhonen. 2013. Statistical Metaphor Processing. *Computational Linguistics*, 39(2):301–353.

Caroline Sporleder and Linlin Li. 2009. Unsupervised Recognition of Literal and Non-Literal Use of Idiomatic Expressions. In *Proceedings of the 12th Conference of the European Chapter of the ACL*, pages 754–762, Athens, Greece.

Peter Turney, Yair Neuman, Dan Assaf, and Yohai Cohen. 2011. Literal and Metaphorical Sense Identification through Concrete and Abstract Context. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 680–690, Edinburgh, UK.

Ulli Waltinger. 2010. German Polarity Clues: A Lexical Resource for German Sentiment Analysis. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*, Valletta, Malta.