

Conversational Metaphors in Use: Exploring the Contrast between Technical and Everyday Notions of Metaphor

Hyeju Jang, Mario Piergallini, Miaomiao Wen, and Carolyn Penstein Rosé

Language Technologies Institute

Carnegie Mellon University

Pittsburgh, PA 15213, USA

{hyejuj, mpiergal, mwen, cprose}@cs.cmu.edu

Abstract

Much computational work has been done on identifying and interpreting the meaning of metaphors, but little work has been done on understanding the motivation behind the use of metaphor. To computationally model discourse and social positioning in metaphor, we need a corpus annotated with metaphors relevant to speaker intentions. This paper reports a corpus study as a first step towards computational work on social and discourse functions of metaphor. We use Amazon Mechanical Turk (MTurk) to annotate data from three web discussion forums covering distinct domains. We then compare these to annotations from our own annotation scheme which distinguish levels of metaphor with the labels: *nonliteral*, *conventionalized*, and *literal*. Our hope is that this work raises questions about what new work needs to be done in order to address the question of how metaphors are used to achieve social goals in interaction.

1 Introduction

Our goal is to understand and characterize the ways that nonliteral language, especially metaphors, play a role in a variety of conversational strategies. In contrast to the large body of work on uncovering the intended propositional meaning behind metaphorical expressions, we are most interested in the illocutionary and perlocutionary force of the same contributions.

People use metaphorical expressions in a variety of ways in order to position themselves socially and express attitudes, as well as to make their point more effective, attractive, and convinc-

ing. Metaphors can be used to describe unfamiliar situations and feelings when the speaker feels that literal description is inadequate. They can also be used to display the speaker's creativity and wit. They can further be used as a tactic for persuasion or manipulation by foregrounding aspects that would not ordinarily be relevant. Cameron (2007) shows that we can understand social interactions and their contexts better by closely looking at these patterns of metaphor use.

Metaphors can vary in how conventionalized they are, from those which have lost their original concrete meanings to completely novel and vivid metaphors. Intuitively, it also makes sense that metaphors which are more conventional and less obviously metaphorical will be used with less conscious thought than more novel or vivid metaphors. There are thus reasons to suspect that distinguishing between levels of metaphoricity could give insight into patterns of use.

In this paper, we are interested in where we can draw a line between levels of metaphoricity. As a first step towards our long-term goal, we present a corpus study in three web discussion forums including a breast cancer support group, a Massive Open Online Course (MOOC), and a forum for street gang members, which cover distinctly different domains and have differing community structure. First, we investigate how laypeople intuitively recognize metaphor by conducting Amazon Mechanical Turk (MTurk) experiments. Second, we introduce a new annotation scheme for metaphorical expressions. In our annotation scheme, we try to map the metaphor spectrum of nonliteralness to three types of language: *nonliteral*, *conventionalized*, and *literal*. Our hope is that this distinction provides some benefit in examining the social and discourse functions of metaphor. Next, we compare MTurk

results with our annotations. Different people will place the dividing line between literal language and metaphorical language in different places. In this work we have the opportunity to gauge how much everyday conceptions of metaphoricity diverge from theoretical perspectives and therefore how much models of metaphoricity may need to be adapted in order to adequately characterize metaphors in strategic use.

The paper is organized as follows. Section 2 relates our work to prior work on annotation and a corpus study. Section 3 describes the data used for annotation. Section 4 illustrates the functions metaphor serves in discourse through a qualitative analysis of our data. Section 5 explains our annotation scheme. Section 6 presents our annotation and MTurk experiments. Section 7 discusses the results. Section 8 concludes the paper.

2 Relation to Prior Work

In this section, we introduce the two main bodies of relevant prior work on metaphor in language technologies: computational metaphor processing and metaphor annotation.

2.1 Computational Work on Metaphor

Much of the computational work on metaphor can be classified into two tasks: automatic identification and interpretation of metaphors.

Metaphor identification has been done using different approaches: violation of selectional preferences (Fass, 1991), linguistic cues (Goatly, 1997), source and target domain words (Stefanowitsch and Gries, 2006), clustering (Birke and Sarkar, 2006; Shutova et al., 2010), and lexical relations in WordNet (Krishnakumaran and Zhu, 2007). Gedigian et al. (2006) and Li and Sporleder (2010) distinguished the literal and non-literal use of a target expression in text. In addition, Mason (2004) performed source-target domain mappings.

Metaphor interpretation is another large part of the computational work on metaphor. Starting with Martin (1990), a number of researchers including Narayanan (1999), Barn- den and Lee (2002), Agerri et al. (2007), and Shutova (2010) have worked on the task. Metaphor identification and interpretation was performed simultaneously in (Shutova, 2013;

Shutova et al., 2013b).

As we have seen so far, much of the computation work has focused on detecting and uncovering the intended meaning behind metaphorical expressions. On the other hand, Klebanov and Flor (2013) paid attention to motivations behind metaphor use, specifically metaphors used for argumentation in essays. They showed a moderate-to-strong correlation between percentage of metaphorically used words in an essay and the writing quality score. We will introduce their annotation protocol in Section 2.2.

However, to the best of our knowledge, not much computational work has been done on understanding the motivation behind the use of metaphor besides that of Klebanov and Flor (2013). Our work hopefully lays additional foundation for the needed computational work.

2.2 Metaphor Annotation

One of the main challenges in computational work on metaphor is the lack of annotated datasets. Annotating metaphorical language is nontrivial because of a lack of consensus regarding annotation schemes and clear definitions. In this section, we introduce some work dedicated to metaphor annotation and a corpus study.

Wallington et al. (2003) conducted experiments to investigate what identifies metaphors. Two different teams annotated the same text with different instructions, one asked to label “interesting stretches” and the other “metaphorical stretches”. They also asked annotators to tag words or phrases that indicated a metaphor nearby, in order to investigate signals of metaphoricity.

Pragglejaz Group (2007) presented a metaphor annotation scheme, called the Metaphor Identification Procedure (MIP), which introduced a systematic approach with clear decision rules. In this scheme, a word is considered to be metaphorical if it is not used according to its most basic concrete meaning, and if its contextual meaning can be understood in comparison with the most basic concrete meaning. This method is relatively straightforward and can give high inter-reliability. Depending on how one decides upon the basic meaning of words, this scheme can be used for different applications. However, defining the basic meaning of a word is nontrivial, and following the def-

inition of basic meaning introduced in the paper tends to result in a large proportion of words being annotated as metaphor. Many of the annotated words would not be considered to be metaphors by a layperson due to their long and widespread usage.

Later works by Steen (2010), Shutova and Teufel (2010), and Shutova et al. (2013a) expanded upon MIP. Steen (2010) discussed the strengths and weaknesses of MIP, and introduced the Metaphor Identification Procedure VU University Amsterdam (MIPVU). Shutova and Teufel (2010) and Shutova et al. (2013a) added a procedure for identifying underlying conceptual mappings between source and target domains.

So far, these presented schemes do not distinguish between degrees of metaphoricity, and were not specifically designed for considering motivations behind metaphor use. Unlike the annotation schemes described above, Klebanov and Flor (2013) built a metaphor annotation protocol for metaphors relevant to arguments in essays. They were interested in identifying metaphors that stand out and are used to support the writer’s argument. Instead of giving a formal definition of a literal sense, the annotators were instructed to mark words they thought were used metaphorically, and to write down the point being made by the metaphor, given a general definition of metaphor and examples. Our work is similar to this work in that both corpus studies pay attention to motivations behind metaphor use. However, our work focuses on more conversational discussion data whereas they focused on essays, which are more well-formed.

3 Data

We conducted experiments using data from three different web forums including a Massive Open Online Course (MOOC), a breast cancer support group (Breastcancer), and a forum for street gang members (Gang). We randomly sampled 21 posts (100 sentences) from MOOC, 8 posts (103 sentences) from Breastcancer and 44 posts (111 sentences) from Gang.

We chose these three forums because they all offer conversational data and they all differ in terms of the social situation. The forums dif-

fer significantly in purpose, demographics and the participation trajectory of members. Therefore, we expect that people will use language differently in the three sets, especially related to metaphorical expressions.

MOOC: This forum is used primarily for task-based reasons rather than socializing. People participate in the forum for a course, and leave when the course ends. As a result, the forum does not have continuity over time; participants do not spend long time with the same people.

Breastcancer: People join this forum for both task-based and social reasons: to receive informational and emotional support. People participate in the forum after they are diagnosed with cancer, and may leave the forum when they recover. This forum is also used episodically by many users, but a small percentage of users stay for long periods of time (2 or more years). Thus, continuity allows shared norms to develop over years centered around an intense shared experience.

Gang: In this forum, members belong to a distinct subculture prior to joining, whereas Breastcancer and MOOC members have less shared identity before entering the forum. This forum is purely social. There is no clear endpoint for participation; members leave the forum whenever they are not interested in it any more. Users may stay for a week or two, or for years.

4 Qualitative Analysis

Metaphors can be used for a number of conversational purposes such as increasing or decreasing social distance or as a tactic of persuasion or manipulation (Ritchie, 2013). In this section, we perform a qualitative analysis on how metaphor functions in our data. We illustrate some examples from each domain with an analysis of how some functions of social positioning are observed.

The choice of metaphor may reflect something about the attitude of the speaker. For example, *journey* is a metaphor frequently used in the breast cancer support discussion forum¹ as seen in examples (2) – (5) from the Breastcancer forum. People compare chemotherapy to a *journey* by using metaphors such as *journey*, *road* and *moves along*. A *journey* has a beginning and a goal one travels towards, but people may take different paths.

¹<http://breastcancer.org>

This conveys the experience of cancer treatment as a process of progressing along a path, struggling and learning, but allows for each person's experience to differ without judgment of personal success or failure (Reisfield and Wilson, 2004). By contrast, another common metaphor compares cancer treatment to battles and war. This metaphor instead conveys an activity rather than passivity, a struggle against a defined foe, which can be won if one fights hard enough. But it also creates negative connotations for some patients, as forgoing treatment could then be seen as equivalent to surrender (ibid.).

- (1) Hello Ladies! I was supposed to start chemo in January, ... I cant start tx until that is done. So I will be *joining you on your journey* this month. I AM SICK OF the ANXIETY and WAITING.
- (2) So Ladies, please add another member to this club. Looks like we well all be *leaning on* each other. But I promise to *pick you up* if you *fall* if you can *catch* me once in a while!
- (3) The *road* seems long now but it really *moves along* fast.
- (4) I split *this journey* into 4 stages and I only deal with one.

In addition, using metaphors can have an effect of increasing empathetic understanding between the participants (Ritchie, 2013). We can see this in examples (1) – (4), where participants in the same thread use similar metaphors relating chemotherapy to a *journey*. Reusing each other's metaphors reduces emotional distance and helps to build empathic understanding and bonding through a shared perception of their situations.

Metaphor also serves to suggest associations between things that one would not normally associate. Example (5) from the MOOC forum frames participation in discussions as stepping into an arena, which refers to an area for sports or competition. By making such an analogy, it conveys an environment of direct competition in front of a large audience. It suggests that a student may be afraid of contributing to discussion because they may make a wrong statement or weak argument

and another person could counter their contributions, and they will be embarrassed in front of their classmates.

- (5) Hi, Vicki, great *point* – I do wish that teachers in my growing up years had been better facilitators of discussion that allowed EVERY-one to practice adn become skill-ful at speaking...I think in the early years some of us need some *hand-holding* in *stepping into the arena* and speaking

Metaphors can also be used simply as a form of wordplay, to display one's wit and creativity. This can be seen in the exchange in examples (6) – (8), from the Gang forum. A common metaphor used on that forum is to refer to someone as *food* to mean that they are weak and unthreatening. The writer in (6) expands on this metaphor to suggest that the other person is especially weak by calling him *dessert*, while the writer in (7) then challenges him to fight by exploiting the meaning of *hungry* as "having a desire for *food*". The first writer (8) then dismisses him as not worth the effort to fight, as he does not *eat vegetables*.

- (6) So If She Is *Food* That Must Make U *Desert*
- (7) if u *hungry* nigga why wait?
- (8) I Dont *Eat Vegetables*.

5 Our Annotation Scheme

When we performed qualitative analysis as in Section 4, we found that more noticeable metaphors such as "journey", "pick you up", and "fall" in (1) and (2) seem more indicative of speaker attitude or positioning than metaphors such as "point" in (5). This might suggest the degree of metaphoricity affects how metaphors function in discourse. In this section, we describe our metaphor annotation scheme, which tries to map this variation among metaphors to a simpler three-point scale of nonliteralness: *nonliteral*, *conventionalized*, and *literal*.

5.1 Basic Conditions

Our annotation scheme targets language satisfying the following three conditions:

1. the expression needs to have an original established meaning.
2. the expression needs to be used in context to mean something significantly different from that original meaning.
3. the difference in meaning should not be hyperbole, understatement, sarcasm or metonymy

These conditions result in metaphorical expressions including simile and metaphorical idioms. We consider simile to be a special case of metaphor which makes an explicit comparison using words such as “like”. We include metaphorical idioms because they are obviously nonliteral and metaphorical despite the fact that they have lost their source domains.

Have an original meaning: The expression or the words within the expression need to have original established meanings. For example, in the sentence “I will be joining you on your journey this month” of (1) in Section 4, the word “journey” refers to chemotherapy given the context, but has a clear and commonly known original meaning of a physical journey from one place to another.

Alter the original and established meanings of the words: The usage needs to change the original meaning of the expression in some way. The intended meaning should be understood through a comparison to the original meaning. For the same example, in “I will be joining you on your journey this month”, the intended meaning can be understood through a comparison to some characteristics of a long voyage. For metaphorical idioms such as “he kicked the bucket,” the nonliteral meaning of “he died” is far from the literal meaning of “he struck the bucket with his foot.”

Should not merely be hyperbole, understatement, sarcasm, or metonymy: To reduce the scope of our work, the usage needs to alter the original meaning of the expression but should not simply be a change in the intensity or the polarity of the meaning, nor should it be metonymy. Language uses like hyperbole and understatement may simply change the intensity of the meaning without otherwise altering it. For sarcasm, the intended meaning is simply the negation of the words used. Metonymy is a reference by association rather than a comparison. For example, in

“The White House denied the rumor”, the White House stands in for the president because it is associated with him, rather than because it is being compared to him. Note that metaphorical expressions used in conjunction with these techniques will still be coded as metaphor.

5.2 Decision Steps

To apply the basic conditions to the actual annotation procedure, we come up with a set of decision questions (Table 1). The questions rely on a variety of other syntactic and semantic distinctions serving as filtering questions. An annotator follows the questions in order after picking a phrase or word in a sentence he or she thinks might be nonliteral language. We describe some of our decisions below.

Unit: The text annotators think might be nonliteral is considered for annotation. We allow a word, a phrase, a clause, or a sentence as the unit for annotation as in (Wallington et al., 2003). We request that annotators include as few words as necessary to cover each metaphorical phrase within a sentence.

Category: We request that annotators code a candidate unit as *nonliteral*, *conventionalized*, or *literal*. We intend the *nonliteral* category to include nonliteral language usage within our scope, namely metaphors, similes, and metaphorical idioms. The *conventionalized* category is intended to cover the cases where the nonliteralness of the expression is unclear because of its extensive usage. The *literal* category is assigned to words that are literal without any doubt.

Syntactic forms: We do not include prepositions or light verbs. We do not consider phrases that consist of only function words such as modals, auxiliaries, prepositions/particles or infinitive markers. We restrict the candidate metaphorical expressions to those which contain content words.

Semantic forms: We do not include single compound words, conventional terms of address, greeting or parting phrases, or discourse markers such as “well”. We also do not include terminology or jargon specific to the domain being annotated such as “twilight sedation” in healthcare, since this may be simply borrowing others’ words.

No.	Question	Decision
1	Is the expression using the primary or most concrete meanings of the words?	Yes = L
2	Does the expression include a light verb that can be omitted without changing the meaning, as in “I take a shower” → “I shower”? If so, the light verb expression as a whole is literal.	Yes = L
3	Is the metaphor composed of a single compound word, like “painkiller”, used in its usual meaning?	Yes = L
4	Is the expression a conventional term of address, greeting, parting phrase or a discourse marker?	Yes = L
5	Is the expression using terminology or jargon very common in this domain or medium?	Yes = L
6	Is the expression merely hyperbole/understatement, sarcasm or metonymy?	Yes = L
7	Is the expression a fixed idiom like “kick the bucket” that could have a very different concrete meaning?	Yes = N
8	Is the expression a simile, using “like” or “as” to make a comparison between unlike things?	Yes = N
9	Is the expression unconventional/creative and also using non-concrete meanings?	Yes = N
10	Is there another common way to say it that would convey all the same nuances (emotional, etc.)? Or, is this expression one of the only conventional ways of conveying that meaning?	If yes to the latter = C
11	If you cannot otherwise make a decision between literal and nonliteral, just mark it as C.	

Table 1: Questions to annotate (N: Nonliteral, C: Conventionalized, L: Literal).

6 Experiment

In this section, we present our comparative study of the MTurk annotations and the annotations based on our annotation scheme. The purpose of this experiment is to explore (1) how laypeople perceive metaphor, (2) how valid the annotations from crowdsourcing can be, and (3) how metaphors are different in the three different domains.

6.1 Experiment Setup

We had two annotators who were graduate students with some linguistic knowledge. Both were native speakers of English. The annotators were asked to annotate the data using our annotation scheme. We will call the annotators *trained annotators* from now on.

In addition, we used Amazon’s Mechanical Turk (MTurk) crowdsourcing marketplace to collect laypeople’s recognition of metaphors. We employed MTurk workers to annotate each sentence with the metaphorical expressions. Each

sentence was given along with the full post it came from. MTurkers were instructed to copy and paste all the metaphors appearing in the sentence to given text boxes. They were given a simple definition of metaphor from Wikipedia along with a few examples to guide them. Each sentence was labeled by seven different MTurk workers, and we paid \$0.05 for annotating each sentence. To control annotation quality, we required that all workers have a United States location and have 98% or more of their previous submissions accepted. We monitored the annotation job and manually filtered out annotators who submitted uniform or seemingly random annotations.

6.2 Results

To evaluate the reliability of the annotations, we used weighted Kappa (Cohen, 1968) at the word level, excluding stop words. The weighted Kappa value for annotations following our annotation scheme was 0.52, and the percent agreement was 95.68%. To measure inter-reliability between two annotators per class, we used Cohen’s Kappa (Co-

hen, 1960). Table 2 shows the Kappa values for each dataset and each class. Table 4 shows the corpus statistics.

Dataset	N	C	N+C	Weighted
all	0.44	0.20	0.49	0.52
breastcancer	0.69	0.20	0.63	0.71
Gang	0.26	0.28	0.39	0.34
MOOC	0.41	0.13	0.47	0.53

Table 2: Inter-reliability between two trained annotators for our annotation scheme.

To evaluate the reliability of the annotations by MTurkers, we calculated Fleiss’s kappa (Fleiss, 1971). Fleiss’s kappa is appropriate for assessing inter-reliability when different items are rated by different judges. We measured the agreement at the word level, excluding stop words as in computing the agreement between trained annotators. The annotation was 1 if the MTurker coded a word as a metaphorical use, otherwise the annotation was 0. The Kappa values are listed in Table 3.

Dataset	Fleiss’s Kappa
all	0.36
breastcancer	0.41
Gang	0.35
MOOC	0.30

Table 3: Inter-reliability among MTurkers.

We also measured the agreement between the annotations based on our scheme and MTurk annotations to see how they agree with each other. First, we made a gold standard after discussing the annotations of trained annotators. Then, to combine the seven MTurk annotations, we give a score for an expression 1 if the majority of MTurkers coded it as metaphorically used, otherwise the score is 0. Then, we computed Kappa value between trained annotators and MTurkers. The agreement between trained annotators and MTurkers was 0.51 for N and 0.40 for N + C. We can see the agreement between trained annotators and MTurkers is not that bad especially for N.

Figure 1 shows the percentage of words labeled as N, C or L according to the number of MTurkers who annotated the word as metaphorical. As seen, the more MTurkers who annotated a word,

Dataset	N	N+ C
all	0.51	0.40
breastcancer	0.64	0.47
Gang	0.36	0.39
MOOC	0.65	0.36

Table 5: Inter-reliability between trained annotators and MTurkers.

the more likely it was to be annotated as N or C by our trained annotators. The distinction between Nonliteral and Conventionalized, however, is a bit muddier, although it displays a moderate trend towards more disagreement between MTurkers for the Conventionalized category. The vast majority of words (>90%) were considered to be literal, so the sample size for comparing the N and C categories is small.

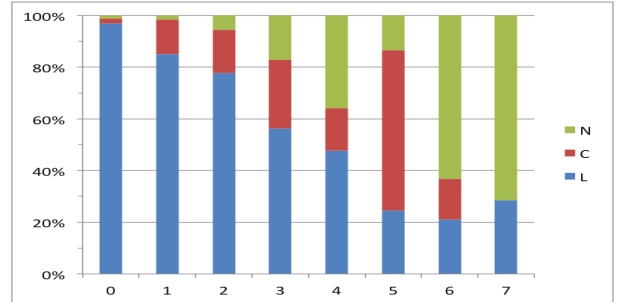


Figure 1: Correspondence between MTurkers and trained annotators. X-axis: the number of MTurkers annotating a word as metaphor.

7 Discussion

In this section, we investigate the disagreements between annotators. A problem inherent to the annotation of metaphor is that the boundary between literal and nonliteral language is fuzzy. Different annotators may draw the line in different places even when it comes to phrases they are all familiar with. It is also true that each person will have a different life history, and so some phrases which are uninteresting to one person will be strikingly metaphorical to another. For example, someone who is unfamiliar with the internet will likely find the phrase “surf the web” quite metaphorical.

Since we did not predefine the words or phrases that annotators could consider, there were often cases where one person would annotate just the

Dataset	Posts	Sent.	Words	Content Words	N	C	N/Sent.	C/Sent.
MOOC	21	100	2005	982	23	59	0.23	0.59
Breastcancer	8	103	1598	797	27	41	0.26	0.4
Gang	44	111	1403	519	30	51	0.27	0.46

Table 4: Data statistics.

noun and another might include the entire noun phrase. If it was part of a conventional multi-word expression, MTurkers seemed likely to include the entire collocation, not merely the metaphorical part. Boundaries were an issue to a lesser extent with our trained annotators.

One of our datasets, the Gang forum, uses a lot of slang and non-standard grammar and spellings. One of our trained annotators is quite familiar with this forum and the other is not. This was the set they had the most disagreement on. For example, the one annotator did not recognize names of certain gangs and rap musicians, and thought they were meant metaphorically. Similarly, the MTurkers had trouble with many of the slang expressions in this data.

Another issue for the MTurkers is the distinction between metaphor and other forms of nonliteral language such as metonymy and hyperbole. For example, in the Gang data, the term “ass” is used to refer to a whole person. This is a type metonymy (synecdoche) using a part to refer to the whole. MTurkers were likely to label such expressions as metaphor. Hyperbolic expressions like “never in a million years” were also marked by some MTurkers.

In a few cases, the sentence may have required more context to decipher, such as previous posts in the same thread. Another minor issue was that some data had words misspelled as other words or grammatical errors, which some MTurkers annotated as metaphors.

Certain categories of conventionalized metaphors that would be annotated in the original presentation of MIP (Pragglejaz-Group, 2007) were never or almost never annotated by MTurkers. These included light verbs such as “make” or “get” when used as causatives or the passive “get”, verbs of sensation used for cognitive meanings, such as “see” meaning “understand”, and demonstratives and prepositions in themselves. This may indicate something about

the relevance of these types of metaphors for certain applications.

8 Conclusion

We annotated data from three distinct conversational online forums using both MTurks and our annotation scheme. The comparison between these two annotations revealed a few things. One is that MTurkers did not show high agreement among themselves, but showed acceptable agreement with trained annotators for the N category. Another is that domain-specific knowledge is important for accurate identification of metaphors. Even trained annotators will have difficulty if they are not familiar with the domain because they may not even understand the meaning of the language used.

Our annotation scheme has room for improvement. For example, we need to distinguish between the Conventionalized and Nonliteral categories more clearly. We will refine the coding scheme further as we work with more annotators.

We also think there may be methods of processing MTurk annotations to improve their correspondence with annotations based on our coding scheme. This could address issues such as inconsistent phrase boundaries or distinguishing between metonymy and metaphor. This could make it possible to use crowdsourcing to annotate the larger amounts of data required for computational applications in a reasonable amount of time.

Our research is in the beginning phase working towards the goal of computational modeling of social and discourse uses of metaphor. Our next steps in that direction will be to work on developing our annotated dataset and then begin to investigate the differing contexts that metaphors are used in. Our eventual goal is to be able to apply computational methods to interpret metaphor at the level of social positioning and discourse functions.

Acknowledgments

This work was supported by NSF grant IIS-1302522, and Army research lab grant W911NF-11-2-0042.

References

- Rodrigo Agerri, John Barnden, Mark Lee, and Alan Wallington. 2007. Metaphor, inference and domain independent mappings. In *Proceedings of RANLP*, pages 17–23. Citeseer.
- John A Barnden and Mark G Lee. 2002. An artificial intelligence approach to metaphor understanding. *Theoria et Historia Scientiarum*, 6(1):399–412.
- Julia Birke and Anoop Sarkar. 2006. A clustering approach for nearly unsupervised recognition of non-literal language. In *EACL*.
- Lynne J Cameron. 2007. Patterns of metaphor use in reconciliation talk. *Discourse & Society*, 18(2):197–222.
- J. Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:37–46.
- Jacob Cohen. 1968. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin*, 70(4):213.
- Dan Fass. 1991. met*: A method for discriminating metonymy and metaphor by computer. *Computational Linguistics*, 17(1):49–90.
- Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378–382.
- Matt Gedigian, John Bryant, Srin Narayanan, and Branimir Ćirić. 2006. Catching metaphors. In *Proceedings of the Third Workshop on Scalable Natural Language Understanding*, pages 41–48. Association for Computational Linguistics.
- Andrew Goatly. 1997. *Language of Metaphors: Literal Metaphorical*. Routledge.
- Beata Beigman Klebanov and Michael Flor. 2013. Argumentation-relevant metaphors in test-taker essays. *Meta4NLP 2013*, pages 11–20.
- Saisuresh Krishnakumaran and Xiaojin Zhu. 2007. Hunting elusive metaphors using lexical resources. In *Proceedings of the Workshop on Computational approaches to Figurative Language*, pages 13–20. Association for Computational Linguistics.
- Linlin Li and Caroline Sporleder. 2010. Using gaussian mixture models to detect figurative language in context. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, HLT ’10, pages 297–300, Stroudsburg, PA, USA. Association for Computational Linguistics.
- James H Martin. 1990. *A computational model of metaphor interpretation*. Academic Press Professional, Inc.
- Zachary J Mason. 2004. Cormet: a computational, corpus-based conventional metaphor extraction system. *Computational Linguistics*, 30(1):23–44.
- Srinivas Narayanan. 1999. Moving right along: A computational model of metaphoric reasoning about events. In *AAAI/IAAI*, pages 121–127.
- Pragglejaz-Group. 2007. Mip: A method for identifying metaphorically used words in discourse. *Metaphor and symbol*, 22(1):1–39.
- Gary M Reisfield and George R Wilson. 2004. Use of metaphor in the discourse on cancer. *Journal of Clinical Oncology*, 22(19):4024–4027.
- SL David Ritchie. 2013. *Metaphor (Key Topics in Semantics and Pragmatics)*. Cambridge university press.
- Ekaterina Shutova and Simone Teufel. 2010. Metaphor corpus annotated for source-target domain mappings. In *LREC*.
- 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.
- Ekaterina Shutova, BarryJ. Devereux, and Anna Korhonen. 2013a. Conceptual metaphor theory meets the data: a corpus-based human annotation study. *Language Resources and Evaluation*, 47(4):1261–1284.
- Ekaterina Shutova, Simone Teufel, and Anna Korhonen. 2013b. Statistical metaphor processing. *Computational Linguistics*, 39(2):301–353.
- 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. Association for Computational Linguistics.
- Ekaterina Shutova. 2013. Metaphor identification as interpretation. *Atlanta, Georgia, USA*, page 276.

Gerard J Steen, Aletta G Dorst, J Berenike Herrmann, Anna Kaal, Tina Krennmayr, and Trijntje Pasma. 2010. *A method for linguistic metaphor identification: From MIP to MIPVU*, volume 14. John Benjamins Publishing.

Anatol Stefanowitsch and Stefan Th Gries. 2006. *Corpus-based approaches to metaphor and metonymy*, volume 171. Walter de Gruyter.

AM Wallington, JA Barnden, P Buchlovsky, L Fellows, and SR Glasbey. 2003. Metaphor annotation: A systematic study. *COGNITIVE SCIENCE RESEARCH PAPERS-UNIVERSITY OF BIRMINGHAM CSRP*.