An approach to automatic figurative language detection: A pilot study

Anna Feldman and Jing Peng
{anna.feldman,jing.peng}@montclair.edu
Montclair State University
USA

Abstract

This pilot study explores a new approach to automatic detection of figurative language. Our working hypothesis is that the problem of automatic identification of idioms (and metaphors, to some extent) can be reduced to the problem of identifying an outlier in a dataset. By an outlier we mean an observation which appears to be inconsistent with the remainder of a set of data.

1 Introduction

The first problem that a researcher interested in figurative language might face is that of identifying and extracting the relevant data from the corpus. This seems a relatively easy task if we are interested in a particular grammatical construction; it is a much more complicated problem, however, if we are interested in a robust analysis of metaphoric and idiomatic expressions. Manual searching in a large corpus is laborious and time consuming; searching for source or domain vocabulary (Stefanow-itsch, 2006), in the case of metaphors, is non-trivial. Searching for metaphors based on "markers of metaphor" is not reliable (see e.g., Goatly (1997); Wallington et al. (2003)). Annotating large corpora with relevant information, such as metaphors, their source and target domains, idioms and their various types, is an extremely ambitious and costly task; though some researchers have explored this avenue (e.g., Steen (2001, 2002); Crisp et al. (2002); Semino (2005), among others). We view idioms (and metaphors, to some extent) as outliers in the data. In this pilot study, we only discuss the task of binary literal/figurative classification. Similarly to Birke and Sarkar (2006), for the purposes of this pilot study we take the simplified view that predicates in literal clauses do not violate accepted selection restrictions or our knowledge of the world. Nonliteral then includes metaphors, idioms, as well phrasal verbs and other anomalous expressions that cannot really be seen as literal. However, we do not make any strong claims about other types of figurative language, such as metonymy, for example.

Idioms Idioms are conventionalized expressions that have figurative meanings that cannot be derived from the literal meaning of the phrase. The prototypical examples of idioms are expressions like *I'll eat my hat*, *He put his foot in his mouth*, *Cut it out*, *I'm going to rake him over the coals*, *a blessing in disguise*, *a chip on your shoulder*, or *kick the bucket*. Researchers have not come up with a single agreed-upon definition of idioms that covers all members of this class (Glucksberg, 1993; Cacciari, 1993; Nunberg et al., 1994; Sag et al., 2002; Villavicencio et al., 2004; Fellbaum et al., 2006). The common property ascribed to the idiom is that it is an expression whose meaning is different from its simple compositional meaning. Some idioms become conventionalized or frozen in usage, and they resist change in syntactic structure. Others, however, do allow some variability in expression (Fellbaum, 2007; Fazly et al., 2009).

Previous studies focusing on the automatic identification of idiom types have often recognized the importance of drawing on their linguistic properties, such as their semantic idiosyncrasy or their restricted flexibility, pointed out above. Some researchers have relied on a manual encoding of idiom-specific knowledge in a lexicon (Copestake et al., 2002; Villavicencio et al., 2004; Odijk, 2004), whereas others have presented approaches for the automatic acquisition of more general (hence less distinctive) knowledge from corpora (McCarthy et al., 2003). Recent work that looks into the acquisition of the distinctive properties of idioms has been limited, both in scope and in the evaluation of the methods proposed (Lin, 1999; Evert et al., 2004).

All these approaches view idioms as multiword expressions (MWEs). All rely crucially on some preexisting lexicons or manually annotated data. All limit the search space by a particular type of linguistic construction (e.g., Verb+Noun combinations).

Metaphors Intuitively, metaphors are just implicit comparisons. In *That flat tire **cost** me an hour; You need to **budget** your time; Don't **spend** too (much/little) time on a slide; I **lost** a lot of time when I got sick*, TIME is compared to MONEY. Different theories have been proposed to explain the mechanisms and functions of metaphors. The conceptual theory of metaphor (Lakoff and Johnson, 1980), for example, deals with conventionalized metaphorical word senses by illustrating mappings between conceptual domains. Some metaphors are "dead" – they get unnoticed by the speakers and lost their original, metaphoric, interpretation, e.g., *fishing for compliments, the growth of the economy*. Some metaphors, in turn, are living (or novel). They are either 1) lexical extensions of conventional conceptual metaphors (e.g., THEORIES are CONSTRUCTED OBJECTS: *He is trying to **buttress** his argument with a lot of irrelevant facts, but it is still so **shaky** that it will easily **fall** apart under criticism*); or 2) analogies (e.g., *My job is a jail*); or 3) rich image metaphors (e.g., *We haven't time to give it more than a **catlick*** [BNC] 'perfunctory wash, similar to that done by cats licking their fur').[1] Often it is difficult to draw a clear

---

[1]Many examples are taken from the course on Figurative Language Processing taught by Birte Lönneker-Rodman at the 19th European Summer School in Logic, Language and Information, Trinity College, Dublin, Ire-

distinction between a dead metaphor and a living metaphor.

Previous work on automated metaphor detection includes (Fass, 1991; Martin, 1990; Dolan, 1995; Goatly, 1997; Mason, 2004; Gedigian et al., 2006; Birke and Sarkar, 2006) among others. Fass (1991), for instance, uses selection preference violation techniques to detect metaphors. However, they rely on hand-coded declarative knowledge bases. Markert and Nissim (2002) present a supervised classification algorithm for resolving metonymy. Metonymy is a closely related figure of speech to metaphors, where a word is substituted by another with which it is associated (e.g., *The **ham sandwich** is waiting for his check* (Lakoff and Johnson (1980)). They use collocation, cooccurrence and grammatical features in their classification algorithm.

Unlike our pilot study to classify unseen sentences as literal or figurative (metaphorical/idiomatic) without limiting ourselves to a particular linguistic construction or metaphor type, most work focuses on identifying metaphorical mappings between concepts in order to be able to interpret sentences that use these metaphors. Exceptions are Gedigian et al. (2006), who use labeled data (verbal targets associated with a set of frames) to train a maximum entropy (ME) classifier to distinguish literal sentences from metaphorical; Birke and Sarkar (2006), who describe a system for automatically classifying literal and nonliteral usages of verbs through nearly unsupervised word-sense disambiguation and clustering techniques; and Krishnakumaran and Zhu (2007),

who propose algorithms to automatically classify sentences into metaphoric or normal usages only relying on the WordNet and bigram counts and dealing with an extremely small subset of novel metaphoric usages that involves subject-object, verb-noun and adjective-noun relationship in sentences.

2 Our Approach

We see text as imperfect data that suffers from "corruption" (i.e., idioms and metaphors) that may affect interpretation and processing, decision making, etc. We hypothesize that in a coherent text, words will be semantically related and idiomatic expressions (and some metaphors) will come out as an outlier in the data. By an outlier we mean an observation which appears to be inconsistent with the remainder of a set of data.

Degand and Bestgen (2003) have identified three important properties of idioms. (1) A sequence with literal meaning has many neighbors, whereas a figurative one has few. (2) Idiomatic expressions should demonstrate low semantic proximity between the words composing them. (3) Idiomatic expressions should demonstrate low semantic proximity between the expression and the preceding and subsequent segments. We also think that many metaphors exhibit a similar behavior.

2.1 Anomaly/Outlier Detection

Measuring and detecting anomaly is challenging because of the insufficient knowledge or representation of the so-called "anomaly" for a given system (Ypma and Duin, 1997). Despite the technical challenge, a number of techniques has

been proposed and shown promising potential. There is a large body of work on anomaly detection in the literature. An extensive review of anomaly detection methods can be found in (Markou and Singh, 2003a,b).

Outlier detection algorithms have been used for a variety of NLP applications, such as text classification, unknown word sense detection, error detection, etc. (see, e.g., Guthrie et al. (2008); Erk (2006); Eskin (2000); Nakagawa and Matsumoto (2002)).

2.2 Figurative language detection based on Principal Component Analysis

The approach we are taking for figurative language detection is based on principal component analysis (PCA) (Jolliffe, 1986; Shyu et al., 2003). PCA has become established as one of the key tools for dimensionality reduction when dealing with real valued data. It has been used for language analysis quite extensively. Woods et al. (1986), for example, use PCA for language test scores. A group of subjects was scored on a battery of language tests, where the subtests measured different abilities such as vocabulary, grammar or reading comprehension. Horvath (1985) analyzes speech samples of Sydney speakers to determine the relative occurrence of five different variants of each of five vowels sounds. Using this data, the speakers clustered according to such factors as gender, age, ethnicity and socio-economic class.

PCA computes a set of mathematical features, called principal components, to explain the variance in the data. These principal components are linear combinations of the original variables describing the data and are orthogonal to each other. The first principal component corresponds to the direction along which the data vary the most. The second principal component corresponds to the direction along which the data vary the second most, and so on. Furthermore, total variance in all the principal components explains total variance in the data.

PCA has several advantages in outlier detection. First, it does not make any assumption regarding data distributions. Many statistical detection methods assume a Gaussian distribution of normal data, which is far from reality. Second, by using a few principal modes to describe data, PCA provides a compact representation of the data, resulting in increased computational efficiency and real time performance.

Let $\mathbf{z} = \{\mathbf{x}_i\}_{i=1}^m$ be a set of data points. Each $\mathbf{x}_i = (x_i^1, \cdots, x_i^q)^t$, where t denotes the transpose operator. That is, each data point is described by q attributes or variables. PCA computes a set of eigenvalue and eigenvector pairs $\{(\lambda_1, e_1, \cdots, (\lambda_q, e_q)\}$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_q$ by performing singular value decomposition of the covariance matrix of the data:

$$\Sigma = \sum_{i=1}^m (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^t, \qquad (1)$$

where $\bar{x} = 1/m \sum_{i=1}^m$. Then the ith principal component of an observation $\mathbf{x}$ is given by

$$y_i = e_i^t(\mathbf{x} - \bar{\mathbf{x}}). \qquad (2)$$

Note that the major components correspond strongly to the attributes having relatively large variance and covariance. Consequently, after projecting the data onto the principal component space, idioms and metaphors that are out-

liers with respect to the major components usually correspond to outliers on one or more of the original attributes. On the other hand, minor (last few) components represent a linear combination of the original attributes with minimal variance. Thus the minor components are sensitive to observations that are inconsistent with the variance structure of the data but are not considered to be outliers with respect to the original attributes (Jobson, 1992). Therefore, a large value along the minor components indicates strongly a potential outlier that otherwise may not be detected based solely on large values of the original attributes.

Our technique computes two functions for a given input $\mathbf{x}$. The first one is computed along major components:

$$f(x) = \sum_{i=1}^{p} \frac{y_i^2}{\lambda_i}. \tag{3}$$

The second one is computed along minor components:

$$g(x) = \sum_{i=q-r+1}^{q} \frac{y_i^2}{\lambda_i}. \tag{4}$$

Here y_i are projections along each component according to Eq. (2). p represents the number of major components and captures sufficient variance in the data, while r denotes the number of minor components. Both p and r can be determined through cross-validation.

It can be seen from our earlier discussion that $f(\mathbf{x})$ captures extreme observations with large values along some original attributes. On the other hand, $g(\mathbf{x})$ measures observations that are outside of the normal variance structure in the data. The strength of our approach is that it detects an outlier that is either extremely valued or

does not confirm to the same variance structure in the data.

Our technique then decides an input $\mathbf{x}$ as outlier if $f(\mathbf{x}) \geq T_f$ or $g(\mathbf{x}) \geq T_g$, where T_f and T_g are outlier thresholds that are associated with the false positive rate α (Kendall et al., 2009). Suppose that the data follow the normal distribution. Define

$$\alpha_f = \Pr\{\sum_{i=1}^{p} \frac{y_i^2}{\lambda_i} > T_f | \mathbf{x} \text{ is normal}\},$$

and

$$\alpha_g = \Pr\{\sum_{i=q-r+1}^{q} \frac{y_i^2}{\lambda_i} | \mathbf{x} \text{ is normal}\}.$$

Then

$$\alpha = \alpha_f + \alpha_g - \alpha_f \alpha_g. \tag{5}$$

The false positive rate has the following bound (Kendall et al., 2009)

$$\alpha_f + \alpha_g - \sqrt{\alpha_f \alpha_g} \leq \alpha \leq \alpha_f + \alpha_g.$$

Different types of outliers can be detected based on the values of α_f and α_g. If $\alpha_f = \alpha_g$ in Eq. (refalpha), α can be determined by solving a simple quadratic equation. For example, if we want a 2% false positive rate (i.e., $\alpha = 0.02$), we obtain $\alpha_f = \alpha_g = 0.0101$.

Note that the above calculation is based on the assumption that our data follow the normal distribution. This assumption, however, is unlikely to be true in practice. We, therefore, determine α_f and α_g values based on the empirical distributions of $\sum_{i=1}^{p} y_i^2/\lambda_i$ and $\sum_{i=q-r+1}^{q} y_i^2/\lambda_i$ in the training data. That is, for a false positive rate of 2%, T_f and T_g represent the 0.9899 quantile of the empirical distributions of $\sum_{i=1}^{p} y_i^2/\lambda_i$ and $\sum_{i=q-r+1}^{q} y_i^2/\lambda_i$, respectively.

3   Dataset

Our training set consists of 1,200 sentences (22,028 tokens) randomly extracted from the British National Corpus (BNC, `http://www.natcorp.ox.ac.uk/`). The first half of the data comes from the social science domain [2] and another half is defined in BNC as "imaginative" [3]. Our annotators were asked to identify clauses containing (any kind of) metaphors and idioms and paraphrase them literally. We used this paraphrased corpus for training. [4] The training data contains 139 paraphrased sentences.

Our test data are 99 sentences extracted from the BNC social science (non-fiction) section, annotated as either literal or figurative and additionally labeled with the information about the figures of speech they contain (idioms (I), dead metaphors (DM), and living metaphors (LM)). The annotator has identified 12 idioms, 22 dead metaphors, and 2 living metaphors in that text.

4   Experiments

We compare the proposed technique, Principal-minor Component Analysis (PmCA), against a random baseline approach. The baseline approach flips a far coin. If the outcome is head, it classifies a given sentence as outlier (idiom, dead metaphor or living metaphor). If the outcome is tail, it classifies a given sentence as a regular sentence. The outlier thresholds T_f and T_g at a given false positive rate are determined from the training data by setting $\alpha_f = \alpha_g$ in Eq. (5).

In this experiment, we treat each sentence as a document. We created a bag-of-words model for the data set, i.e., we use TF-IDF to represent the data. Single value decomposition is then applied to the bag of words and the number of principal modes for representing the latent semantics space is calculated that capture 100% variance in the data.

5   First Results

Table 5 shows the detection rates of the two competing methods at a given false positive rate. The results reported here were based on 10% of major components (p in Eq. (3)) and 0.1% minor components (r in Eq. (4)). It turns out that the technique is not sensitive to p values, while r represents a trade-off between detection and precision. That is, precision increases with increasing number of minor components, at the expense of lower detection. This is expected because when estimated over a large number of minor components, the averaging effect reduces sensitivity of observations to the variance structure of the data.

6   Discussion

This pilot study only describes a binary classification task: sentences are classified as either literal or figurative. Our long-term goal is to develop a method that is able to identify metaphors and idioms automatically. We understand that metaphor and idiom detection is

---

[2] This is an excerpt from *Cities and Plans: The Shaping of Urban Britain in the Nineteenth and Twentieth Centuries* by Gordon Emanuel Cherry.

[3] This is an excerpt taken from *Heathen*, a thriller novel written by Shaun Hutson.

[4] We understand that this task is highly subjective. Future experiments will measure the interannotator agreement on this task.

FP Rate	I+M	I	M	DM	LM	B
1	47	44	45	43	100	50
2	53	44	55	52	100	50
4	63	56	63	62	100	50
6	70	67	73	71	100	50
8	73	77	73	71	100	50
10	**87**	**89**	**86**	**86**	**100**	50

Table 1: Detection rates (%) of the PmCA method on idioms and metaphors (I+M), idioms (I), metaphors (M) (regardless of the type), dead metaphor (DM), living metaphor (LM); B is the baseline.

an extremely challenging task due to "issues including context sensitiveness, emergence of novel metaphoric forms, and the need for semantic knowledge about the sentences." (Krishnakumaran and Zhu, 2007) Moreover, Gibbs (1984) suggests (based on experimental evidence) that the distinctions between literal and metaphoric meanings have little psychological validity. He views literal and figurative expressions as end points of a single continuum. This makes the task of idioms and metaphors detection even more challenging because, perhaps, there is no objective clear boundary between metaphors/idioms and literal expressions.

Below we provide output sentences identified by our algorithm as either figurative or literal. Some cases are really fuzzy and clear literal/figurative demarcation is difficult. For example, the (false negative) sentence given in (4) below, marked by the annotator as metaphoric, was not identified by PmCA as such. When doing error analysis, we informally consulted several human subjects and asked them explicitly whether they thought the sentence contained a metaphor — surprisingly enough, most of them did not notice the (dead) metaphor and classified that sentence as literal.

1. True Positives (TP): Figurative sentences identified as figurative by PmCA

   - *This is an over-statement, but combined with an increasing preference by Government to curtail the content of structure plans, and recently by a Government less disposed to plan making by the public sector anyway, to withdraw structure plans all together, the **heady days** of 1960s optimism were lost.*
   - *As we have seen, the ideals of private suburbia were **deeply rooted**, particularly in England, and in design terms the **Unwinesque** tradition of vernacular cottage architecture and the predilection for low-density layouts had been articulated in the inter-war council estate.*
   - *Subsidies were resumed in 1961 and thereafter had **a chequered history** with changing problems in housing supply and differing political judgements (Burnett, 1978).*
   - *Planning policies tend **to run behind** developments and trends, and all too often **the planning machine** has given the impression of existing more for the benefit of those who run it (professionals and politicians) than those who are served by it.*

2. False Positives (FP): Literal sentences identified as figurative by PmCA

   - *Community disturbance was considerable, with very high annual transference*

rates of people from slum housing to alternative accommodation of very different style, quality, location and community setting.

3. True Negatives (TN): Literal sentences identified as literal by PmCA

   - *The manual was a complete guide to the building of six types of dwellings: the kitchen-living room house, the working kitchen house, the dining kitchen house, old people's dwellings, three-storey terrace houses and flats and maisonettes.*

4. False Negatives (FN): Figurative sentences identified as literal by PmCA

   - *The **attack on the slums** proved a very significant event in post-war urban planning.*

6.1 Related Work: Collocation Extraction

Words in natural language tend to occur in clusters. A group of words which frequently occur together is called a collocation. Some main criteria that characterize the notion of collocation are 1) (relative) non-compositionality; 2) non-substitutability; and 3) non-modifiability. Nevertheless, the term collocation is applied to a wide range of phenomena, including light verbs, verb particle constructions, proper names, terminological expressions, which do not necessarily satisfy all the three criteria. There has been extensive, serious work on collocation extraction (e.g., Berry-Rogghe (1973); Choueka et al. (1983); Church and Hanks (1989); Benson (1989); Church et al. (1991); Brundage et al. (1992); Smadja (1993); Dras and Johnson (1996); Daille (1996); Lin (1998); Evans and Zhai (1996); Shimohata et al. (1997); Goldman et al. (2001); Kilgariff and Tugwell (2001)) A detailed survey of various methods of collocation extraction is provided in Manning and Schütze (1999).

Naturally, idioms and metaphors (at least, partially) seem to fall under the collocation umbrella. Idioms, however, are the extreme case because their interpretation is mostly incomprehensible if previously unheard. In addition, collocations are word groups which frequently appear in the same context. This does not necessarily apply to metaphors, especially to novel ones – their context is much more difficult to predict.

In this work, we have not attempted to extract idioms and metaphors from the corpus. Our task was to classify sentences into literal and non-literal.

7 Conclusion

In this pilot study we did not want to restrict ourselves to a particular linguistic form of an idiom or a metaphor. We applied this method to English, but in principle, the technique is language-independent and can be applied to an arbitrarily selected language.

This binary classification approach has multiple applications. It is useful for indexing purposes in information retrieval (IR) as well as for increasing the precision of IR systems. Knowledge of which clauses should be interpreted literally and which figuratively will also improve text summarization and machine translation systems.

With regard to the future, we will focus on improving the detection rates and make the classification more fine-grained (metaphors vs. idioms, dead vs. living metaphors).

References

Benson, M. (1989). The Structure of the Collocational Dictionary. *International Journal of Lexicography 2*(1), 1–14.

Berry-Rogghe, G. (1973). *The Computer and Literary Studies*, Chapter The Computation of Collocations and their Relevant to Lexical Studies, pp. 103–112. Edinburgh: University Press.

Birke, J. and A. Sarkar (2006). A clustering approach to the nearly unsupervised recognition of nonliteral language. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL'06)*, Trento, Italy, pp. 329–226.

Brundage, J., M. Kresse, U. Schwall, and A. Storrer (1992). Multiword Lexemes: A Monoloingual and Constrastive Typology for Natural Language Processing and Machine Translation. Technical Report 232. Institut fuer Wissensbasierte Systeme, IBM Deutschland GmbH, Heidelberg.

Cacciari, C. (1993). The Place of Idioms in a Literal and Metaphorical World. In C. Cacciari and P. Tabossi (Eds.), *Idioms: Processing, Structure, and Interpretation*, pp. 27–53. Lawrence Erlbaum Associates.

Choueka, Y., S. T. Klein, and E. Neuwitz (1983). Automatic Retrieval of Frequent Idiomatic and Collocational Expressions in a Large Corpus. *Journal of the Association for Literary and Linguistic Computing 16*(1), 22–29.

Church, K. W., W. Gale, P. Hanks, and D. Hindle (1991). *Lexical Acquisition: Exploting On-line Resources to Build a Lexicon*, Chapter Using Statistics in Lexical Analysis, pp. 115–164. Hillsdale, NJ: Lawrence Erlbaum.

Church, K. W. and P. Hanks (1989). Word association norms, mutual information and lexicography. In *ACL 27*, pp. 76–83.

Copestake, A., F. Lambeau, A. Villavicencio, F. Bond, T. Baldwin, I. A. Sag, and D. Flickenger (2002). Multiword Expressions: Linguistic Precision and Reusability. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, Las Palmas, Canary Islands, pp. 1941–1947.

Crisp, P., J. Heywood, and G. J. Steen (2002). Metaphor Identification and Analysis, Classification and Quantification. *Language and Literature 11*, 55–69.

Daille, B. (1996). *The Balancing Act: Combining Symbolic and Statistical Approaches to Language*, Chapter Study and Implementation of Combined Techniques for Automatic Extraction, pp. 49–66. Cambridge, Masschusetts: MIT Press.

Degand, L. and Y. Bestgen (2003). Towards Automatic Retrieval of Idioms in French Newspaper Corpora. *Literary and Linguistic Computing 18*(3), 249–259.

Dolan, W. B. (1995). Metaphor as An Emrgent Property of Machine-readable Dictionaries. In *AAAI 1995 Spring Symposium*, Volume 95, pp. 27–32.

Dras, M. and M. Johnson (1996). Death and Lightness: Using a Demographic MOdel to Find Support Verbs. In *Proceedings of the 5th International Conference on the Cognitive Science of Natural Language Processing*.

Erk, K. (2006). Unknown word sense detection as outlier detection. In *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, pp. 128–135. Association for Computational Linguistics.

Eskin, E. (2000). Detecting errors within a corpus using anomaly detection. In *Proceedings of the first conference on North American chapter of the Association for Computational Linguistics*, San Francisco, CA, USA, pp. 148–153. Morgan Kaufmann Publishers Inc.

Evans, D. A. and C. Zhai (1996). Noun-phrase Analysis in Unrestricted Text for Information Retrieval. In *ACL 34*, pp. 17–24.

Evert, S., U. Heid, and K. Spranger (2004). Identifying Morphosyntactic Preferences in Collocations. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC'04)*, Lisbon, Portugal, pp. 907–910.

Fass, D. (1991). Met∗: A Method for Discriminating Metonymy and Metaphor by Computer. *Computational Linguistics 17*(1), 49–90.

Fazly, A., P. Cook, and S. Stevenson (2009). Unsupervised Type and Token Identification of

Idiomatic Expressions. *Computational Linguistics 35*(1), 61–103.

Fellbaum, C. (2007). The Ontological Loneliness of Idioms. In A. Schalley and D. Zaefferer (Eds.), *Ontolinguistics*. Mouton de Gruyter.

Fellbaum, C., A. Geyken, A. Herold, F. Koerner, and G. Neumann (2006). Corpus-based Studies of German Idioms and Light Verbs. *International Journal of Lexicography 19*(4), 349–360.

Gedigian, M., J. Bryant, S. Narayanan, and B. Ciric (2006). Catching Metaphors. In *Proceedings of Scalable Natural Language Understanding Conference*, Boston.

Gibbs, R. W. (1984). Literal Meaning and Psychological Theory. *Cognitive Science 8*, 275–304.

Glucksberg, S. (1993). Idiom Meanings and Allusional Content. In C. Cacciari and P. Tabossi (Eds.), *Idioms: Processing, Structure, and Interpretation*, pp. 3–26. Lawrence Erlbaum Associates.

Goatly, A. (1997). *The Language of Metaphors*. London: Routeledge.

Goldman, J. P., L. Nerima, and E. Wehrli (2001). Collocation Extraction Using a Syntacti Parser. In *Workshop Proceedings: COLLOCATION: Computational Extraction, Analysis and Exploitation*, Toulouse, France.

Guthrie, D., L. Guthrie, and Y. Wilks (2008). An Unsupervised Approach for the Detectin of Outliers in Corpora. In *Proceedings of the sixth international conference on Language Resources and Evaluation (LREC 2008)*, Marrakech, Morocco.

Horvath, B. M. (1985). *Variation in Australian English*. Cambridge: Cambridge University PRess.

Jobson, J. (1992). *Applied Multivariate Data Analysis, Volume II: Categorical and Multivariate Methods*. Springer Verlag.

Jolliffe, I. (1986). *Principal Component Analysis*. New York: Springer-Verlag.

Kendall, M., A. Stuart, and J. Ord (2009). *Kendall's Advanced Theory of Statistics: Volume 1:Distribution Theory*. John Wiley and Sons.

Kilgariff, A. and D. Tugwell (2001). *Machine Translation in the Information Age*, Chapter WASP-Bench: an MT Lexicographer's Workstation Supporting State-of-the-art Lexical Disambiguation. Geneva: European Association for Machine Translation.

Krishnakumaran, S. and X. Zhu (2007, April). Hunting Elusive Metaphors Using Lexical Resources. In *Proceedings of the Workshop on Computational Approaches to Figurative Language*, Rochester, New York, pp. 13–20. Association for Computational Linguistics.

Lakoff, G. and M. Johnson (1980). *Metaphors We Live By*. University of Chicago Press.

Lin, D. (1998). Extracting Collocations from Text Corpora. In *First Workshop on Computational Terminology*, Montreal, Canada.

Lin, D. (1999). Automatic Identification of Non-compositional Phrases. In *Proceedings of ACL*, College Park, Maryland, pp. 317–324.

Manning, C. D. and H. Schütze (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.

Markert, K. and M. Nissim (2002). Metonymy Resolution as a classification task. In *Proceedings of ACL-02 Conference on Empirical Methods in Natural Language Processing*, pp. 204–213.

Markou, M. and S. Singh (2003a). Novelty detection: A review part1: Statistical approaches. *Signal Processing*.

Markou, M. and S. Singh (2003b). Novelty detection: A review part2: Neural network-based approaches. *Signal Processing*.

Martin, J. (1990). *Computational Model of Metaphor Interpretation*. Academic Press.

Mason, Z. J. (2004). Cormet: A Computational, Corpus-based Conventional Metaphor Extraction System. *Computational Linguistics 30*(1), 23–44.

McCarthy, D., B. Keller, and J. Carroll (2003). Detecting a Continuum of Compositionality in Phrasal Verbs. In *Proceedings of the ACL-SIGLEX Workshop in Multiword Expressions: Analysis, Acquisition and Treatment*, Sapporo, Japan, pp. 73–80.

Nakagawa, T. and Y. Matsumoto (2002). Detecting errors in corpora using support vector machines. In *Proceedings of COLING 2002: The 19th International Conference on Computational Linguistics*.

Nunberg, G., I. A. Sag, and T. Wasow (1994). Idioms. *Language 70*(3), 491–538.

Odijk, J. (2004). A Proposed Standard for the Lexical Representations of Idioms. In *Proceedings of Euralex'04*, Lorient, France, pp. 153–164.

Sag, I. A., T. Baldwin, F. Bond, A. Copestake, and D. Flickinger (2002). Multiword expressions: A Pain in the Neck for NLP. In *Proceedings of the 3rd International Conference on Intelligence Text Processing and Computational Linguistics (CICLing 2002)*, Mexico City, Mexico, pp. 1–15.

Semino, E. (2005). The Metaphorical Construction of Complex Domains: the Case of Speech Activity in English. *Metaphor and Symbol*, 35–70.

Shimohata, S., T. Sugio, and J. Nagata (1997). Retrieving Collocations by Co-occurrences and Word Order Constraints. In *ACL 35/EACL 8*, pp. 476–481.

Shyu, M., S. Chen, K. Sarinnapakorn, and L. Chang (2003). A novel anomaly detection scheme based on principal component classifier. In *Proceedings of IEEE International Conference on Data Mining*.

Smadja, F. (1993). Retrieving Collocations from Text: Xtract. *Computational Linguistics 19*(1), 143–177.

Steen, G. J. (2001). A Reliable Procedure for Metaphor Identification. In J. Barnden, M. Lee, and K. Markert (Eds.), *Proceedings of the Workshop on Corpus-based and Processing Approaches to Figurative Language*, pp. 67–75. Lancaster: University Centre for Computer Corpus Research on Language.

Steen, G. J. (2002). Towards a Procedure for Metaphor Identification. *Language and Literature 7*, 243–269.

Stefanowitsch, A. (2006). *Corpus-based Approaches to Metaphor and Metonymy*, Chapter Words and Their Metaphors, pp. 63–105. Mouton de Gruyter.

Villavicencio, A., A. Copestake, B. Waldron, and F. Lambeau (2004). Lexical Encoding of MWEs. In *Proceedings of the Second ACL Workshop on Multiword Expressions: Integrating Processing*, Barcelona, Spain, pp. 80–87.

Wallington, A., J. A. Barnden, P. Buchlovsky, L. Fellows, and S. R. Glasbey (2003). Metaphor Annotation: A Systematic Study. Technical Report CSRP-03-4. The University of Birmingham, School of Computer Science.

Woods, A., P. Fletcher, and A. Hughes (1986). *Statistics in Language Studies*. Cambridge: Cambridge University Press.

Ypma, A. and R. P. Duin (1997). *Novelty Detection Using Selforganizing Maps*. London: Springer.