

Recent advances in neural metaphor processing: A linguistic, cognitive and social perspective

Xiaoyu Tong[♣], Ekaterina Shutova[♣], Martha Lewis^{◇♣}

[♣]ILLC, University of Amsterdam, the Netherlands

[◇]Department of Engineering Mathematics, University of Bristol, United Kingdom

xiaoyu.tong@student.uva.nl, e.shutova@uva.nl,
martha.lewis@bristol.ac.uk

Abstract

Metaphor is an indispensable part of human cognition and everyday communication. Much research has been conducted elucidating metaphor processing in the mind/brain and the role it plays in communication. In recent years, metaphor processing systems have benefited greatly from these studies, as well as the rapid advances in deep learning for natural language processing (NLP). This paper provides a comprehensive review and discussion of recent developments in automated metaphor processing, in light of the findings about metaphor in the mind, language, and communication, and from the perspective of downstream NLP tasks.

1 Introduction

Metaphor is a figurative device that allows us to portray one domain, often abstract, in terms of another that is typically more concrete (Lakoff and Johnson, 1980). In the sentences *His eyes lightened at the news*, or *Her mood darkened*, the abstract domain of emotion is described in terms of the more concrete domain of brightness. Through cross-domain associations and comparisons, metaphor enables us to communicate complex abstract concepts, to convey our affective states, to reinforce our arguments and, ultimately, to forge entirely novel word meanings. Metaphor is also ubiquitous in language, occurring on average in every three sentences, according to corpus-linguistic research (Steen et al., 2010b; Shutova, 2011). This makes accurate interpretation of metaphorical language essential for many NLP tasks and applications.

A systematic and comprehensive survey of metaphor processing systems was published five years ago (Shutova, 2015). Since then, automated metaphor processing has shifted focus towards neural approaches and there has been much activity in the area, including two shared tasks (Leong et al., 2018, 2020). The development has kept pace with the development of deep learning techniques, and

the state-of-the-art performance has been advanced very quickly, stressing the need for a new survey focusing on this recent progress.

Moreover, the design and evaluation of metaphor processing systems should be informed by both the cognition of metaphor and its role in communication, as well as the potential applications of these systems. It is common practice in the NLP community to divide metaphor processing into two sub-tasks, metaphor identification and interpretation, and the two shared tasks encourage the use of a common task definition and particular metaphor corpora for the development of metaphor identification systems. However, there has been little discussion about whether or how well the existing systems address theoretical and cognitive considerations in the use of metaphor and how insights and models produced by this work could inform and benefit the wider natural language understanding (NLU) research.

This paper aims to provide a starting point for such a discussion, by reviewing recent advances in metaphor processing in the context of contemporary theories of metaphor, its sociolinguistic aspects and the needs of NLP applications. We begin with an overview of theories and findings about metaphor in the mind/brain, language, and communication. We then review metaphor processing approaches published in the last 5 years, as well as the datasets they used. We provide an analysis of these models from the linguistic and cognitive perspective and propose some future directions for modelling social aspects of metaphor use (e.g. the intent behind metaphor use, extended metaphor and metaphoric framing), as well as the underaddressed tasks of metaphor interpretation and generation. Finally, we discuss how elements of metaphor processing can be integrated in the real-world applications, such as machine translation, opinion mining, dialogue modelling, and modelling argumentative discourse (e.g., political discourse).

2 Theoretical accounts of metaphor

2.1 Conceptual metaphors

According to the conceptual metaphor theory (CMT), linguistic metaphors have their roots in *conceptual metaphors*, cross-domain mappings in one's conceptual system (Lakoff and Johnson, 1980). Based on comparable properties and relations in *the target domain* and *the source domain*, conceptual metaphors invite one to conceptualise the former through the latter. For example, *He attacked every point in her argument* instantiates the conceptual metaphor ARGUMENT IS WAR, in which ARGUMENT is the target domain and WAR the source domain. It uses an event in the domain of WAR to describe an event about ARGUMENT.

The target and source domains of a metaphor usually involve abstract and concrete concepts respectively, which has to do with concept representation in the brain. According to the theory of embodied cognition (Barsalou, 1999), concepts are represented within sensorimotor circuitry: GRASPING, for instance, is represented in areas that control hand movements (Gallese and Lakoff, 2005). When using expressions such as *grasp a point*, the same areas are involved for processing the metaphor (Aziz-Zadeh and Damasio, 2008).

Conceptual metaphors that are entrenched in one's conceptual system are termed *conventional metaphors*; those that are not entrenched are termed *novel metaphors*. ARGUMENT IS WAR is a typical conventional metaphor; the expression *attack one's argument*, for example, has entered contemporary dictionaries of English and is unlikely to be considered novel by the native speakers. On the contrary, *Fear coiled around his heart* instantiates a novel metaphor, FEAR IS A SNAKE. The verb has a single sense in the Macmillan dictionary: a particular movement of 'something long and thin'. Fear is an abstract concept, incapable of any movement to be seen. The usage is therefore novel.

2.2 Linguistic metaphors

Informed by CMT, Steen et al. (2010b) proposed the Metaphor Identification Procedure Vrije Universiteit (MIPVU) for manually identifying metaphor-related words (MRWs) in a text. MRWs are word uses that can be attributed to underlying conceptual metaphors. What are usually called metaphorically used words are a subset of MRWs termed *indirect metaphors*. The source-domain word in *attack one's argument* is an indirect metaphor: it refers

to a target-domain act, and is therefore indirectly related to the underlying conceptual metaphor, ARGUMENT IS WAR.

Other types of MRWs are direct metaphors, implicit metaphors, and metaphor flags. *Direct metaphors* refer to source-domain entities or events. They often co-occur with *metaphor flags*, signals that a metaphor is used. In (1), the linking verb *is* signals a use of the BEAUTY AS FLOWER metaphor. The noun *flower* is a direct metaphor, referring directly to a source-domain concept.

- (1) Beauty is but a flower / which wrinkles will devour; [...] (Thomas Nashe)

Implicit metaphors are substitutions or ellipses that are directly or indirectly related to underlying conceptual metaphors. In (2), for example, the pronoun *that* co-refers with the preceding metaphorically used *antidote*; the determiner *this* in the following sentence is also used metaphorically: it refers to what is talked about in the previous sentence, which is an abstraction of its basic, physical meaning (Steen et al., 2010b). Metaphorical usage of substitutions are important for rendering cohesive discourse (Steen et al., 2010b).

- (2) Fortunately, there is a single *antidote* effective against both these myths; and *that* is to start all over again *This* antidote is effective against the romantic-individualist myth (BNC)

Discourse-level information is essential for identifying *extended metaphors*, sustained use of the same metaphors in a discourse fragment. A typical example is the 'All the World's a Stage' speech written by William Shakespeare. As is presented below, the speech begins with a WORLD AS STAGE metaphor, and proceeds with exploring various target-source pairs within the metaphor. Note that the last two lines could be mistaken as literal if presented as independent sentences.

- (3) All the world's a stage, / And all the men and women merely players: / They have their exits and their entrances; / And one man in his time plays many parts, [...]

2.3 Metaphor comprehension

There has been a debate about whether metaphor comprehension is a comparison or a categorisation process. According to the comparison view (e.g.,

Ortony, 1979), when one processes (1), for example, one searches for the shared properties and/or relational structures of the two domains, such as being pleasing to the eye, and the possibility of being fragile. Scholars advocating the categorisation view (e.g., Glucksberg and Keysar, 1990), however, would argue that the object, *a flower*, refers to a super-ordinate category that both BEAUTY and FLOWER belong to.

The Career of Metaphor Theory (Bowdle and Gentner, 2005) suggests that both comparison and categorisation are possible paths of metaphor processing; which path is chosen depends on an interaction between the conventionality of the metaphor and its linguistic realisation. More specifically, novel metaphors are processed through comparison; as a metaphor becomes conventionalised, people tend to process it through categorisation, which is less cognitively demanding than a comparison process (Bowdle and Gentner, 2005).

An inevitable result of metaphor processing is the emergence of features not inherent in the target or the source domain (Tourangeau and Rips, 1991). This notion of *emergent meaning* corresponds to the connotations or inference of metaphorical language in linguistics literature. Consider the difference in meaning between *attack one's argument* and *criticise one's argument*. The first contains far richer shades of meaning than the second. It might be possible to express a similar meaning without using non-literal language, but the expression is unlikely to be as concise as the use of a single word, *attack*.

2.4 Metaphor use in communication

The linguistic expression of emotional states often employs metaphor (Fainsilber and Ortony, 1987; Fussell and Moss, 1998). As emotion is an abstract domain, it goes with CMT and embodied cognition that we employ more concrete domains, such as physical or bodily experience, to conceptualise it. Moreover, since metaphor gives rise to emergent meaning, metaphorical language has a stronger emotional effect than literal language, regardless of the source and target domains involved (Blanchette et al., 2001; Crawford, 2009; Mohammad et al., 2016). For instance, Citron and Goldberg (2014) found that metaphorical expressions involving taste (e.g., *She looked at him sweetly*), evoke a higher emotional response than their literal counterparts (e.g., *She looked at him kindly*).

Metaphor has also proved to be an effective persuasive device (Sopory and Dillard, 2002; van Stee, 2018). The persuasive power of metaphors is pronounced in metaphoric framing effect. Since metaphors encourage a particular way to conceptualise the target domain, repeated use of the same metaphors throughout discourses in mass media tends to affect how the public perceives and reacts to societal issues that belong to the target domain (Lakoff, 1991; Entman, 2003; Lakoff and Wehling, 2012). For instance, participants in a series of studies (Thibodeau and Boroditsky, 2011, 2013) favoured different social solutions to crime after reading articles that associate CRIME with different source domains. Moreover, while the participants could identify the implicitly advocated solutions given a metaphor, they were unaware of the influence of the metaphors on their own preference.

The Deliberate Metaphor Theory is an attempt to deal with the intention behind metaphor use systematically (Steen, 2008, 2017). The theory defines *deliberate metaphor use* as the intentional introduction of a topic shift or perspective change to the discourse. Deliberate metaphor use is associated with online metaphor processing, the construction of conceptual metaphors during text comprehension (Steen, 2017). Examples of deliberate metaphors include conventional metaphors instantiated as copula metaphors and all novel metaphors. A systematic procedure for the identification of potential deliberate metaphors has also been proposed (Reijnierse et al., 2018).

3 Metaphor datasets

Metaphoricity annotations Tsvetkov et al. (2014) released a dataset (henceforth: TSV) consisting of an equal number (884) of metaphorical and non-metaphorical adjective-noun (AN) phrases collected from the web. The phrases were stated to be verified by multiple annotators, but the criteria for metaphor annotation were not provided.

The dataset released by Mohammad et al. (2016) (henceforth: MOH) consists of 1639 (1230 literal and 409 metaphorical) sentences extracted from WordNet, manifesting the use of 440 verbs. Metaphoricity annotation of the verb uses was obtained through crowdsourcing. Note that the WordNet sentences are mainly instances of conventional metaphor. The specified association between word senses and metaphoricity makes it easier to determine the source and target domains involved.

The Language Computer Corporation metaphor datasets (LCC) (Mohler et al., 2016) contain linguistic metaphors extracted from web corpora for a small set of target domains in four languages: English, Spanish, Russian, and Farsi. Metaphoricity is annotated at sentence level, and the conceptual metaphors and affect information is specified.

VU Amsterdam Metaphor Corpus (VUA) (Steen et al., 2010b) is the fruit of the authors' application of MIPVU. The metaphoricity of each lexical unit (187,570 in total) in a subset of the British National Corpus (BNC Consortium, 2007) was annotated. VUA is the only metaphor corpus used in studies of automated metaphor identification that is built by cognitive linguists, and the only one that deals with the metaphoricity of function words.

Metaphor paraphrase datasets Mohammad et al. (2016) also obtained literal paraphrases of the sentences, in which the metaphorically used verbs are replaced by their synonyms (171 pairs of sentences in total), selected by the authors. Note that the literal paraphrases were considered to convey less emotion than the original metaphorical sentences in their experiment. It is therefore questionable to what extent the paraphrases capture the connotations of the metaphorical sentences.

Bizzoni and Lappin (2018) built a metaphor paraphrase dataset containing 200 sets of 5 sentences; 4 paraphrases at varying levels of aptness are provided for each metaphorical sentence. Apart from verbs, the dataset also includes metaphorical uses of adjectives, copula metaphors, and multi-word metaphors. The dataset takes into account the connotations of the metaphorical sentences to some extent. For instance, candidate paraphrases for the copula metaphor *My job is a dream* include *I love my job* and *I hate my job*, which indicate opposite sentiment poles. A metaphor processing system will need to infer the sentiment to select the apt paraphrase.

4 Recent metaphor processing systems

4.1 Automated metaphor identification

4.1.1 Neural architectures

Most neural models treat metaphor identification as a sequence labelling task, outputting a sequence of metaphoricity labels for a sequence of input words (usually a sentence) (Bizzoni and Ghanimifard, 2018; Chen et al., 2020; Dankers et al., 2019; Gao et al., 2018; Gong et al., 2020; Mao et al., 2019;

Mykowiecka et al., 2018; Pramanick et al., 2018; Su et al., 2020; Wu et al., 2018). The first sequence labelling systems typically represented an input sentence as a sequence of pre-trained word embeddings and produced a task- and context-specific sentence representation through bidirectional long short-term memory (BiLSTM) (Dankers et al., 2019; Gao et al., 2018; Mykowiecka et al., 2018; Pramanick et al., 2018). Bizzoni and Ghanimifard (2018) experimented with separating long sentences into smaller chunks, which led to a 6% increase in F-score when using a BiLSTM architecture. Their BiLSTM system outperformed their compositional system, which employs a sequence of fully-connected neural networks (NNs) and essentially performs bigram phrase composition to modulate the representation of input words with respect to their neighbours. BiLSTM models also outperformed bidirectional gated recurrent unit (BiGRU) models in the study of Mykowiecka et al. (2018). From Gao et al. (2018), the contextualised Embeddings from Language Models (ELMo) (Peters et al., 2018) began to be used in addition to the context-free Global Vectors (GloVe) (Pennington et al., 2014) for representing input sentences (Dankers et al., 2019; Gao et al., 2018; Mao et al., 2019). The most recent systems adopt a fine-tuning approach, employing pre-trained contextual language models such as Bidirectional Encoder Representations from Transformers (BERT) (Chen et al., 2020; Dankers et al., 2019) and RoBERTa (Gong et al., 2020). A summary of the best performing recent metaphor identification systems is presented in Table 1 (a summary of all systems can be found in Appendix A).

Several BiLSTM-based systems consider both contextualised and pre-trained representations in the classification layers (Mao et al., 2019; Swarnkar and Singh, 2018). The Di-LSTM Contrast system (Swarnkar and Singh, 2018) encodes the left- and right-side context of a target word using forward and backward LSTMs. The classification is based on a concatenation of the target-word vector and its difference with the encoded context. Mao et al. (2019) combined GloVe and BiLSTM hidden states for sequence labelling, which outperformed the best model in the 2018 VUA All POS track.

Wu et al. (2018) and Su et al. (2020) employed separate encoding of local and long-range context. Wu et al. (2018) used a convolutional neural network (CNN) and a BiLSTM to extract local and

sentence context respectively. [Su et al. \(2020\)](#) used separate Transformer encoder layers to encode global and local text features for each word. The two systems achieved the best performance on the VUA All POS tracks in their respective shared tasks ([Leong et al., 2018, 2020](#)).

Modelling metaphor in discourse Several recent approaches have also incorporated wider discourse properties in their models. [Mu et al. \(2019\)](#) focused on the metaphoricity of verbs. They used general-purpose word, sentence and document embedding methods (e.g. GloVe, ELMo, doc2vec ([Le and Mikolov, 2014](#)), skip-thought ([Kiros et al., 2015](#))) to represent the surrounding paragraphs. Their system feeds into a gradient boosting decision tree classifier ([Chen and Guestrin, 2016](#)) a concatenation of three feature vectors, representing 1) the lemma of the target verb, 2) its subject and direct object, and 3) its surrounding paragraph. All the representations are learned from embedding methods. Representing the features with ELMo led to the highest F-score (0.668), using the VUA Verbs 2018 shared task data.

[Dankers et al. \(2020\)](#) fine-tuned a BERT model that receives a discourse fragment as input. Using hierarchical attention (which computes both token- and sentence-level attention) after the encoded layers achieved better performance than applying general attention to all tokens. Both [Dankers et al. \(2020\)](#) and [Mu et al. \(2019\)](#) thus demonstrated the importance of context beyond sentence for word-level metaphor identification. Their qualitative analysis shows that co-reference resolution is one of the driving factors in the performance increase.

Note that the above systems use discourse to aid in detecting linguistic metaphors whose metaphoricity is otherwise ambiguous, but do not monitor whether a metaphor is sustained throughout a fragment of discourse. Thus, they are unlikely to be directly applicable to the identification of extended metaphors or metaphoric frames.

4.1.2 Cognitively-inspired approaches

Categorical features Metaphor processing is concerned with how concepts are organised in the brain/mind, and is closely related to categorisation. It therefore makes sense to employ categorical features for metaphor identification. [Tekirođlu et al. \(2015\)](#) tested the use of sensorial categories (the five human senses) for identifying AN synaes-

thetic metaphors (e.g., *sweet music*, *soft light*). Using sensorial categories in addition to WordNet supersenses, concreteness, and imageability led to improved performance (accuracy 0.890 vs 0.845 on TSV). [Mykowiecka et al. \(2018\)](#) used another general-purpose resource, the Harvard IV psychosocial dictionary¹, which includes categories of emotions, people and animals, objects, places, etc. However, it did not lead to consistent improvement in model performance. [Bulat et al. \(2017\)](#) compared property-based and linguistic embeddings for input word representation. They obtained property-based word embeddings by mapping linguistic word embeddings onto a conceptual space, using as training data a frequency-based human property-norm dataset ([McRae et al., 2005](#)). Using property-based word embeddings led to a 4% increase in F-score on TSV.

Sensory features An important function of conceptual metaphors, according to CMT, is to use bodily experience to understand abstract concepts; concreteness features have therefore also proved useful for automated metaphor identification ([Bizzone and Ghanimifard, 2018](#); [Turney et al., 2011](#)). [Shutova et al. \(2016\)](#) tested combination of visual and linguistic embeddings on MOH and TSV. The multimodal system outperformed the monomodal systems in both tests. [Gong et al. \(2020\)](#) also included both categorical and sensory features. In addition to RoBERTa, the system employs concreteness features, topic distributions, WordNet classes, VerbNet classes, verb clusters, and part of speech (POS), which led to improvements in performance on VUA All POS.

Word-context incongruity The neural model of [Swarnkar and Singh \(2018\)](#) computes the difference between a target word and its context. This operation can be associated with the comparison view of metaphor, including CMT, and is reflected in MIPVU. [Shutova et al. \(2016\)](#) used cosine similarity between word or phrase embeddings to predict the metaphoricity of verb-noun (VN) or AN pairs; a word pair is marked metaphorical if the cosine similarity is below a trained threshold. The systems reached an F-score of 0.71 on MOH and 0.76 on TSV using linguistic embeddings alone. [Rei et al. \(2017\)](#) proposed a supervised similarity network, which learns to calculate weighted co-

¹<http://www.wjh.harvard.edu/~inquirer/homecat.htm>

| Study | Input | Input features | Architecture | Dataset | Performance (default F) |
|--------------------------------|-----------|--|-------------------------|---------------|-------------------------|
| Tekiroğlu et al. (2015) | AN | Token, Sensorial, Word-Class, Conc, Image ★ | Random Forest | TSV | A 0.890 |
| Shutova et al. (2016) | AN/VN | Token, VisualEmb ★ | skip-gram; CNN ★ | TSV/MOH | 0.79/0.75 |
| Bulat et al. (2017) | AN | AttributeEmb ★ | SVM | TSV | 0.77 |
| Rei et al. (2017) | AN/VN | Token, Attribute-Emb (AN-TSV) ★ | Supervised similarity ★ | TSV/MOH | 0.811 /0.742 |
| Mao et al. (2018) | S | Token | Cosine sim. ★ | MOH | 0.75 |
| Gao et al. (2018) | S | Token (GloveElmo) | BiLSTM | VUA Verbs | 0.697 |
| Gong et al. (2020) | S | Token | RoBERTa | VUA Verbs | 0.771 |
| Bizzoni and Ghanimifard (2018) | Chunk | Token | BiLSTM | VUA All POS | 0.621 |
| Wu et al. (2018) | S | Lemma, POS, WordCluster ★ | CNN-BiLSTM (ensemble) | VUA All POS | 0.651 |
| Dankers et al. (2020) | Discourse | Token | BERT base | VUA All POS | 0.715 |
| Chen et al. (2020) | S | Token | BERT large | VUA All POS | 0.718 |
| Gong et al. (2020) | S | Token, POS, Conc, TopicDistr, WordClass, WordCluster ★ | RoBERTa (ensemble) | VUA All POS | 0.730 |
| Mao et al. (2019) | S | Token (GloveElmo) | BiLSTM + Glove | VUA All POS | 0.740 |
| Su et al. (2020) | S | Token, POS | Transformers (ensemble) | VUA All POS | 0.769 |
| Pramanick et al. (2018) | S | Token, Lemma, POS, WordLemma | BiLSTM-CRF | VUA all words | 0.6740 |
| Gao et al. (2018) | S | Token (GloveElmo) | BiLSTM | VUA all words | 0.726 |
| Dankers et al. (2019) | S | Token | BERT + emotion ★ | VUA all words | 0.769 |

Table 1: Summary metaphor identification systems, sorted by dataset and performance. S: sentence. Chunk: sentence fragment. Conc: concreteness. Image: imageability. Emb: embeddings. A: accuracy. ★: use of a cognitively-inspired method.

sine similarity in a task-specific vector space. This allows the model to learn which dimensions of similarity are most relevant in particular metaphoric comparisons. The system outperformed Shutova et al. (2016) without the use of visual representations. Mao et al. (2018) dealt with word-level metaphor identification in sentences. Given a target word in a sentence, the system searches WordNet for the synonym or direct hypernym of the target word most similar to the context words. The target word is metaphorically used if its cosine similarity with the selected word is below a threshold.

Metaphor and emotion Motivated by the close relationship between metaphor use and the expression of emotions, Gargett and Barnden (2015) successfully used emotion features, amongst others, for metaphor identification. Kozareva (2013) and Strzalkowski et al. (2014) model the affect carried by metaphors in texts in different languages. Most recently, Dankers et al. (2019) employed multitask learning (MTL) to train models of metaphor identification and emotion prediction jointly. Models were based on BiLSTM and BERT, with a range of MTL architectures. The emotion prediction task used the Valence-Arousal-Dominance model

(Mehrabian, 1996), and each of these was considered separately in a MTL setup. The best performance was achieved with the BERT architecture. Dankers et al. (2019) found that while predicting dominance was the most challenging task on the emotion side, it also provided the greatest and most consistent improvements to metaphor identification, and vice versa.

4.2 Automated metaphor interpretation

Recent research on automated metaphor interpretation mainly followed Shutova (2010) in treating the problem as a paraphrasing task. Su et al. (2017) proposed a property transfer process for the interpretation of copula metaphors in Chinese and English. Given a target-source pair, the system extracts source-domain properties from hand-crafted databases², represented as adjectives. It then selects the property that contributes the most to the semantic relatedness of the target-source pair. The resultant pair of target-domain word and property is taken as interpretation of the copula metaphor. The metaphor LOVE IS TIDE, for instance, was interpreted as *The love is unstoppable*. Su et al.

²For instance, the adjective taxonomy provided by Sardonius: <http://bonnat.ucd.ie/sardonius/>.

(2017) thus took into account the emergent meaning of metaphors. Note that the literal explanations they obtained can be regarded as explanations of conceptual metaphors as well.

The above-mentioned system of Mao et al. (2018) performs both metaphor interpretation and identification: the synonym or hypernym of the metaphoric target word that matches the context can be considered the literal counterpart of the target word. The output interpretations were evaluated through a English-Chinese machine translation task: words classified as metaphorically used were paraphrased prior to translation. The system improved the accuracy of both Google Translation (0.60 vs 0.34) and Bing Translation (0.66 vs 0.42) on the metaphorical sentences in MOH. The experiment thus demonstrated the value of metaphor interpretation for machine translation.

Bizzoni and Lappin (2018) presented a neural model that detects paraphrases of sentences containing metaphor use. Given a metaphorical sentence and a candidate paraphrase, the system uses parallel CNN-LSTM blocks to encode the two sentences separately. The sentence representations are then merged and passed through fully-connected layers to produce a prediction. The system reached an F-score of 0.746 in a binary classification task and a Pearson correlation of 0.553 in a paraphrase ordering task, on the authors' own dataset.

5 Discussion

Metaphor identification Word-level metaphor identification is the most studied topic in automated metaphor processing. The systems utilise the latest NLP techniques and findings about metaphor in cognitive sciences; the state-of-the-art performance has been advanced quickly. A limitation in these studies, however, is that the systems do not deal with different types of metaphors. The high percentage of conventional metaphors in VUA (Steen et al., 2010a,b) implies that systems trained on the corpus may not capture conventional and novel metaphor uses equally well. It is the processing of novel metaphors that is a particular challenge for NLU systems, and the current experimental setup does not allow us to explicitly benchmark metaphor identification systems in this ability. The same goes for deliberate metaphor use, which is essential for modelling metaphor in communication and applies to both conventional and novel metaphors.

Whilst there have been huge advances in perfor-

mance in word-level metaphor identification, the broader tasks of identifying conceptual metaphors, extended metaphors, and metaphoric framing, have been largely ignored. Conceptual metaphors form the framework for both extended metaphors and metaphoric framing, from a cognitive perspective, and these latter two are a key part of the intentional use of metaphor in communication.

Metaphor interpretation Recent studies on automated metaphor interpretation largely focused on literal paraphrases of metaphorical sentences. The inference of emergent meaning is, however, an integral part of metaphor comprehension and should be captured by metaphor interpretation systems. To capture emergent meaning is to reveal the properties and relations of the target domain emphasised by the metaphor and any additional connotations and emotional effects that arise in the interpretation process. In this respect, the existing paraphrase datasets (Shutova, 2010; Mohammad et al., 2016) are unlikely to reflect the emergent meaning of the metaphors. Bizzoni and Lappin (2018) consider the sentiment poles of metaphorical expressions, but it is questionable to what extent the apt paraphrases also reflect other dimensions, such as the intensity of an emotion.

One way to deal with the inference of linguistic metaphors is to convert it to the task of paraphrasing copula metaphors. As is the case with Su et al. (2017) and Bizzoni and Lappin (2018), the paraphrasing of copula metaphors has been dealt with as obtaining the property of the target domain that is indicated by the metaphor. We can thus deal with the inference of any linguistic metaphor by first identifying its conceptual metaphor and then treating the conceptual metaphor as a copula metaphor. To tackle the task, we will need datasets of the emergent meaning of a wide range of conceptual metaphors. Such datasets can be built from data obtained in psycholinguistic studies about emergent meaning, in which participants were asked to provide concepts or properties that were related to metaphors as well as their target and source domains alone (e.g., Gineste et al., 2000).

A more comprehensive model of metaphor interpretation would also consider the complexity of linguistic metaphors, which essentially arises from the social aspects of metaphor use, such as the various roles that metaphor plays in communication.

Metaphor generation To the best of our knowledge, automated metaphor generation has largely remained an unexplored area. As metaphor is an important tool for reasoning and linguistic expression, the task is a crucial step toward the generation of human-like language. Note, however, that despite the scarcity of explicit attempts at metaphor generation, it is likely to be an inherent part of existing natural language generation (NLG) systems already. Metaphorical use of function words, for instance, are unavoidable for linguistic expression. Highly conventionalised metaphorical use of some content words may also be captured by current NLG techniques. For instance, the use of *bright* for the description of colours (e.g., *bright blue eyes*) is considered metaphorical in VUA whereas it is one of the most frequent usages of the adjective. It is likely that data-driven NLG (including neural models) instantiate this usage of the word without ‘knowing’ that it is a metaphor.

It is thus in the intentional use of metaphors that the difficulty of metaphor generation lies. The development of metaphor generation systems therefore relies on the modelling of metaphor use in communication, which we discuss below.

Implications for other NLP tasks That linguistic metaphors have to do with concepts in two different domains implies that their automated processing is potentially indispensable for resolving referents in text. This is important for a number of NLP applications, such as document summarisation, question answering, and machine translation. The resolution of the actual referents of MRWs is mainly concerned with indirect metaphors, which use source-domain words to refer to target-domain concepts. Note that implicit metaphors can be indirect metaphors as well. It is therefore necessary to take the metaphoricity of function words such as pronouns and determiners into account when it comes to resolving the referent of metaphorically used words. We suggest that the processing of implicit metaphors, especially substitutions, can be facilitated by combining metaphor identification with coreference resolution, as the metaphorically used pronouns essentially refer to the same referents as a preceding metaphorical expression.

Furthermore, there has not been much discussion of the form of metaphor interpretation that would be most beneficial for downstream NLP tasks. Most research on automated metaphor interpretation focuses on producing human-readable

literal paraphrases. However, it is likely that internal model representations that capture the rich meaning of metaphorical language can be more elegantly and successfully integrated in the task models. Consider, for instance, machine translation. After obtaining an internal representation of a metaphorical sentence in the source language, the system can then search for a literal or metaphorical expression in the target language that best matches the representation. We suggest that the development of such a system can begin with investigating the hidden states of existing deep learning models of metaphor identification: the vectors used to output metaphoricity labels may already capture the emergent meaning of the input sentences.

Metaphor in communication In addition to the semantic aspects, metaphor use often also has social implications. Metaphor processing is an essential component of opinion mining: since metaphorical language tends to express stronger emotions than literal language, the presence of metaphors and their inferences would contribute to the sentiment conveyed by a text. It is also worth considering distinguishing deliberate and non-deliberate metaphor uses: focusing on the former may achieve a better approximation of the speaker/writer’s sentiment and communicative goals.

Metaphor processing should thus also be incorporated into dialogue modelling. Metaphors could be extended or even evolve with the turn taking in a conversation, as speakers seek mutual understanding, or the resolution of a difference of opinion in argumentative discourse such as debates. Systems that are concerned with speaker’s intent will therefore benefit from the identification and interpretation of metaphorical language.

Given the relationship between metaphor and persuasiveness, it may be beneficial to model metaphor use in other types of argumentative discourse, such as political speeches and news articles. A recent development in this regard is a MTL approach to modelling metaphor, emotion, and framing in political texts (Huguet Cabot et al., 2020).

6 Conclusion

In this paper, we reviewed recent advances in automated metaphor processing and brought to light its less explored aspects, pertaining to cognitive processing of metaphor, its role in communication and its connection with downstream NLP applications. We hope that our work will encourage further in-

terdisciplinary research on metaphor and facilitate progress in this important and fascinating area.

Acknowledgements

We are grateful to the NAACL reviewers for their helpful feedback. Martha Lewis is supported by NWO Veni ‘Metaphorical Meanings for Artificial Agents’

References

- Lisa Aziz-Zadeh and Antonio Damasio. 2008. [Embodied semantics for actions: Findings from functional brain imaging](#). *Journal of Physiology - Paris*, 102(1-3).
- Lawrence W Barsalou. 1999. Perceptual symbol systems. *Behavioral and brain sciences*, 22(4):577–660.
- 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, New Orleans, Louisiana. Association for Computational Linguistics.
- Yuri Bizzoni and Shalom Lappin. 2018. [Predicting human metaphor paraphrase judgments with deep neural networks](#). In *Proceedings of the Workshop on Figurative Language Processing*, pages 45–55, New Orleans, Louisiana. Association for Computational Linguistics.
- Isabelle Blanchette, Kevin Dunbar, John Hummel, and Richard Marsh. 2001. [Analogy use in naturalistic settings: The influence of audience, emotion and goals](#). *Memory and Cognition*, 29(5).
- BNC Consortium. 2007. [The British National Corpus](#). BNC XML Edition.
- Brian F. Bowdle and Dedre Gentner. 2005. [The career of metaphor](#). *Psychological Review*, 112(1):193–216.
- Luana Bulat, Stephen Clark, and Ekaterina Shutova. 2017. [Modelling metaphor with attribute-based semantics](#). In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 523–528, Valencia, Spain. Association for Computational Linguistics.
- Tianqi Chen and Carlos Guestrin. 2016. [XGBoost: A scalable tree boosting system](#). In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’16, page 785–794, New York, NY, USA. Association for Computing Machinery.
- Xianyang Chen, Chee Wee (Ben) 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, Online. Association for Computational Linguistics.
- Francesca MM Citron and Adele E Goldberg. 2014. Metaphorical sentences are more emotionally engaging than their literal counterparts. *Journal of cognitive neuroscience*, 26(11):2585–2595.
- Elizabeth Crawford. 2009. [Conceptual Metaphors of Affect](#). *Emotion Review*, 1(2).
- Verna Dankers, Karan Malhotra, Gaurav Kudva, Volodymyr Medentsiy, and Ekaterina Shutova. 2020. [Being neighbourly: Neural metaphor identification in discourse](#). In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 227–234, Online. Association for Computational Linguistics.
- 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, Hong Kong, China. Association for Computational Linguistics.
- Erik-Lân Do Dinh and Iryna Gurevych. 2016. [Token-level metaphor detection using neural networks](#). In *Proceedings of the Fourth Workshop on Metaphor in NLP*, pages 28–33, San Diego, California. Association for Computational Linguistics.
- Robert Entman. 2003. Language: The Loaded Weapon. *Political Communication*, (20):415–432.
- Lynn Fainsilber and Andrew Ortony. 1987. [Metaphorical uses of language in the expression of emotions](#). *Metaphor and Symbolic Activity*, 2(4):239–250.
- Susan R Fussell and Mallie M Moss. 1998. Figurative language in emotional communication. *Social and cognitive approaches to interpersonal communication*, pages 113–141.
- Vittorio Gallese and George Lakoff. 2005. The brain’s concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive neuropsychology*, 22(3-4):455–479.
- Ge Gao, Eunsol Choi, Yejin Choi, and Luke Zettlemoyer. 2018. [Neural metaphor detection in context](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 607–613, Brussels, Belgium. Association for Computational Linguistics.

- Andrew Gargett and John Barnden. 2015. [Modeling the interaction between sensory and affective meanings for detecting metaphor](#). In *Proceedings of the Third Workshop on Metaphor in NLP*, pages 21–30, Denver, Colorado. Association for Computational Linguistics.
- Marie-Dominique Gineste, Bipin Indurkha, and Veronique Scart. 2000. [Emergence of features in metaphor comprehension](#). *Metaphor and Symbol*, 15(3):117–135.
- Sam Glucksberg and Boaz Keysar. 1990. [Understanding metaphorical comparisons: Beyond similarity](#). *Psychological Review*, 97(1):3–18.
- 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, Online. Association for Computational Linguistics.
- Pere-Lluís Huguet Cabot, Verna Dankers, David Abadi, Agneta Fischer, and Ekaterina Shutova. 2020. [The Pragmatics behind Politics: Modelling Metaphor, Framing and Emotion in Political Discourse](#). In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4479–4488, Online. Association for Computational Linguistics.
- Hyeju Jang, Seungwhan Moon, Yohan Jo, and Carolyn Rosé. 2015. [Metaphor detection in discourse](#). In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 384–392, Prague, Czech Republic. Association for Computational Linguistics.
- Ryan Kiros, Yukun Zhu, Russ R Salakhutdinov, Richard Zemel, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. 2015. [Skip-thought vectors](#). In *Advances in neural information processing systems*, pages 3294–3302.
- Zornitsa Kozareva. 2013. [Multilingual affect polarity and valence prediction in metaphor-rich texts](#). In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 682–691, Sofia, Bulgaria. Association for Computational Linguistics.
- George Lakoff. 1991. [Metaphor and war: The metaphor system used to justify war in the Gulf](#). *Peace Research*, 23:25–32.
- George Lakoff and Mark Johnson. 1980. *Metaphors we live by*. University of Chicago Press.
- George Lakoff and Elisabeth Wehling. 2012. *The Little Blue Book: The Essential Guide to Thinking and Talking Democratic*. Free Press, New York.
- Quoc Le and Tomas Mikolov. 2014. [Distributed representations of sentences and documents](#). In *Proceedings of the 31st International conference on machine learning*, pages 1188–1196.
- Chee Wee Leong, Beata Beigman Klebanov, and Ekaterina Shutova. 2018. [A report on the 2018 VUA metaphor detection shared task](#). In *Proceedings of the Workshop on Figurative Language Processing*, pages 56–66, New Orleans, Louisiana. Association for Computational Linguistics.
- Chee Wee (Ben) Leong, Beata Beigman Klebanov, Chris Hamill, Egon Stemle, Rutuja Ubale, and Xianyang Chen. 2020. [A report on the 2020 VUA and TOEFL metaphor detection shared task](#). In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 18–29, Online. Association for Computational Linguistics.
- Rui Mao, Chenghua Lin, and Frank Guerin. 2018. [Word embedding and WordNet based metaphor identification and interpretation](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1222–1231, Melbourne, Australia. Association for Computational Linguistics.
- Rui Mao, Chenghua Lin, and Frank Guerin. 2019. [End-to-end sequential metaphor identification inspired by linguistic theories](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3888–3898, Florence, Italy. Association for Computational Linguistics.
- Ken McRae, George S. Cree, Mark S. Seidenberg, and Chris Mcnorgan. 2005. [Semantic feature production norms for a large set of living and nonliving things](#). *Behavior Research Methods*, 37:547–559.
- Albert Mehrabian. 1996. [Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament](#). *Current Psychology*, 14(4):261–292.
- Saif Mohammad, Ekaterina Shutova, and Peter Turney. 2016. [Metaphor as a medium for emotion: An empirical study](#). In *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics*, pages 23–33, Berlin, Germany. Association for Computational Linguistics.
- Michael Mohler, Mary Brunson, Bryan Rink, and Marc Tomlinson. 2016. [Introducing the LCC metaphor datasets](#). In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, pages 4221–4227, Portorož, Slovenia. European Language Resources Association (ELRA).
- Jesse Mu, Helen Yannakoudakis, and Ekaterina Shutova. 2019. [Learning outside the box: Discourse-level features improve metaphor identification](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 596–601, Minneapolis, Minnesota. Association for Computational Linguistics.

- Agnieszka Mykowiecka, Aleksander Wawer, and Malgorzata Marciniak. 2018. [Detecting figurative word occurrences using recurrent neural networks](#). In *Proceedings of the Workshop on Figurative Language Processing*, pages 124–127, New Orleans, Louisiana. Association for Computational Linguistics.
- Andrew Ortony. 1979. Beyond literal similarity. *Psychological Review*, 86(3):161–180.
- Jeffrey Pennington, Richard Socher, and Christopher 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, Doha, Qatar. Association for Computational Linguistics.
- Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. [Deep contextualized word representations](#). In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237, New Orleans, Louisiana. Association for Computational Linguistics.
- Malay Pramanick, Ashim Gupta, and Pabitra Mitra. 2018. [An LSTM-CRF based approach to token-level metaphor detection](#). In *Proceedings of the Workshop on Figurative Language Processing*, pages 67–75, New Orleans, Louisiana. Association for Computational Linguistics.
- Marek Rei, Luana Bulat, Douwe Kiela, and Ekaterina Shutova. 2017. [Grasping the finer point: A supervised similarity network for metaphor detection](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1537–1546, Copenhagen, Denmark. Association for Computational Linguistics.
- W. Gudrun Reijniere, Christian Burgers, Tina Krennmayr, and Gerard J. Steen. 2018. [Dmip: A method for identifying potentially deliberate metaphor in language use](#). *Corpus Pragmatics*, 2(2):129–147.
- Ekaterina Shutova. 2010. [Automatic metaphor interpretation as a paraphrasing task](#). In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 1029–1037, Los Angeles, California. Association for Computational Linguistics.
- Ekaterina Shutova. 2011. *Computational Approaches to Figurative Language*. Ph.D. thesis, University of Cambridge, UK.
- Ekaterina Shutova. 2015. [Design and evaluation of metaphor processing systems](#). *Computational Linguistics*, 41(4):579–623.
- 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, San Diego, California. Association for Computational Linguistics.
- Pradeep Sopyry and James Price Dillard. 2002. [The persuasive effects of metaphor: A meta-analysis](#). *Human Communication Research*, 28(3):282–419.
- Gerard J. Steen. 2008. [The paradox of metaphor: Why we need a three-dimensional model of metaphor](#). *Metaphor and Symbol*, 23(4):213–241.
- Gerard J. Steen. 2017. [Deliberate Metaphor Theory: Basic assumptions, main tenets, urgent issues](#). *Intercultural Pragmatics*, 14(1):1–24.
- Gerard J. Steen, Aletta G. Dorst, J. Berenike Herrmann, Anna A. Kaal, and Tina Krennmayr. 2010a. [Metaphor in usage](#). *Cognitive Linguistics*, 21(4):765–796.
- Gerard J. Steen, Aletta G. Dorst, J. Berenike Herrmann, Anna A. Kaal, Tina Krennmayr, and Tryntje Pasma. 2010b. [A Method for Linguistic Metaphor Identification: From MIP to MIPVU](#). John Benjamins.
- Tomek Strzalkowski, Samira Shaikh, Kit Cho, George Aaron Broadwell, Laurie Feldman, Sarah Taylor, Boris Yamrom, Ting Liu, Ignacio Cases, Yuliya Peshkova, and Kyle Elliot. 2014. [Computing affect in metaphors](#). In *Proceedings of the Second Workshop on Metaphor in NLP*, pages 42–51, Baltimore, MD. Association for Computational Linguistics.
- Chang Su, Shuman Huang, and Yijiang Chen. 2017. [Automatic detection and interpretation of nominal metaphor based on the theory of meaning](#). *Neurocomputing*, 219:300–311.
- Chuangdong Su, Fumiyo Fukumoto, Xiaoxi Huang, Jiye Li, Rongbo Wang, and Zhiquan Chen. 2020. [Deep-Met: A reading comprehension paradigm for token-level metaphor detection](#). In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 30–39, Online. Association for Computational Linguistics.
- Krishnkant Swarnkar and Anil Kumar Singh. 2018. [Di-LSTM contrast : A deep neural network for metaphor detection](#). In *Proceedings of the Workshop on Figurative Language Processing*, pages 115–120, New Orleans, Louisiana. Association for Computational Linguistics.
- Serra Sinem Tekiroğlu, Gözde Özbal, and Carlo Straparava. 2015. [Exploring sensorial features for metaphor identification](#). In *Proceedings of the Third Workshop on Metaphor in NLP*, pages 31–39, Denver, Colorado. Association for Computational Linguistics.

- Paul H. Thibodeau and Lera Boroditsky. 2011. [Metaphors we think with: The role of metaphor in reasoning](#). *PLOS ONE*, 6(2):1–11.
- Paul H. Thibodeau and Lera Boroditsky. 2013. [Natural language metaphors covertly influence reasoning](#). *PLOS ONE*, 8(1):1–7.
- Roger Tourangeau and Lance Rips. 1991. [Interpreting and evaluating metaphors](#). *Journal of Memory and Language*, 30(4):452–472.
- 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, Baltimore, Maryland. Association for Computational Linguistics.
- Peter 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 Empirical Methods in Natural Language Processing*, pages 680–690, Edinburgh, Scotland, UK. Association for Computational Linguistics.
- Stephanie K. van Stee. 2018. [Meta-analysis of the persuasive effects of metaphorical vs. literal messages](#). *Communication Studies*, 69(5):545–566.
- Chuhan Wu, Fangzhao Wu, Yubo Chen, Sixing Wu, Zhigang Yuan, and Yongfeng Huang. 2018. [Neural metaphor detecting with CNN-LSTM model](#). In *Proceedings of the Workshop on Figurative Language Processing*, pages 110–114, New Orleans, Louisiana. Association for Computational Linguistics.

A Appendix: summary of metaphor identification systems

| Study | Input | Input features | Architecture | Dataset | Performance (default F) |
|--------------------------------|-----------|---|-------------------------|---------------|----------------------------|
| Tekirođlu et al. (2015) | AN | Token, Sensorial, WordClass, Conc, Image ★ | Random Forest | TSV | A 0.890 |
| Shutova et al. (2016) | AN/VN | Token, VisualEmb ★ | Skip-gram; CNN ★ | TSV/MOH | 0.79/0.75 |
| Bulat et al. (2017) | AN | AttributeEmb ★ | SVM | TSV | 0.77 |
| Rei et al. (2017) | AN/VN | Token, AttributeEmb (AN-TSV) ★ | Supervised similarity ★ | TSV/MOH | 0.811 /0.742 |
| Mao et al. (2018) | S | Token | Cosine similarity ★ | MOH | 0.75 |
| Gao et al. (2018) | S | Token (GloveElmo) | BiLSTM | VUA Verbs | 0.697 |
| Gong et al. (2020) | S | Token | RoBERTa (ensemble) | VUA Verbs | 0.771 |
| Pramanick et al. (2018) | S | Token, Lemma, POS, WordLemma | BiLSTM-CRF | VUA All POS | 0.5362 |
| Do Dinh and Gurevych (2016) | Chunk | Token, POS | MLP | VUA All POS | 0.5550 |
| Swamkar and Singh (2018) | S | Token | Di-LSTM Contrast ★ | VUA All POS | 0.570 |
| Mykowiecka et al. (2018) | S | Token | BiLSTM | VUA All POS | 0.583 |
| Bizzoni and Ghanimifard (2018) | Chunk | Token | BiLSTM | VUA All POS | 0.621 |
| Wu et al. (2018) | S | Lemma, POS, WordCluster ★ | CNN-BiLSTM (ensemble) | VUA All POS | 0.651 |
| Dankers et al. (2020) | Discourse | Token | BERT base | VUA All POS | 0.715 |
| Chen et al. (2020) | S | Token | BERT large | VUA All POS | 0.718 |
| Gong et al. (2020) | S | Token, POS, Conc, TopicDistr, WordClass, WordCluster ★ | RoBERTa (ensemble) | VUA All POS | 0.730 |
| Mao et al. (2019) | S | Token (GloveElmo) | BiLSTM + Glove/MHCA | VUA All POS | 0.740/0.743 |
| Su et al. (2020) | S | Token, POS | Transformers (ensemble) | VUA All POS | 0.769 |
| Pramanick et al. (2018) | S | Token, Lemma, POS, WordLemma | BiLSTM-CRF | VUA all words | 0.6740 |
| Gao et al. (2018) | S | Token (GloveElmo) | BiLSTM | VUA all words | 0.726 |
| Dankers et al. (2019) | S | Token | BERT + emotion ★ | VUA all words | 0.769 |
| Dankers et al. (2019) | S | Token | BERT + emotion ★ | LCC | r 0.614 |

Table 2: Metaphor identification systems, sorted by dataset and performance. S: sentence. Chunk: sentence fragment. Conc: concreteness. Image: imageability. Emb: embeddings. A: Accuracy. ★: use of a cognitively-inspired method.

B Appendix: effects of cognitively-inspired approaches

| Study | Categorical | Sensory | Incongruity | Emotion |
|---------------------------|---|------------------------|-------------|------------------|
| Tekirođlu et al. (2015) | Sensorial ⁺ ; WordClass | Conc; Image | - | - |
| Bulat et al. (2017) | AttributeEmb ⁺ | - | - | - |
| Rei et al. (2017) | AttributeEmb ⁺⁻ | - | CosSim | - |
| Mykowiecka et al. (2018) | Psychosocial ⁻ | - | - | - |
| Shutova et al. (2016) | - | VisualEmb ⁺ | CosSim | - |
| Gong et al. (2020) | TopicDiff; WordClass; WordCluster | Conc ⁺⁻ | - | - |
| Dankers et al. (2019) | - | - | - | MTL ⁺ |
| Swarnkar and Singh (2018) | - | - | CosSim | - |
| Mao et al. (2018) | - | - | CosSim | - |
| Jang et al. (2015) | WordClass | Conc | CosSim | - |

Table 3: Cognitively-inspired approaches for metaphor identification. Effect notation: positive⁺; negative⁻; not tested. Conc: concreteness. Image: imageability. Emb: embeddings. CosSim: cosine similarity. TopicDiff: cosine similarity between topic distributions.