

# Modality Enriched Neural Network for Metaphor Detection

Mingyu Wan<sup>1,2</sup> and Baixi Xing<sup>3</sup>

<sup>1</sup>School of Foreign Languages, Peking University

<sup>2</sup>Department of Chinese Bilingual Studies, The Hong Kong Polytechnic University

<sup>3</sup>School of Information Science Language, Beijing Language and Culture University  
wanmy@pku.edu.cn, xingbaixi@gmail.com

## Abstract

Metaphor as a cognitive mechanism in human’s conceptual system manifests itself an effective way for language communication. Although being intuitively sensible for human, metaphor detection is still a challenging task due to the subtle ontological differences between metaphorical and non-metaphorical expressions. This work proposes a modality enriched deep learning model for tackling this unsolved issue. It provides a new perspective for understanding metaphor as a modality shift, as in ‘*sweet voice*’. It also attempts to enhance metaphor detection by combining deep learning with effective linguistic insight. Extending the work at Wan et al. (2020), we concatenate word sensorimotor scores (Lynott et al., 2019) with word vectors as the input of attention-based Bi-LSTM using a benchmark dataset—the VUA corpus. The experimental results show great F1 improvement (above 0.5%) of the proposed model over other methods in record, demonstrating the usefulness of leveraging modality norms for metaphor detection.

## 1 Introduction

Metaphor is a common linguistic mechanism that encodes versatile ontological information accompanied by phenomena of *e.g.* domain transfer (Ahrens and Jiang, 2020), concreteness change (Maudslay et al., 2020) or semantic surprise (Zhang and Barnden, 2013) etc. In general, metaphor usually involves certain concept transfer from one domain (Source) to another (Target), as demonstrated in the example ‘*sweet smile*’ (using the taste modality to describe the vision modality). It is a fundamental way for people to relate our physical and familiar social experiences to a multitude of other subjects and contexts that are more complex, implicit or less known (Lakoff and Johnson, 1980; Lakoff and Johnson, 2008).

Detecting metaphors in texts is essential for capturing the authentic meaning of a language unit (*e.g.* word, phrases or even sentences) when the language is beyond literal interpretations, which can benefit a lot of language processing applications, such as machine translation, dialogue systems and sentiment analysis (Tsvetkov et al., 2014). Although detecting metaphors is subconsciously effortless for human, it is actually a great challenge to existing technologies due to the subtle semantic and ontological differences between metaphorical and non-metaphorical ones (Leong et al., 2018; Leong et al., 2020).

This work proposes a modality-enriched deep learning model for metaphor detection based on the idea that the basic senses (touch, hearing, smell, taste, vision and interception) and action affectors (mouth/throat, hand/arm, foot/leg, head) indicated by the sensorimotor norms (Lynott et al., 2019) of words provide crucial information for metaphoricity inference. As a next step of Wan et al. (2020) which only adopts simple statistical models, we build a more complex model combining neural networks with modality information. For standard reference, we adopt the benchmark dataset of the first and second shared tasks of metaphor detection—the VUA corpus (Klebanov et al., 2018)<sup>1</sup> for experiments.

---

This work is licensed under a Creative Commons Attribution 4.0 International Licence. Licence details: <http://creativecommons.org/licenses/by/4.0/>.

<sup>1</sup><http://www.vismet.org/metcor/documentation/home.html>

## 2 Related Work

Automatic detection of metaphors has been intensively studied in the recent decade among the Natural Language Processing community. Many approaches have been proposed with systems such as traditional machine learning, deep neural networks, transformer and sequential models etc. Commonly utilized features include n-grams, word2vecs, parts-of-speech, semantic classes, lexical concreteness, constructions and frames, among others (Bizzoni and Ghanimifard, 2018; Hong, 2016; Klebanov et al., 2014; Klebanov et al., 2015; Wilks et al., 2013).

The primary methods in metaphor detection, including the earliest attempts, mainly adopt supervised machine learning paradigm with training on pre-labelled datasets (Blanchard et al., 2014; Steen, 2010) based on various features mentioned above. This can be attested by the recent system papers for the 1<sup>st</sup> and 2<sup>nd</sup> shared tasks of Metaphor Detection (Leong et al., 2018; Leong et al., 2020). Although 90% of the systems adopt deep learning architectures (Dankers et al., 2019; Gao et al., 2018; Gutierrez et al., 2017), there are substantial linguist resources employed for system refinement, such as data with information of concreteness and imageability, semantic encoding of WordNet, FrameNet, VerbNet, SUMO ontology, property norms, syntactic dependency patterns, sensorial and vision-based information, etc. Such external resources provide different linguistic cues for metaphoricity inference, demonstrating their effectiveness for metaphor detection to various degrees (Alnafesah et al., 2020; Klebanov et al., 2016; Shutova et al., 2016).

To name a few advances, Brooks and Youssef (2020) build up an ensemble model of RNNs together with attention-based Bi-LSTMs. Chen et al. (2020) adopt BERT to obtain sentence embeddings, and then apply a linear layer with softmax to each token for metaphoricity predictions. Gong et al.(2020) use RoBERTa to obtain word embeddings and concatenate it with linguistic features (e.g. WordNet, VerbNet, POS, topicality, concreteness), and then feed them into a fully-connected Feedforward network to make predictions. Maudslay et al.(2020) combines the concreteness of a word with its static and contextual embeddings as inputs into a deep Multi-layer Perceptron network for predicting metaphoricity.

Despite the above advances, metaphor detection remains a difficult task because the semantic and ontological differences between metaphorical and non-metaphorical expressions are often subtle and their perception may vary from person to person. Existing methods show different strengths on detecting metaphors, yet each has its respective disadvantages, such as having generalization problems or lack association of their results with the intrinsic properties of metaphors. Besides, the performance of metaphor detection in record is still unsatisfactory (0.6-0.8 F1). In this work, we propose an innovative deep learning model for metaphor detection combining the strengths of neural networks with sensorimotor norms. Our work provides a new perspective for understanding metaphors as a linguistic mechanism accompanied by concept shifts indicated by the scores of sensorimotor norms. It also aims to testify the effectiveness of such scores encoded in words for the token-level metaphor detection task. Ultimately, we hope to improve the task of metaphor detection in a consistent way.

## 3 Data

### 3.1 The VUA Corpus

We adopt the VU Amsterdam Metaphor Corpus (VUA) (Tekiroğlu et al., 2015) for the experiments. This corpus is a benchmark dataset released for the shared tasks of metaphor detection (Leong et al., 2018; Leong et al., 2020), which is publicly available for standard reference. It is a subcorpus of the British National Corpus with manually annotated labels indicating the metaphoricity of each token in the corpus. It consists of 117 text fragments sampled across four genres: Academic, News, Conversation, and Fiction. The data has been annotated using the MIPVU procedure (Steen, 2010) with a strong inter-annotator agreement ( $k > 0.8$ ). Examples of the sentences in the corpus are demonstrated in Figure 1 below. Each token in the corpus is encoded with a unique id composed by the text fragment id, sentence id and the token sequence number in each sentence. Thus, the word ‘*corporate*’ in the first sentence of Figure 1 has the id of ‘ale-fragment01\_1\_2’. Each metaphorical expression is marked by the label of ‘M\_’ for distinction. Based on these gold labels, we are able to conduct supervised training and testing.

```

txt_id sentence_id sentence_txt
ale-fragment01 1 Latest corporate unbundler M_reveals laid-back M_approach
ale-fragment01 2 By FRANK KANE
ale-fragment01 3 IT SEEMS that Roland Franklin , the latest unbundler to ap
ale-fragment01 4 He has not properly investigated the M_target 's dining fa
ale-fragment01 5 The 63-year-old M_head of Pembridge Investments , M_throug
ale-fragment01 6 If he had M_taken his own rule seriously , he would have i
ale-fragment01 7 There are other M_things he has , M_on his own M_admissio
ale-fragment01 8 When the bid was M_launched last week , Mr Franklin M_face
ale-fragment01 9 He M_regards the M_charges as unfounded .
ale-fragment01 10 M_On property , he is M_blunt .

```

Figure 1: Sample of the annotated data in the VUA corpus

### 3.2 The Sensorimotor Norms

One key innovation of this study is that we propose to leverage modality information for metaphor detection as we are motivated by a common phenomenon existing in metaphorical expressions. That is, we observe many mismatched modalities among the metaphorical word and its neighbouring words. For instance, the word ‘head’ in the 5<sup>th</sup> line of Figure 1 has a dominant modality of ‘Vision’ and ‘Haptic’, while its following token ‘Investments’ has no obvious modality senses which is thought of weakly stimulated in terms of perception. The apparent modality difference between the two words implies a possible metaphorical usage, as proven by the annotated label in the corpus.

Thus, the Lancaster Sensorimotor norms collected by Lynott (2019) is used for enriching the deep learning model with modality scores. The data include measures of sensorimotor strength (0-5 scale indicating different degrees of sense modalities and action effectors) for around 40K English words across six perceptual modalities: touch, hearing, smell, taste, vision and interception, as well as five action effectors: mouth/throat, hand/arm, foot/leg, head (excluding mouth/throat), torso.<sup>2</sup>

To demonstrate the structure of the sensorimotor norms, we provide five example words and their six sensory scores<sup>3</sup> in Table 1 for illustration.

Word	Auditory	Gustatory	Haptic	Visual	Olfactory	Interoceptive	...
Adopt	1.222	0.056	1.056	<b>1.889</b>	0.111	1.222	...
Big	0.944	0.167	2.722	<b>3.889</b>	0.111	0.333	...
Daze	0.455	0.000	0.000	1.953	0.000	<b>3.253</b>	...
Eat	1.263	<b>4.526</b>	2.158	2.632	2.421	2.474	...
Learn	<b>3.941</b>	0.765	1.765	3.882	0.588	1.529	...

Table 1: Examples of the modality norms (six main senses)

The modality with the highest scores (highlighted) among the six senses of the words marks the dominant sense modality for each word, such as ‘**Visual**’ for the word ‘*Big*’ and ‘**Gustatory**’ for the word ‘*Eat*’. The six basic senses as well as motor affectors are directly linked to human experience reflecting the two basic infrastructures in human’s perceptual and conceptual system and can be used as very useful resource for linguistic modelling (Chersoni et al., 2020; WAN et al., 2020).

## 4 The Modality Enriched Model

In the model, words are processed with the integration of modality scores and word embeddings, as demonstrated in Figure 2 below. We map the words to the sensorimotor norms and obtain the modality representations (64 dimension for each word). At the same time we get the word vectors (300 dimension for each word) using GloVe (Pennington et al., 2014) and then concatenate them as inputs to neural networks. For those words not mapped in the sensorimotor norms, we assigned three kinds of values, including all zeros, random values following normal distribution, as well as average scores of the sensory

<sup>2</sup><https://osf.io/7emr6/>

<sup>3</sup>There are 64 dimension of scales for each word in the data, including measures of senses and actions and other subcategories. We demonstrate the six major dimension for concept explanation.

words in the corpus. In the end, we chose to use the average score for each dimension of the out-of-dictionary words due to optimization reasons.

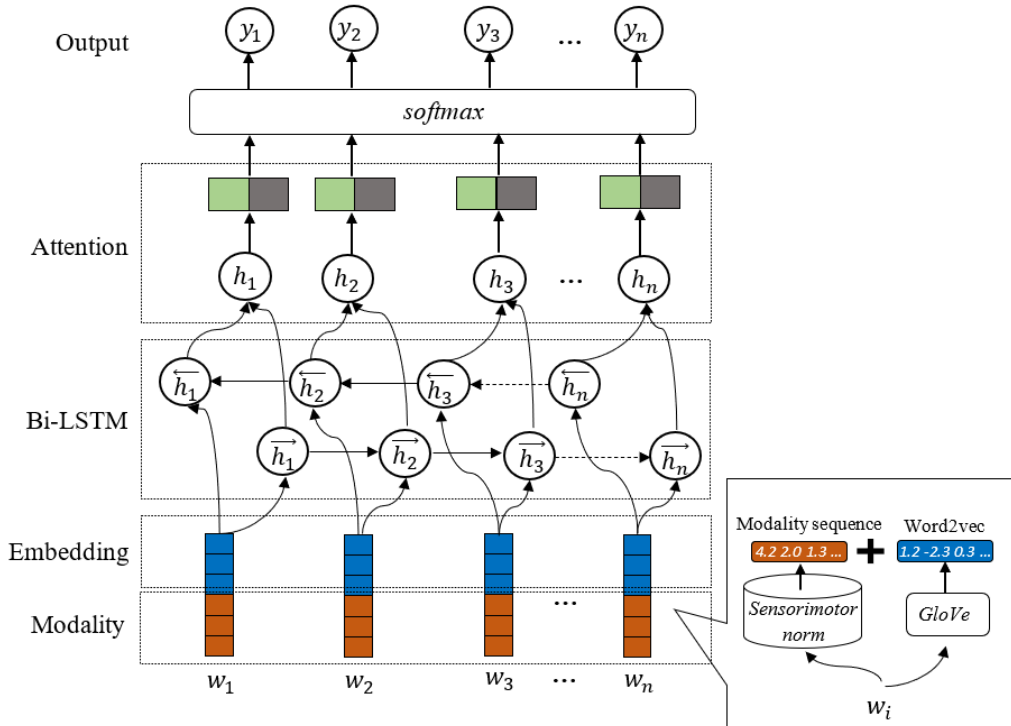


Figure 2: The Modality Enriched Model

Bi-LSTM layer produces a hidden status of each word in a sentence. We use these status to calculate an attention weight which will be multiplied with the output of Bi-LSTM layer. Let  $\mathbf{H} \in \mathbb{R}^{d \times N}$  be a matrix consisting of hidden vectors  $[\vec{h}_1, \vec{h}_2, \dots, \vec{h}_N]$  that is produced by LSTM, where  $d$  is the size of hidden layers and  $N$  is the length of the given sentence. The attention mechanism will generate an attention weight  $\alpha$ . The final sentence representation is given by:

$$\mathbf{h} = \mathbf{H} \times \alpha^T$$

We also add an additional Linear layer. The final probability distribution is:

$$y = \text{softmax}(\mathbf{W}_s \mathbf{h} + b_s)$$

Let  $y$  be the target distribution for sentence,  $\hat{y}$  be the predicted metaphoricity distribution. Train to minimize the cross-entropy error between  $y$  and  $\hat{y}$  for all sentences.

$$\text{loss} = - \sum_i \sum_j y_i^j \log \hat{y}_i^j + \lambda \|\theta\|^2$$

Finally, we get a probability distribution of 0-1 label to train the model and as the prediction result.

## 5 Results and Discussion

### 5.1 Evaluation Results

In order to evaluate the effectiveness of the proposed model for metaphor detection, we randomly select a development set (4,380 tokens) from the training set (17,240 tokens) in proportion to the Train/Test ratio of the task in Leong et al. (2020). The evaluation results are summarized in Table 2 below:

Category	Approach	P	R	F1
Baseline	uni-gram + logistic regression	0.521	0.657	0.581
Linguistic	modality + linear	0.605	0.556	0.579
	modality + LSTM	0.699	0.675	0.687
Neural	Glove + LSTM	0.744	0.748	0.746
Enriched	modality + Glove + LSTM	<b>0.767</b>	<b>0.755</b>	<b>0.761</b>

Table 2: Evaluation Results of the System

In Table 2, the baseline of using unigram as features and logistic regression (LR) as the classifier is implemented for a basic comparison. It is a commonly adopted baseline in the tasks of metaphor detection. We also implement several sub-categories of approaches before trying the enriched model, including the linguistic and neural networks in separate and also in combination. The results show an 18% F1 improvement of the enriched model over the baseline, a 7% F1 improvement over pure linguistic model, a 1.5% F1 improvement over the pure neural network model, and this superiority is salient and consistent in terms of both P (Precision) and R (Recall).

## 5.2 Peer Comparisons

To further demonstrate the effectiveness of our method, this following table presents the comparisons of our system to some highly related recent works on the same task. All the results are publicly available, as reported in Leong et al. (2020). The detailed results are displayed in Table 3 below:

Work	Method	F1
Wan et al. (2020)	modality + other features + LR	0.652
Kuo and Carpuat (2020)	Bi-LSTM+Embeddings+Unigram Lemmas+Spell Correction	0.686
Kumar & Sharma (2020)	Character embeddings+Similarity Networks+Bi-LSTM+Transformer	0.717
Liu et al. (2020)	BERT, XNET + POS tags + Bi-LSTM	0.730
Li et al. (2020)	ALBERT + Bi-LSTM	0.755
Top base: Devlin et al. (2018)	BERT: Pre-training of deep bidirectional transformers	0.756
<b>The current study</b>	<b>modality + Glove + LSTM</b>	<b>0.761</b>
Top 1: Su et al. (2020)	Global and local text information+Transformer stacks	0.804

Table 3: Comparison of Results of Our System to State-of-the-art Works

Our method obtains very promising results: it outperforms 6/7 highly related works to a great extent (0.5%-11% F1 gain), also approaching a reachable performance (a 4% F1 discrepancy) to the Top 1 work in record (Su et al., 2020). Moreover, our results are consistently superior to the top baseline and other linguistically-based or deep learning approaches. This suggests the effectiveness of leveraging modality norms in neural networks for metaphor detection, echoing the hypothesis in Wan et al. (2020) that metaphor manifests a concept mismatch (modality shift in particular) between source and target.

## 6 Conclusion and Future Work

We presented a linguistically enhanced method for metaphor detection using modality features plus word embeddings on the basis of attention-based neural network, as a further evidence of Wan et al. (2020)’s first proposal of utilizing conceptual norms for metaphor detection. The proposed model with access to modality scores achieves outperforming results that 1) are significantly higher than one without such information, and 2) show leading performances than most related works in record. The results of the study are in line with previous work, and strengthen the conclusion that conceptual modality is crucial to metaphors (relevant to both our understanding of metaphors and NLP). In future, we will adopt more advanced deep learning architecture and introduce modality using other paradigms rather than concatenating embedding with the hope to further enrich the content.

## Acknowledgements

We owe gratitude to the anonymous reviewers of COLING'2020 for their constructive comments and suggestions for improving this work. Special thanks to Prof. Ahrens, Kathleen who recommended the 2<sup>nd</sup> Shared Task of Metaphor Detection to the authors and provided valuable advice for the initial work. Finally, this work is supported by Beijing Natural Science Foundation (4192057) and a Post-doctoral project (no. 4-ZZKE) at the Hong Kong Polytechnic University.

## References

- Kathleen Ahrens and Menghan Jiang. 2020. Source domain verification using corpus-based tools. *Metaphor and Symbol*, 35(1):43–55.
- Ghadi Alnafesah, Harish Tayyar Madabushi, and Mark Lee. 2020. Augmenting neural metaphor detection with concreteness. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 204–210.
- Yuri Bizzoni and Mehdi Ghanimifard. 2018. Bigrams and bilstms two neural networks for sequential metaphor detection. In *Proceedings of the Workshop on Figurative Language Processing*, pages 91–101.
- D Blanchard, J Tetreault, D Higgins, A Cahill, and M Chodorow. 2014. Ets corpus of non-native written english ldc2014t06. *Philadelphia: Linguistic Data Consortium*.
- Jennifer Brooks and Abdou Youssef. 2020. Metaphor detection using ensembles of bidirectional recurrent neural networks. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 244–249.
- Xianyang Chen, Chee Wee Leong, Michael Flor, and Beata Beigman Klebanov. 2020. Go figure! multi-task transformer-based architecture for metaphor detection using idioms: Ets team in 2020 metaphor shared task. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 235–243.
- R. Chersoni, E. and Xiang, Q. Lu, and C.R. Huang. 2020. Automatic learning of modality exclusivity norms via crosslingual word embeddings. In *Proceedings of the 9th Joint Conference on Lexical and Computational Semantics (\*SEM2020)*.
- Verna Dankers, Marek Rei, Martha Lewis, and Ekaterina Shutova. 2019. Modelling the interplay of metaphor and emotion through multitask learning. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2218–2229.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Ge Gao, Eunsol Choi, Yejin Choi, and Luke Zettlemoyer. 2018. Neural metaphor detection in context. *arXiv preprint arXiv:1808.09653*.
- Hongyu Gong, Kshitij Gupta, Akriti Jain, and Suma Bhat. 2020. Illinimet: Illinois system for metaphor detection with contextual and linguistic information. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 146–153.
- E Dario Gutierrez, Guillermo A Cecchi, Cheryl Corcoran, and Philip Corlett. 2017. Using automated metaphor identification to aid in detection and prediction of first-episode schizophrenia. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2923–2930.
- Jisup Hong. 2016. Automatic metaphor detection using constructions and frames. *Constructions and frames*, 8(2):295–322.
- 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.
- Beata Beigman Klebanov, Chee Wee Leong, and Michael Flor. 2015. Supervised word-level metaphor detection: Experiments with concreteness and reweighting of examples. In *Proceedings of the Third Workshop on Metaphor in NLP*, pages 11–20.
- Beata Beigman Klebanov, Chee Wee Leong, E Dario Gutierrez, Ekaterina Shutova, and Michael Flor. 2016. Semantic classifications for detection of verb metaphors. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 101–106.

- Beata Beigman Klebanov, Chee Wee Leong, and Michael Flor. 2018. A corpus of non-native written english annotated for metaphor. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 86–91.
- Tarun Kumar and Yashvardhan Sharma. 2020. Character aware models with similarity learning for metaphor detection. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 116–125.
- Kevin Kuo and Marine Carpuat. 2020. Evaluating a bi-lstm model for metaphor detection in toefl essays. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 192–196.
- George Lakoff and Mark Johnson. 1980. *Metaphors we live by*. Chicago, IL: University of Chicago.
- George Lakoff and Mark Johnson. 2008. *Metaphors we live by*. University of Chicago press.
- Chee Wee Leong, Beata Beigman Klebanov, and Ekaterina Shutova. 2018. A report on the 2018 via metaphor detection shared task. In *Proceedings of the Workshop on Figurative Language Processing*, pages 56–66.
- Chee Wee Leong, Beata Beigman Klebanov, Chris Hamill, Egon Stemle, Rutuja Ubale, and Xianyang Chen. 2020. A report on the 2020 via and toefl metaphor detection shared task. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 18–29.
- Shuqun Li, Jingjie Zeng, Jinhui Zhang, Tao Peng, Liang Yang, and Hongfei Lin. 2020. Albert-bilstm for sequential metaphor detection. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 110–115.
- Jerry Liu, Nathan O’Hara, Alexander Rubin, Rachel Draelos, and Cynthia Rudin. 2020. Metaphor detection using contextual word embeddings from transformers. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 250–255.
- Dermot Lynott, Louise Connell, Marc Brysbaert, James Brand, and James Carney. 2019. The lancaster sensorimotor norms: multidimensional measures of perceptual and action strength for 40,000 english words. *Behavior Research Methods*, pages 1–21.
- Rowan Hall Maudslay, Tiago Pimentel, Ryan Cotterell, and Simone Teufel. 2020. Metaphor detection using context and concreteness. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 221–226.
- Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.
- Ekaterina Shutova, Douwe Kiela, and Jean Maillard. 2016. Black holes and white rabbits: Metaphor identification with visual features. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 160–170.
- Gerard Steen. 2010. *A method for linguistic metaphor identification: From MIP to MIPVU*, volume 14. John Benjamins Publishing.
- Chuangdong Su, Fumiyo Fukumoto, Xiaoxi Huang, Jiyi Li, Rongbo Wang, and Zhiqun Chen. 2020. Deepmet: A reading comprehension paradigm for token-level metaphor detection. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 30–39.
- Serra Sinem Tekiroğlu, Gözde Özbal, and Carlo Strapparava. 2015. Exploring sensorial features for metaphor identification. In *Proceedings of the Third Workshop on Metaphor in NLP*, pages 31–39.
- Yulia Tsvetkov, Leonid Boytsov, Anatole Gershman, Eric Nyberg, and Chris Dyer. 2014. Metaphor detection with cross-lingual model transfer. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 248–258.
- Mingyu WAN, Kathleen Ahrens, Emmanuele Chersoni, Menghan Jiang, Qi Su, Rong Xiang, and Chu-Ren Huang. 2020. Using conceptual norms for metaphor detection. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 104–109, Online, July. Association for Computational Linguistics.
- Yorick Wilks, Adam Dalton, James Allen, and Lucian Galescu. 2013. Automatic metaphor detection using large-scale lexical resources and conventional metaphor extraction. In *Proceedings of the First Workshop on Metaphor in NLP*, pages 36–44.
- Li Zhang and John Barnden. 2013. Towards a semantic-based approach for affect and metaphor detection. *International Journal of Distance Education Technologies (IJDET)*, 11(2):48–65.