Exploring Sensorial Features for Metaphor Identification

Serra Sinem Tekiroğlu
University of Trento
Fondazione Bruno Kessler
Trento, Italy
tekiroglu@fbk.eu

Gözde Özbal
Fondazione Bruno Kessler
Trento, Italy
gozbalde@gmail.com

Carlo Strapparava
Fondazione Bruno Kessler
Trento, Italy
strappa@fbk.eu

Abstract

Language is the main communication device to represent the environment and share a common understanding of the world that we perceive through our sensory organs. Therefore, each language might contain a great amount of sensorial elements to express the perceptions both in literal and figurative usage. To tackle the semantics of figurative language, several conceptual properties such as concreteness or imegeability are utilized. However, there is no attempt in the literature to analyze and benefit from the sensorial elements for figurative language processing. In this paper, we investigate the impact of sensorial features on metaphor identification. We utilize an existing lexicon associating English words to sensorial modalities and propose a novel technique to automatically discover these associations from a dependency-parsed corpus. In our experiments, we measure the contribution of the sensorial features to the metaphor identification task with respect to a state of the art model. The results demonstrate that sensorial features yield better performance and show good generalization properties.

1 Introduction

Languages include many lexical items that are connected to sensory modalities in various semantic roles. For instance, while some words can be used to describe a perception activity (e.g., *to sniff*, *to watch*, *to feel*), others can simply be physical phenomena that can be perceived by sensory receptors (e.g., *light*, *song*, *salt*, *smoke*). Common usage of language, either figurative or literal, can be very dense in terms of sensorial words. As an example, the sentence "*I heard a harmonic melody.*" contains three sensorial words: *to hear* as a perception activity, *harmonic* as a perceived sensorial feature and *melody* as a perceivable phenomenon. The connection to the sense modalities of the words might not be mutually exclusive, that is to say a word can be associated with more than one sense. For instance, the adjective *sweet* could be associated with both *taste* and *smell*.

The description of one kind of sense impression by using words that normally describe another is commonly referred to as *linguistic synaesthesia*[1]. As an example, we can consider the slogans "*The taste of a paradise*" where the sense of sight is combined with the sense of taste or "*Hear the big picture*" where *sight* and *hearing* are merged. Synaesthesia strengthens creative thinking and it is commonly exploited as an imagination boosting tool in advertisement slogans (Pricken, 2008).

Synaesthesia is also commonly used in metaphors. Synaesthesic metaphors use words from one type of sensory modality, such as *sight*, *hearing*, *smell*, *taste* and *touch*, to describe a concept from another modality. In conceptual metaphor theory, metaphor is defined as a systematic mapping between two domains; namely target (or tenor) and source (or vehicle) domains (Lakoff and Johnson, 1980). Such mappings are asymmetric and might not correlate all features from the source domain to the target domain. Systematic studies on synaesthetic metaphors propose that there is a

---

[1] http://ahdictionary.com/

certain directionality of sense modality mappings. (Ullman, 1957), in a very early study, presented this directionality as a linear hierarchy of lower and higher sense modalities. In this hieararchy, modalities are ordered from lower to higher as touch, taste, smell, sound and color. Ullman (1957) proposes that lower modalities tend to occur as the source domain, while higher modalities tend to occur as the target domain. For instance, in the synaesthetic metaphor "*soft light*", the target domain of *seeing* is associated with the source domain of *touching*, while the target domain of *hearing* is associated with the source domain of *tasting* in "*sweet music*". However, later studies (Williams, 1976; Shen, 1997) propose that the mapping in the synaesthetic metaphorical transfer is more complex among the sensory modalities. Williams (1976) constitutes a generalized mapping for the synaesthetic metaphorical transfer by means of the diachronic semantic change of sensorial adjectives. Having regard to the citation dates of adjective meanings from Oxford English Dictionary[2] and Middle English Dictionary[3], the regular transfer rules among the sensorial modalities are introduced.

Several techniques for metaphor identification have been explored, including selectional preference violations (Fass, 1991; Neuman et al., 2013) or verb and noun clustering (Shutova et al., 2010; Birke and Sarkar, 2006; Shutova and Sun, 2013), supervised classification (Gedigian et al., 2006; Mohler et al., 2013; Tsvetkov et al., 2014a). As well as the identification techniques, different cognitive properties such as imageability (Broadwell et al., 2013; Tsvetkov et al., 2014a) and concreteness of the metaphor constituents (Neuman et al., 2013; Turney et al., 2011; Tsvetkov et al., 2014a), or lexical semantic properties such as supersenses (Hovy et al., 2013; Tsvetkov et al., 2014a) have been exploited.

While detecting and interpreting metaphors, imageability and concreteness features are generally utilized to identify the metaphorical transfer from a more concrete to a less concrete or from a more imageable to a less imageable word. However, in synaesthetic metaphors, the imageability or concreteness levels of both tenor and vehicle (or tar-

get and source) words can be similar. For instance, according to the MRC Psycholinguistic Database (MRCPD) (Coltheart, 1981) the concreteness (C) and imageability (I) values for target *smell* and source *cold* in the sentence "*The statue has a cold smell.*" are *C:450, I:477* and *C:457, I:531* respectively. Likewise, in the noun phrase "*Sweet silence*" the values are very close to each other (*C:352, I:470* for *silence* and *C:463, I:493* for *sweet*). As demonstrated by these examples, while both imageability and concreteness are related to human senses, these features alone might not be sufficient to model synaesthetic metaphors.

In this paper, we fill in this gap by measuring the contribution of the sensorial features to the identification of metaphors in the form of adjective-noun pairs. We explicitly integrate features that represent the sensorial associations of words for metaphor identification. To achieve that, we both utilize an existing sensorial lexicon and propose to discover these associations from a dependency-parsed corpus. In addition, we exploit the synaesthetic directionality rules proposed by Williams (1976) to encode a degree to which an adjective-noun pair is consistent with the synaesthetic metaphorical transfer. Our experiments show that sensorial associations of words could be useful for the identification of metaphorical expressions.

The rest of the paper is organized as follows. We first review the relevant literature to this study in Section 2. Then in Section 3, we describe the word-sense association resources. In Section 4, we describe the features that we introduce and detail the experiments that we conducted. Finally, in Section 5, we draw our conclusions and outline possible future directions.

2 Related Work

Mohler et al. (2013) exploit a supervised classification approach to detect linguistic metaphors. In this work, they first produce a domain-specific semantic signature which can be found to be encoded in the semantic network (linked senses) of WordNet, Wikipedia[4] links and corpus collocation statistics. A set of binary classifiers are actuated to detect metaphoricity within a text by comparing its seman-

tic signature to the semantic signatures of a set of known metaphors.

Schulder and Hovy (2014) consider the term relevance as an indicator of being non-literal and propose that novel metaphorical words are less prone to occur in the typical vocabulary of a text. The performance of this approach is evaluated both as a standalone metaphor classifier and as a component of a classifier using lexical properties of the words such as part-of-speech roles. The authors state that term relevance could improve the random baselines for both tasks and it could especially be useful in case of a sparse dataset.

Rather than an anomaly in the language or a simple word sense disambiguation problem, a cognitive linguistic view considers metaphor as a method for transferring knowledge from a concrete domain to a more abstract domain (Lakoff and Johnson, 1980). Following this view, Turney et al. (2011) propose an algorithm to classify adjectives and verbs as metaphorical or literal based on their abstractness/concreteness levels in association with the nouns they collocate with. The authors describe words as concrete if they are things, events, and properties that can be perceivable by human senses.

Neuman et al. (2013) extend the abstractness/concreteness model of Turney et al. (2011) with a selectional preference approach in order to detect metaphors consisting of concrete concepts. They focus on three types of metaphors including i) a subject noun and an object noun associated by the verb *to be* (e.g., "God is a king"), ii) the metaphorical verb representing the act of a subject noun on an object noun (e.g., "The war absorbed his energy"), iii) metaphorical adjective-noun phrases (e.g., "sweet kid").

Beigman Klebanov et al. (2014) propose a supervised approach to predict the metaphoricity of all content words with any part-of-speech in a running text. The authors propose a model combining unigram, topic models, POS, and concreteness features. While unigram features contribute the most, concreteness features are found to be effective only for some of the sets.

Based on the hypothesis that on the conceptual level, metaphors are shared across languages, rather than being lexical or language specific, Tsvetkov et al. (2014a) propose a metaphor detection system with cross-lingual model transfer for English that exploits several conceptual semantic features; abstractness and imageability, semantic supersenses, vector space word representations. They focus on two types of metaphors with the subject-verb-object (SVO) and adjective-noun (AN) syntactic relations. As another contribution, they create new metaphor-annotated corpora for English and Russian. In addition, they support the initial hypothesis by showing that the model trained in English can detect metaphors in Spanish, Farsi and Russian by projecting the features from the English model into another language using a bilingual dictionary. To the best of our knowledge, this system is the current state of the art for metaphor detection in English and constitutes the baseline for our experiments.

3 Word-Sense Associations

Following the hypothesis of Broadwell et al. (2013) that "Metaphors are likely to use highly imageable words, and words that are generally more imageable than the surrounding context", we introduce a novel hypothesis that metaphors are likely to also use sensorial words. To extract the sensorial associations of words, we use the following two resources.

3.1 Sensicon

This resource (Tekiroglu et al., 2014) is a large sensorial lexicon that associates 22,684 English words with human senses. It is constructed by employing a two phased computational approach.

In the first phase, a bootstrapping strategy is performed to generate a relatively large set of sensory seed words from a small set of manually selected seed words. Following an annotation task to select the seed words from FrameNet (Baker et al., 1998), WordNet relations are exploited to expand the sensory seed synsets that are acquired by mapping the seed words to WordNet synsets. At each bootstrapping cycle, a five-class sensorial classifier model is constructed over the seed synsets defined by their WordNet glosses. The expansion continues until the prediction performance of the model steadily drops.

In the second phase, a corpus based method is utilized to estimate the association scores in the final lexicon. Each entry in the lexicon consists of a lemma and part-of-speech (POS) tag pair and their

associations to the five human senses (i.e. sight, hearing, taste, smell and touch) measured in terms of normalized pointwise mutual information (NPMI). Each sensorial association provided by the lexicon is a float value in the range of -1 and 1.

Due to the way it is constructed, Sensicon might tend to give high association values for metaphorical sense associations of words as well as the literal ones. For instance, while adjective *dark* is related to *sight* as the literal sense association, Sensicon assigns very high association values to both *sight* and *taste*. While this tendency would be helpful as a hint for identifying synaesthetic words, metaphor identification task would need a complementary word-sense association resource that could highlight the literal sense association of a word.

3.2 Dependency-parsed corpus (DPC)

As an alternative to *Sensicon* for building word-sense associations, we extract this information from a corpus of dependency-parsed sentences. To achieve that, we follow a similar approach to Özbal et al. (2014) and use a database that stores, for each relation in the dependency treebank of LDC Giga-Word 5th Edition corpus[5]), its occurrences with specific "governors" (heads) and "dependents" (modifiers). To determine the sensorial load of a noun n, we first count how many times n occurs with the verb lemmas 'see', 'smell', 'hear', 'touch' and 'taste' in a direct object (*dobj*) syntactic relation in the database. Then, we divide each count by the number of times n appears in a direct object syntactic relation independently of the head that it is connected to. More specifically, the probability that n is associated to sense s is calculated as:

$$\mathrm{p}(s,n) = \frac{\mathrm{c}_{dobj}(v_s,n)}{\sum_{h_i} \mathrm{c}_{dobj}(h_i,n)} \qquad (1)$$

where $c_r(h,m)$ is the number of times that m depends on h in relation r (in this case, $r = dobj$) in the dependency database, v_s is the most representative verb for sense s (e.g., the verb 'hear' for the sense of hearing) and each h_i is a different governor of n in a *dobj* relation as observed in the database.

Our hypothesis is that nouns frequently acting as a direct object of a verb representing a human sense s are highly associated to s.

Similarly, to extract the sensorial load of an adjective a, we calculate the number of times a occurs with the verb lemmas 'look', 'smell', 'sound', 'feel' and 'taste' in an adjectival complement (*acomp*) syntactic relation in the database. Then, we divide each count by the number of times a appears in an *acomp* syntactic relation. More specifically, the probability that a is associated to sense s is calculated as:

$$\mathrm{p}(s,a) = \frac{\mathrm{c}_{acomp}(v_s,a)}{\sum_{h_i} \mathrm{c}_{acomp}(h_i,a)} \qquad (2)$$

The two resources capture different properties of words with respect to their sensorial load. While Sensicon yields indirect sensorial associations by modeling distributional properties of the lexicon, DPC attempts to directly model these associations independently of the context. For instance, while Sensicon associates the noun *plate* with *taste* as it frequently occurs in contexts involving eating, DPC assigns the highests scores to *sight* and *touch*.

4 Evaluation

In this section, we demonstrate the impact of sensorial associations of words on the classification of adjective-noun pairs as metaphorical or literal expressions.

4.1 Dataset

As an initial attempt to investigate the impact of sensorial associations of words in metaphor identification, we target metaphorical expressions which can easily be isolated from their context. In this study, we focus on adjective-noun (AN) pairs which could also well suit a common definition of the synaesthetic metaphors as adjective metaphors where an adjective associated to one sense modality describes a noun related to another modality (Utsumi and Sakamoto, 2007). To this end, we experiment with the AN dataset constructed by Tsvetkov et al. (2014a). The dataset consists of literal and metaphorical AN relations collected from public resources on the web and validated by human annotators. For instance, it includes *green energy*, *straight*

answer as metaphorical relations and *bloody nose*, *cool air* as literal relations. To be able to compare our model with the state-of-the-art, we use the same training and test split as Tsvetkov et al. (2014a). More precisely, 884 literal and 884 metaphorical AN pairs are used for training, while 100 literal and 100 metaphorical AN pairs are used for testing.

4.2 Classifier and Features

We perform a literal/metaphorical classification task by adding sensorial features on top of the features proposed by Tsvetkov et al. (2014a), which constitute our baseline: concreteness, imageability, supersenses and vector space word representations. As we discussed earlier, *imageability* (I) and *concreteness* (C) are highly effective in metaphor identification task. We obtain the I and C scores of each word from the resource constructed by Tsvetkov et al. (2014a) by projecting I and C values of words in MRCPD onto 150,114 English words. *Supersenses* are coarse semantic representations that could reflect the conceptual mappings between adjective and noun components of a relation. We attain noun supersenses from the lexicographer files of WordNet, such as *noun.phenomenon, noun.feeling, verb.perception*, and adjective supersenses from the resource generated by Tsvetkov et al. (2014b). As the last baseline feature, *Vector Space Word Representations* can be considered as lexical-semantic properties where each word is represented by a vector and semantically similar words have similar vectors. The detailed description of how the baseline features are extracted can be found in Tsvetkov et al. (2014a).

As the main focus of this study, we extract the sensorial features from Sensicon and a dependency-parsed corpus (DPC). For each adjective and noun in an AN relation, we add as features its five sense associations according to the two resources. This results in 10 features (S) coming from Sensicon and 10 features (D) coming from DPC. From S and D, we derive two more features (p_S and p_D respectively) computed as the Pearson correlation between the sense features for the noun and the adjective.

As the third type of sensorial feature, we add a feature (R) which encodes the degree to which the adjective noun pair is consistent with William's theory of sense modality directionality in synaes-
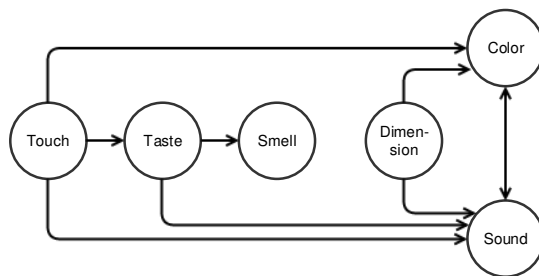

Figure 1: Directionality of sensory modalities as proposed by Williams(1976).

thetic metaphors (Williams, 1976). According to Williams, the mapping between the source and target sense of a synaesthetic adjective is more likely to flow in some directions and not in others, as exemplified in Figure 1. For example, while synaesthetic metaphors could be constructed with touch related adjectives and taste related nouns, the opposite direction, a taste related adjective and touch related noun, is less likely to occur. In our study, we employed simplified version of the directionality mapping in Figure 1 by identifying sight modality with *dimension* and *color*. For an AN relation, we first assign a sense to each component (i.e., adjective and noun) by choosing the highest sense association in DPC. We decided to employ DPC instead of Sensicon in the definition of this feature since by construction it provides a more direct association between words and senses. The value of R is set to 1.0 if the sense associations of the adjective and noun satisfies a direction in Figure 1. If the associations violate the directions in the figure, the value of the feature is set to 0.5. In all other cases it is set to 0.

Another sensorial feature set (W) is constructed by checking if the constituents of an AN pair appear in the Sensicon seed set, which consists of 4,287 sensorial words. For each adjective and noun, we add 5 binary features (one for each sense) and if the word is listed among the seeds for a specific sense, the feature for that sense is set to 1. In the same way, we construct another feature set (L) from the resource described in (Lynott and Connell, 2013; Lynott and Connell, 2013). This resource contains 1,000 nouns and object properties annotated with the five senses. Table 1 summarizes the features used in the classification task.

Feature Name	Abbreviation	# of the Features
Baseline	B	183
Baseline - VSM	B'	55
Sensicon	S	10
Sensicon Pearson	p_S	1
DPC	D	10
DPC Pearson	p_D	1
Sensicon Seeds	W	10
Lynott-Connell Sense Words	L	10
Directionality Rules	R	1
All sensorial features	A	43

Table 1: Feature sets used in the experiments.

To replicate the experimental setup of Tsvetkov et al. (2014a) as closely as possible, for our experiment we also use a Random Forest classifier, which was demonstrated to outperform other classification algorithms and to be robust to overfitting (Breiman, 2001). To fine tune the classifier and find the best Random Forest model for each feature set combination, we perform a grid search over the number of the generated trees (in the range between 50 and 300) and the maximum depth of the tree (in the range between 0 and 50) using 10-fold cross validation on AN training data. We choose the best model for each feature combination based on the maximum *average cross validation accuracy - standard deviation* value obtained by applying the given parameters.

4.3 Evaluation of the Baseline Features

The first row in Table 2 demonstrates the accuracy obtained with the complete set of baseline features. As it can be observed from the results, there is a significant drop of accuracy when moving from training to test data. We suspect that this performance loss might be due to the high dimensionality of the vector space feature set. Since according to Tsvetkov et al. (2014a) these features were designed mostly to deal with the multilinguality of their experimental setting, we evaluate the performance of the baseline excluding the vector space features. The row labeled B' reports the resulting accuracy values. The figures show that this simpler model has better generalization performance on monolingual English data. Hence, we decide to add our sensorial features on top of the simplified B' baseline.

Features	Cross-validation	Test
B	0.851	0.798
B'	0.831	0.845

Table 2: The cross validation and test accuracies of the baseline with and without vector space features.

4.4 Evaluation of the Sensorial Features

The second row labeled 'All' in Table 3 shows the cross validation and test accuracies of the sensorial features added on top of B'. The following rows show the outcome of the ablation experiments in which we remove each feature set at a time. The results that are marked with one or more $*$ indicate a statistically significant improvement in comparison to B' according to McNemar's test (McNemar, 1947). From the results it can be observed that the model including all sensorial features outperforms the baseline in both cross-validation and testing even though the difference on test data is not significant.

According to the ablation experiments, sensorial transaction rules (R) yield the highest contribution. While the Pearson correlation value calculated with Sensicon (p_S) results in an improvement, the feature representing the correlation with DPC (p_D) causes a decrease in the performance of the model. In general, all models using any tested subset of the sensorial features outperform the very competitive baseline even though the difference is significant only in two cases. To have more conclusive insights about the importance of each feature, an analysis on a larger dataset would be necessary. Overall, all the results demonstrate the useful contribution of the sensorial features to the task.

4.5 Error Analysis

The analysis that we performed on the test results shows that the noticeable performance differences among test results arise from the number of the instances in the test set. Indeed, a more comprehensive and bigger test set would provide better insights about the performance of sensorial features in the metaphor identification task.

---

[6]For two classifiers that have the same accuracy, McNemar test can yield different results with respect to the same baseline, depending on the tendency of each classifier to make the same errors as the baseline.

Features	Cross-validation	Test
B'	0.831	0.845
All	0.852**	0.875
All-S	0.850**	0.870
All-D	0.838	0.875
All-p_S	**0.855***	0.870
All-p_D	0.851**	**0.890***
All-R	0.838	0.865
All-L	0.853**	0.880
All-W	0.853**	0.880*[6]

Table 3: Performance of the B' baseline in combination with the different sets of sensorial features. Statistical significance: ***, $p < .001$; **, $p < .01$; *, $p < .05$.

Regarding the impact of the sensorial features, the test results indicate that sensorial association of the words could be beneficial in resolving the metaphors that include at least one sensorial component. For instance, the best configuration All-p_D could identify the *quiet revolution* as metaphorical while identifying *quiet voice* as literal with the sensorial adjective *quiet*.

A highly observable problem that causes error in the predictions is the limited coverage of the sensorial association resources. As an example, the literal AN pair *woolly mammoth* could not be resolved, since the adjective *woolly*, which is highly related to touch modality, can not be found in either Sensicon or DPC.

As another type of error, for less direct relations to sensory modalities, DPC might not provide the right information. For instance, in the literal AN relation *blind man*, the adjective *blind* is associated with taste as the highest sensory relation while associating *man* with sight modality. This might lead to the classification of this literal pair as metaphorical.

Considering the shortcomings of the current sensorial resources, a better sensorial lexicon differentiating various aspects of sensorial words such as direct sensorial properties (e.g., *coldness*, *odor* or *touch*), perceptibility of the concepts such as the visible concept (e.g., *cloud*), or tasteable concept (e.g., *food*), and also deeper cognitive relations of the words with senses such as *microphone* with hearing or *blind* with sight, could increase the perfor-

mance of the metaphor identification systems.

5 Conclusion

In this paper, we investigated the impact of sensorial features on the identification of metaphors in the form of adjective-noun pairs. We adopted a lexical approach for feature extraction in the same vein as the other cognitive features employed in metaphor identification, such as imageability and concreteness. To this end, we first utilized a state-of-the-art lexicon (i.e. *Sensicon*) associating English words to sensorial modalities. Then, we proposed a novel technique to automatically discover these associations from a dependency-parsed corpus. In our experiments, we evaluated the contribution of the sensorial features to the task when added to a state-of-the art model. Our results demonstrate that sensorial features are beneficial for the task and they generalize well as the accuracy improvements observed on the training data constantly reflect on test performance. To the best of our knowledge, this is the first model explicitly using sensorial features for metaphor detection. We believe that our results should encourage the community to explore further ways to encode sensorial information for the task and possibly to also use such features for other NLP tasks.

As future work, we would like to investigate the impact of sensorial features on the classification of other metaphor datasets such as VU Amsterdam Metaphor Corpus (Steen et al., 2010) and TroFi (Trope Finder) Example Base[7]. It would also be interesting to explore the contribution of these features for other figure of speech types such as similes. Furthermore, we plan to extend DPC approach with the automatic discovery of sensorial associations of verbs and adverbs in addition to adjectives and nouns. These efforts could result in the compilation of a new sensorial lexicon.

[7]Available at http://www.cs.sfu.ca/anoop/students/jbirke/

References

Collin F Baker, Charles J Fillmore, and John B Lowe. 1998. The berkeley framenet project. pages 86–90. Association for Computational Linguistics.

Beata Beigman Klebanov, Ben Leong, Michael Heilman, and Michael Flor. 2014. Different texts, same metaphors: Unigrams and beyond. In *Proceedings of the Second Workshop on Metaphor in NLP*, pages 11–17, Baltimore, MD, June. Association for Computational Linguistics.

Julia Birke and Anoop Sarkar. 2006. A clustering approach for nearly unsupervised recognition of nonliteral language. In *EACL*.

Leo Breiman. 2001. Random forests. *Machine learning*, 45(1):5–32.

George Aaron Broadwell, Umit Boz, Ignacio Cases, Tomek Strzalkowski, Laurie Feldman, Sarah Taylor, Samira Shaikh, Ting Liu, Kit Cho, and Nick Webb. 2013. Using imageability and topic chaining to locate metaphors in linguistic corpora. In *Social Computing, Behavioral-Cultural Modeling and Prediction*, pages 102–110. Springer.

Max Coltheart. 1981. The mrc psycholinguistic database. *The Quarterly Journal of Experimental Psychology*, 33(4):497–505.

Dan Fass. 1991. met*: A method for discriminating metonymy and metaphor by computer. *Computational Linguistics*, 17(1):49–90.

Matt Gedigian, John Bryant, Srini Narayanan, and Branimir Ciric. 2006. Catching metaphors. In *Proceedings of the Third Workshop on Scalable Natural Language Understanding*, pages 41–48. Association for Computational Linguistics.

Dirk Hovy, Shashank Srivastava, Sujay Kumar Jauhar, Mrinmaya Sachan, Kartik Goyal, Huiying Li, Whitney Sanders, and Eduard Hovy. 2013. Identifying metaphorical word use with tree kernels. *Meta4NLP 2013*, page 52.

George Lakoff and Mark Johnson. 1980. *Metaphors We Live By*. University of Chicago press.

Dermot Lynott and Louise Connell. 2013. Modality exclusivity norms for 400 nouns: The relationship between perceptual experience and surface word form. *Behavior research methods*, 45(2):516–526.

Quinn McNemar. 1947. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12(2):153–157, jun.

Michael Mohler, David Bracewell, David Hinote, and Marc Tomlinson. 2013. Semantic signatures for example-based linguistic metaphor detection. *Meta4NLP 2013*, page 27.

Yair Neuman, Dan Assaf, Yohai Cohen, Mark Last, Shlomo Argamon, Newton Howard, and Ophir Frieder. 2013. Metaphor identification in large texts corpora. *PloS one*, 8(4):e62343.

Gözde Özbal, Daniele Pighin, and Carlo Strapparava. 2014. Automation and evaluation of the keyword method for second language learning. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 352–357. Association for Computational Linguistics.

Mario Pricken. 2008. *Creative Advertising Ideas and Techniques from the World's Best Campaigns*. Thames & Hudson, 2nd edition.

Marc Schulder and Eduard Hovy. 2014. Metaphor detection through term relevance. *ACL 2014*, page 18.

Yeshayahu Shen. 1997. Cognitive constraints on poetic figures. *Cognitive Linguistics (includes Cognitive Linguistic Bibliography)*, 8(1):33–72.

Ekaterina Shutova and Lin Sun. 2013. Unsupervised metaphor identification using hierarchical graph factorization clustering. In *HLT-NAACL*, pages 978–988.

Ekaterina Shutova, Lin Sun, and Anna Korhonen. 2010. Metaphor identification using verb and noun clustering. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 1002–1010. Association for Computational Linguistics.

Gerard J Steen, Aletta G Dorst, J Berenike Herrmann, Anna A Kaal, and Tina Krennmayr. 2010. Metaphor in usage. *Cognitive Linguistics*, 21(4):765–796.

Serra Sinem Tekiroglu, Gözde Özbal, and Carlo Strapparava. 2014. Sensicon: An automatically constructed sensorial lexicon. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1511–1521, Doha, Qatar, October. Association for Computational Linguistics.

Yulia Tsvetkov, Leonid Boytsov, Anatole Gershman, Eric Nyberg, and Chris Dyer. 2014a. Metaphor detection with cross-lingual model transfer. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pages 248–258. Association for Computational Linguistics.

Yulia Tsvetkov, Nathan Schneider, Dirk Hovy, Archna Bhatia, Manaal Faruqui, and Chris Dyer. 2014b. Augmenting english adjective senses with supersenses. In *Proc. of LREC*.

Peter D Turney, Yair Neuman, Dan Assaf, and Yohai Cohen. 2011. Literal and metaphorical sense identification through concrete and abstract context. In *Proceedings of the 2011 Conference on the Empirical Methods in Natural Language Processing*, pages 680–690.

Shimon Ullman. 1957. Panchronistic tendencies in synaesthesia. *The principles of semantics*, pages 266–289.

Akira Utsumi and Maki Sakamoto. 2007. Computational evidence for two-stage categorization as a process of adjective metaphor comprehension. In *Proceedings of the Second European Cognitive Science Conference (EuroCogSci2007)*, pages 77–82.

Joseph M Williams. 1976. Synaesthetic adjectives: A possible law of semantic change. *Language*, pages 461–478.