

# Computing Affect in Metaphors

Tomek Strzalkowski<sup>1,2</sup>, Samira Shaikh<sup>1</sup>, Kit Cho<sup>1</sup>, George Aaron Broadwell<sup>1</sup>, Laurie Feldman<sup>1</sup>, Sarah Taylor<sup>3</sup>, Boris Yamrom<sup>1</sup>, Ting Liu<sup>1</sup>, Ignacio Cases<sup>1</sup>, Yuliya Peshkova<sup>1</sup> and Kyle Elliot<sup>4</sup>

<sup>1</sup>State University of  
New York - Univer-  
sity at Albany

<sup>2</sup>Polish Academy of  
Sciences

<sup>3</sup>Sarah M. Taylor  
Consulting LLC

<sup>4</sup>Plessas Experts  
Network

[tomek@albany.edu](mailto:tomek@albany.edu)

## Abstract

This article describes a novel approach to automated determination of affect associated with metaphorical language. Affect in language is understood to mean the attitude toward a topic that a writer attempts to convey to the reader by using a particular metaphor. This affect, which we will classify as positive, negative or neutral with various degrees of intensity, may arise from the target of the metaphor, from the choice of words used to describe it, or from other elements in its immediate linguistic context. We attempt to capture all these contributing elements in an Affect Calculus and demonstrate experimentally that the resulting method can accurately approximate human judgment. The work reported here is part of a larger effort to develop a highly accurate system for identifying, classifying, and comparing metaphors occurring in large volumes of text across four different languages: English, Spanish, Russian, and Farsi.

## 1 Introduction

We present an approach to identification and validation of affect in linguistic metaphors, i.e., metaphorical expressions occurring in written language. Our method is specifically aimed at isolating the affect conveyed in metaphors as opposed to more broad approaches to sentiment classification in the surrounding text. We demonstrate experimentally that our basic Affect Calculus captures metaphor-related affect with a high degree of accuracy when applied to neutral metaphor targets. These are targets that themselves do not carry any prior valuations. We sub-

sequently expanded and refined this method to properly account for the contribution of the prior affect associated with the target as well as its immediate linguistic context.

## 2 Metaphor in Language

Metaphors are mapping systems that allow the semantics of a familiar Source domain to be applied to a new Target domain so as to invite new frameworks for reasoning (usually by analogy) to emerge in the target domain. The purpose of a metaphor is (a) to simplify or enable reasoning and communication about the target domain that would otherwise be difficult (because of technical complexity) or impossible (due to lack of agreed upon vocabulary) (e.g., Lakoff & Johnson, 1980; 2004); or (b) to frame the target domain in a particular way that enables one form of reasoning while inhibiting another (e.g., Thibodeau & Boroditsky, 2011). The two reasons for using metaphors are not necessarily mutually exclusive, in other words, (a) and (b) can operate at the same time. The distinction suggested above has to do with affect: a metaphor formed through (a) alone is likely to be neutral (e.g., client/server, messenger DNA), while a metaphor formed using (b) is likely to have a polarizing affect (e.g., tax's burden).

The Source and Target domains that serve as endpoints of a metaphoric mapping can be represented in a variety of ways; however, in a nutshell they are composed of two kinds of things: *concepts* and *relations*. In a Target domain the concepts are typically abstract, disembodied, often fuzzy concepts, such as *crime*, *mercy*, or *violence*, but may also include more concrete, novel, or elaborate concepts such as *democracy* or *economic inequality*. In a Source domain, the concepts are typically concrete and physical; however, mapping between two abstract domains is

also possible. (E.g., *crime* may be both a target and a source domain.)

The *relations* of interest are those that operate between the concepts within a Source domain and can be “borrowed” to link concepts within the Target domain, e.g., “Crime<sub>(TARGET)</sub> *spread* to<sub>(RELATION)</sub> previously safe areas” may be borrowing from a DISEASE or a PARASITE source domain.

### 3 Related Research: metaphor detection

Most current research on metaphor falls into three groups: (1) theoretical linguistic approaches (as defined by Lakoff & Johnson, 1980; and their followers) that generally look at metaphors as abstract language constructs with complex semantic properties; (2) quantitative linguistic approaches (e.g., Charteris-Black, 2002; O’Halloran, 2007) that attempt to correlate metaphor semantics with their usage in naturally occurring text but generally lack robust tools to do so; and (3) social science approaches, particularly in psychology and anthropology that seek to explain how people produce and understand metaphors in interaction, but which lack the necessary computational tools to work with anything other than relatively isolated examples.

In computational investigations of metaphor, knowledge-based approaches include MetaBank (Martin, 1994), a large knowledge base of metaphors empirically collected. Krishnakumaran and Zhu (2007) use WordNet (Felbaum, 1998) knowledge to differentiate between metaphors and literal usage. Such approaches entail the existence of lexical resources that may not always be present or satisfactorily robust in different languages. Gedigan et al. (2006) identify a system that can recognize metaphor; however their approach is only shown to work in a narrow domain (*The Wall Street Journal*, for example).

Computational approaches to metaphor (largely AI research) to date have yielded only limited scale, often hand designed systems (Wilks, 1975; Fass, 1991; Martin, 1994; Carbonell, 1980; Feldman & Narayan, 2004; Shutova & Teufel, 2010; inter alia, also Shutova, 2010b for an overview). Baumer et al. (2010) used semantic role labels and typed dependency parsing in an attempt towards computational metaphor identification. However, they describe their own work as an initial exploration and hence, inconclusive. Shutova et al. (2010a) employ an unsupervised method of metaphor identification using nouns and verb clustering to automatically impute met-

aphoricity in a large corpus using an annotated training corpus of metaphors as seeds. Their method relies on annotated training data, which is difficult to produce in large quantities and may not be easily generated in different languages. Several other similar approaches were recently reported at the Meta4NLP<sup>1</sup> workshop, e.g., (Mohler et al., 2013; Wilks et al., 2013; Hovy et al., 2013).

Most recently, a significantly different approach to metaphor understanding based on lexical semantics and discourse analysis was introduced by Strzalkowski et al. (2013). Space constraints limit our discussion about their work in this article, however in the foregoing, our discussion is largely consistent with their framework.

### 4 Affect in Metaphors

Affect in language is understood to mean the attitude toward a topic that a speaker/writer attempts to convey to the reader or audience via text or speech (van der Sluis and Mellish 2008). It is expressed through multiple means, many of which are unrelated to metaphor. While affect in text is often associated, at least in theory, with a variety of basic emotions (anger, fear, etc.), it is generally possible to classify the set of possible affective states by polarity: positive, negative, and sometimes neutral. Affect is also considered to have a graded strength, sometimes referred to as intensity.

Our approach to affect in metaphor has been vetted not only by our core linguistic team but also by an independent team of linguist-analysts with whom we work to understand metaphor across several language-culture groups. Our research continues to show no difficulties in comprehension or disagreement across languages concerning the concept of linguistic affect, of its application to metaphor, and of its having both polarity and intensity.

### 5 Related Research: sentiment and affect

There is a relatively large volume of research on sentiment analysis in language (Kim and Hovy, 2004; Strapparava and Mihalcea, 2007; Wiebe and Cardie, 2005; inter alia) that aim at detecting polarity of text, but is not specifically concerned with metaphors. A number of systems were developed to automatically extract writer’s senti-

---

<sup>1</sup> The First Workshop on Metaphor in NLP.  
<http://aclweb.org/anthology//W/W13/W13-09.pdf>

ment towards specific products or services such as movies or hotels, from online reviews (e.g., Turney, 2002; Pang and Lee, 2008) or social media messages (e.g., Thelwall et al., 2010). None of these techniques has been applied specifically to metaphorical language, and it is unclear if these alone would be sufficient due to the relatively complex semantics involved in metaphor interpretation. Socher et al. (2013 cite) have recently used recursive neural tensor networks to classify sentences into positive/negative categories. However, the presence of largely negative concepts such as “poverty” in a given sentence overwhelms the sentiment for the sentence in their method. Other relevant efforts in sentence level sentiment analysis include Sem-Eval Task<sup>2</sup>. While presence of affect in metaphorical language is well documented in linguistic and psycholinguistic literature (e.g., Osgood, 1980; Pavio and Walsh, 1993; Caffi and Janney, 1994; Steen, 1994), relatively little work was done to detect affect automatically. Some notable recent efforts include Zhang and Barnden (2010), Veale and Li (2012), and Kozareva (2013), who proposed various models of metaphor affect classification based primarily on lexical features of the surrounding text: specifically the word polarity information. In these and other similar approaches, which are closely related to sentiment analysis, affect is attributed to the entire text fragment: a sentence or utterance containing a metaphor, or in some cases the immediate textual context around it.

In contrast, our objective is to isolate affect due to the metaphor itself, independently of its particular context, and also to determine how various elements of the metaphoric expression contribute to its polarity and strength. For example, we may want to know what is the affect conveyed about the Government as a target concept of the metaphor in “*Government regulations are crushing small businesses.*” and how it differs in “*Government programs help to eradicate poverty in rural areas.*” or in “*Feds plan to raise the tax on the rich.*” In all these examples, there is a subtle interplay between the prior affect associated with certain words (e.g., “*crush*”, “*poverty*”) and the semantic role they occupy in the sentence (e.g., agent vs. patient vs. location, etc.). Our objective is to develop an approach that can better explain such differences. Not surprisingly, in one of the target domains we are investigating, the Economic Inequality domain,

there is considerable agreement on the basic attitudes across cultures towards the key target concepts: poverty is negative, wealth is positive, taxation is largely negative, and so on. This is in a marked contrast with another Target domain, the Governance domain where the target concepts tend to be neutral (e.g. bureaucracy, regulations etc.)

Another important motivation in developing our approach (although not discussed in this paper) is to obtain a model of affect that would help to explain empirically why metaphorically rich language is considered highly influential. Persuasion and influence literature (Soppory and Dillard, 2002) indicates messages containing metaphorical language produce somewhat greater attitude change than messages that do not. However, some recent studies (e.g., Broadwell et al., 2012) found that lexical models of affect, sentiment, or emotion in language do not correlate with established measures of influence, contrary to expectations. Therefore, a different approach to affect is needed based both on lexical and semantic features. We describe this new model below, and show some preliminary results in applications to metaphors interpretation.

## 6 Basic Affect Calculus

The need for a new approach to affect arises from the inability of the current methods of sentiment analysis to capture the affect that is conveyed by the metaphor itself, which may be only a part of the overall affect expressed in a text. Affect conveyed in metaphors, while often more polarized than in literal language, is achieved using subtler, less explicit, and more modulated expressions. This presents a challenge for NLP approaches that base affect determination upon the presence of explicit sentiment markers in language that may mask affect arising from a metaphor. This problem becomes more challenging when strong, explicit sentiment markers are present in a surrounding context or when the attitude of the speaker/writer towards the target concept is considered.

Our initial objective is thus to detect and classify the portion of affect that the speaker/writer is trying to convey by choosing a specific metaphor. The observables here are the linguistic metaphors that are actually uttered or written; therefore, our method must be able to determine affect present in the linguistic metaphors first and then extrapolate to the conceptual metaphor based on evidence across multiple uses of the

<sup>2</sup> <https://www.cs.york.ac.uk/semeval-2013/task2/>

same metaphor. Conceptual metaphors are posited by instances of linguistic metaphors that point to the same source domain. We choose initially to model the speaker/writer perspective; however, it may also be important to determine the effect that a metaphor has on the reader/listener, which we do not address here.

Affect in metaphor arises from the juxtaposition of a Source and a Target domain through the relations explicated in linguistic metaphors. These relations typically involve one or more predicates from the source domain that are applied to a target concept. For example, in “*Government regulations are crushing small businesses.*” the relation “*crushing*” is borrowed from a concrete source domain (e.g., Physical Burden), and used with an abstract target concept of “*government regulation*” which becomes the agentive argument, i.e., *crushed*(GovReg, *X*), where *X* is an optional patientive argument, in this case “*small businesses*”. Thus, government regulation is said to be doing something akin to “*crushing*”, a harmful and negative activity according to the Affective Norms in English (ANEW) psycholinguistic database (Bradley and Lang, 1999). Since “*government regulation*” is doing something negative, the polarity of affect conveyed about it is also negative. The ANEW lexicon we are using contains ratings of ~100K words. The original ANEW lexicon by Bradley and Lang was expanded following the work done by Liu et al. (2014) in expanding the MRC imageability lexicon. While other sources of valence judgments exist such as NRC (Mohammad et al., 2013) and MPQA (Weibe and Cardie, 2005), there are limitations – for instance – NRC lexicon rates each word on a positive or negative scale, which does not allow for more fine-grained analysis of strength of valence.

Calculation from Table 1 is further generalized by incorporating the optional second argument of the relation and the role of the target concept (i.e., agentive or patientive). Thus, if *X*=‘small business’ as in the example above, the complete relation becomes *crushed*(GovReg,

*SmBus*), which retains negative affect assuming that ‘small business’ is considered positive or at least neutral, an assessment that needs to be established independently.

The above calculations are captured in the Affect Calculus (AC), which was derived from the sociolinguistic models of topical positioning and disagreement in discourse (Broadwell et al., 2013).

The Affect Calculus was conceived as a hypothetical model of metaphorical affect, involving the metaphor target, the source relation, as well as the arguments of this relation, one of which is the target itself. The basic version of the AC is shown in Table 1. We should note that the AC allows us to make affect inferences about any of the elements of the metaphoric relation given the values of the remaining elements. We should also note that this calculus does not yet incorporate any discernable prior affect that the target concept itself may carry. When the target concept may be considered neutral (as is “*government regulation*” when taken out of context) this table allows us to compute the affect value of any linguistic metaphor containing it. This is unlike the target concepts such as “*poverty*” which bring their prior affect into the metaphor. We will return to this issue later.

In the Affect Calculus table, *Relation* denotes a unary or binary predicate (typically a verb, an adjective, or a noun). In the extended version of the AC (Section 6) *Relation* may also denote a compound consisting of a predicate and one or more *satellite arguments*, i.e., arguments other than AGENT or PATIENT, such as ORIGIN or DESTINATION for *motion verbs*, etc.

## 7 Extended Affect Calculus

The basic Affect Calculus does not incorporate any prior affect that the target concept might bring into a metaphor. This is fine in some domains (e.g., Government), where most target concepts may be considered neutral. But in other target domains, such as the Economic Inequality domain, many of the target concepts have a

Relation type	Type 1 (proper-tive) <i>Rel</i> (Target)	Type 2 (agentive) <i>Rel</i> (Target, <i>X</i> )		Type 3 (patientive) <i>Rel</i> ( <i>X</i> , Target)	
<i>Relation/X</i>		$X \geq neutral$	$X < neutral$	$X \geq neutral$	$X < neutral$
<i>Positive</i>	POSITIVE	POSITIVE	$\leq UNSYMP$	POSITIVE	$\leq SYMPAT$
<i>Negative</i>	NEGATIVE	$\leq UNSYMP$	$\geq SYMPAT$	$\leq SYMPAT$	$\geq SYMPAT$
<i>Neutral</i>	NEUTRAL	NEUTRAL	$\leq NEUTRAL$	NEUTRAL	$\leq NEUTRAL$

Table 1. A simple affect calculus specifies affect polarity for linguistic metaphors using a 5-point polarity scale [negative < unsympathetic < neutral < sympathetic < positive]. *X* is the second argument.

strong prior affect in most cultures (e.g., ‘poverty’ is universally considered negative). We thus need to incorporate this prior affect into our calculation whenever an affect-loaded target concept is invoked in a metaphor. Where the basic Affect Calculus simply imposes a context-borne affect upon a neutral target concept, the Advanced Affect Calculus must combine it with the prior affect carried by the target concept, depending upon the type of semantic context. As already discussed, we differentiate 3 basic semantic contexts (and additional contexts in the extended Affect Calculus discussed in the next section) where the target concept is positioned with respect to other arguments in a metaphorical expression:

- *Propertive* context is when a property of a Target is specified (e.g. *deep poverty*, *sea of wealth*)
- *Agentive* context is when the Target appears as an agent of a relation that may involve another concept (Argument X) in the patient role (e.g. *Government regulations are crushing...*, *Government programs help...*)
- *Patientive* context is when the Target appears in the patient role that involves another concept (possibly implicit, Argument X) in the agent role. (e.g. *...eradicate poverty.*, *....navigate government bureaucracy*)

Table 1 (in the previous section) specifies how to calculate the affect expressed towards the target depending upon the affect associated with the Relation and the Argument X. In the Advanced Affect Calculus, this table specifies the *context-borne affect* that interacts with the affect associated with the target. When the target prior affect is unknown or assumed neutral, the AC table is applied directly, as explained previously. When the target has a known polarized affect, either positive or negative, the values in the AC table are used to calculate the final affect by combining the prior affect of the target with an appropriate value from the table. This is necessary for affect-loaded target concepts such as “poverty” or “wealth” that have strong prior affect and cannot be considered neutral.

In order to calculate the combined affect we define two operators  $\oplus$  and  $\otimes$ . These operators form simple polarity algebra shown in Table 2. When the Target is in a Patientive relation, we use  $\otimes$  to combine its affect with the context value from the AC table; otherwise, we use  $\oplus$ . In the table for  $\oplus$  operator, we note that combining opposing affects from the Target and the Rela-

tion causes the final affect to be undetermined (UND). In such cases we will take the affect of the stronger element (more polarized score) to prevail.

$\otimes$	pos	neg	neu		$\oplus$	pos	neg	neu
pos	pos	neg	pos		pos	pos	UND	pos
neg	neg	pos	neg		neg	UND	neg	neg
neu	pos	neg	neu		neu	pos	neg	neu

Table 2: Polarity algebra for extended affect calculus

More specifically, in order to determine the combined polarity score in these cases, we compute the distance between each element’s ANEW score and the closest boundary of the neutral range of scores. For example, ANEW scores are assigned on a 10-point continuum (derived from human judgments on 10-point Likert scale) from most negative (0) to most positive (9). Values in the range of 3.0 to 5.0 may be considered neutral (this range can be set differently for target concepts and relations):

- Poverty affect score = 1.67 (ANEW) – 3 (neutral lower) = -1.33
- Grasp affect score = 5.45 (ANEW) – 5 (neutral upper) = +0.45

Consider the expression “poverty’s grasp”. Since poverty is a polarized target concept in Propertive position, we use  $\oplus$  operator to combine its affect value with that of Relation (*grasp*). The result is negative:

- “Poverty’s grasp” affect score (via  $AC \oplus$ ) = -1.33 + 0.45 = -0.82 (negative)

When the combined score is close to 0 (-0.5 to +0.5) the final affect is neutral.

## 7.1 Exceptions

The above calculus works in a majority of cases, but there are exceptions requiring specialized handling. An incomplete list of these is below (and cases will be added as we encounter them):

Reflexive relations. In some cases the target is in the agentive position but semantically it is also a patient, as in “poverty is spreading”. These cases need to be handled carefully – although the current AC may be able to handle them in some contexts. When interpreted as an agentive rela-

tion, the affect of “poverty is spreading” comes out as undetermined but would likely be output as negative on the basis of the strong negative affect associated with poverty (vs. weaker positive affect of “spreading”). When handled as a patientive relation (an unknown force is spreading poverty), it comes out clearly and strongly negative. Similarly, “wealth is declining” is best handled through patientive relation. Therefore, for this AC we will treat intransitive relations as patientive.

Causative relations. Some relations denoted by causative verbs such as “alleviate”, “mitigate” or “ease” appear to presuppose that their patient argument has negative affect, and their positive polarity already incorporates this assumption. Thus, “alleviate” is best interpreted as “reduce the negative of”, which inserts an extra negation into the calculation. Without considering this extra negation we would calculate “alleviate(+) poverty(-)” as negative (doing something positive to a negative concept), which is not the expected reading. Therefore, the proposed special handling is to treat “alleviate” and similar relations as always producing positive affect when applied to negative targets.

## 8 Extensions to Basic Affect Calculus

The basic model presented in the preceding section oversimplifies certain more complex cases where the metaphoric relation involves more than 2 arguments. Consequently, we are considering several extensions to the basic Affect Calculus as suggested below. The foregoing should be treated as hypotheses subject to validation.

One possible extension involves relations represented by verbs of motion (which is a common source domain) that involve satellite arguments such as ORIGIN and DESTINATION in addition to the main AGENT and PATIENT roles. Any polarity associated with these arguments may impact affect directed at the target concept appearing in one of the main role positions. Likewise, we need a mechanism to calculate affect for target concepts found in one of the satellite roles. In “*Federal cuts could push millions into poverty*” the relation ‘push into’ involves three arguments: AGENT (*Federal cuts*), PATIENT (*millions [people]*) and DESTINATION (*poverty*). In calculating affect towards ‘Federal cuts’ it is not sufficient to consider the polarity of the predicate “push” (or “push into”), but instead one must consider the polarity of “push into (*poverty*)” as the composite agentive relation involving ‘federal cuts’.

The polarity of this composite, in turn, depends upon the polarity of its destination argument. In other words:

$$polarity(Rel(DEST)) = polarity(DEST)$$

Thus, if ‘poverty’ is negative, then pushing someone or something into poverty is a harmful relation. Assuming that ‘*millions [people]*’ is considered at least neutral, we obtain negative affect for ‘*Federal cuts*’ from the basic Affect Calculus table.

An analogous situation holds for the ‘ORIGIN’ argument, with the polarity reversed. Thus:

$$polarity(Rel(ORIGIN)) = \sim polarity(ORIGIN)$$

In other words, the act of removing something from a bad place is helpful and positive. For example, in “*Higher retail wages would lift Americans out of poverty*” the relation compound “lift out of (*poverty*)” is considered helpful/positive. Again, once the polarity of the relation compound is established, the basic affect calculus applies as usual, thus we obtain positive affect towards ‘*higher retail wages*’. In situations when both arguments are present at the same time and point towards potentially conflicting outcomes, we shall establish a precedence order based on the evidence from human validation data.

Another class of multi-argument relations we are considering includes verbs that take an INSTRUMENT argument, typically signaled by ‘with’ preposition. In this case, affect inference for the relation compound is postulated as follows:

$$\begin{aligned} polarity(Rel(INSTR)) \\ &= polarity(INSTR) \text{ if } polarity(INSTR) < neutral \\ &= polarity(Rel) \text{ otherwise} \end{aligned}$$

In other words, using a negative (bad) instrument always makes the relation harmful, while using a positive or neutral instrument has no effect on the base predicate polarity.

Other types of multi-argument relations may require similar treatment, and we are currently investigating further possible extensions. In all cases not explicitly covered in the extended Affect Calculus, we shall assume the default condition that other satellite arguments (such as TIME, LOCATION, etc.) will have no impact on the polarity of the source relation compound. In other words:

$$polarity(Rel(s-role)) =_{\text{default}} polarity(Rel)$$

## 9 Evaluation and Results

For an evaluation, our objective is to construct a test that can evaluate the ability of an automated system to correctly identify and classify the af-

fect associated with linguistic and conceptual metaphors. A series of naturally occurring text samples containing a linguistic metaphor about a target concept are presented as input to the system. The system outputs the affect associated with the metaphor, as positive, negative, or neutral. The system output is then compared to human generated answer key resulting in an accuracy score. The evaluation thus consists of two components:

1. Determining the ground truth about affect in test samples;
2. Measuring the automated system’s ability to identify affect correctly.

Step 1 is done using human assessors who judge affect in a series of test samples. Assessors are presented with brief passages where a target concept and a relation are highlighted. They are asked to rank their responses on a 7-point scale for the following questions, among others:

- *To what degree does the above passage use metaphor to describe the highlighted concept?*
- *To what degree does this passage convey an idea that is either positive or negative?*

It is strictly necessary that input to the system be metaphorical sentences, since affect may be associated with non-metaphoric expressions as well; in fact, some direct expressions may carry stronger affect than subtle and indirect metaphors. This is why both questions on the survey are necessary: the first focuses the assessor’s attention on the highlighted metaphor before asking about affect. If the purpose of the test is to measure the accuracy of assigning affect to a metaphor, then accuracy should be measured against the subset of expressions judged to be metaphorical.

The judgments collected from human assessors are tested for reliability and validity. Reliability among the raters is computed by measuring intra-class correlation (ICC) (McGraw & Wong, 1996; Shrout & Fleiss, 1979). Typically, a coefficient value above 0.7 indicates strong agreement. In general, our analyses have shown that we need approximately 30 or more subjects in order to obtain a reliability coefficient of at least 0.7. In addition, certain precautions were taken to ensure quality control in the data. We used the following criteria to discard a subject’s data: (1) completed the task too quickly (i.e., averaged fewer than 10 seconds for each passage); (2) gave the same answer to 85% or more of the test items; (3) did not pass a simple language proficiency test; or (4) did not provide correct answers to a set of randomly inserted control pas-

sages which have been previously judged by experts to be unequivocally literal or metaphorical. Human judgments are collected using Amazon’s Mechanical Turk services. For each passage in surveys, we would collect at least 30 viable judgments. In addition, we have native language speakers who have been rigorously trained to provide expert judgments on metaphor and affect identification task. Table 3 shows the intra-class correlations for affect determination amongst Mechanical Turk subjects. Experiments were conducted in 4 languages: English, Spanish, Russian, and Farsi.

	English	Spanish	Russian	Farsi
Metaphor	0.864	0.853	0.916	0.720
Affect	0.924	0.791	0.713	0.797

Table 3: Intra-class correlations for metaphor and affect assessment by Mechanical Turk subjects

In Figure 1, we present partial evidence that the human assessment collection method captures the phenomenon of affect associated with metaphors. The chart clearly shows that affect tends to be more polarized in metaphors than in literal expressions. The chart is based on more than 11,000 affect judgments for English linguistic metaphors and literal expressions about Governance concepts. We see a highly pronounced tendency towards the polarization of affect (both positive and negative). Ratings of affect (y-axis) in metaphoric expressions (columns 5-7) are judged to be stronger, and in particular more negative than the literal expressions (columns 1-3). A similar trend occurs with other target concepts as well as other languages, although the data are less reliable due to smaller test samples. Once an answer key is established using the aforementioned procedures, system accuracy can be determined from a confusion matrix as shown in Table 4. In Table 4, we show system assignment of affect versus answer key for English Governance and Economic Inequality target metaphors. Overall accuracy across positive, negative and neutral affect for English test set of 220 samples is 74.5%. Analogous confusion matrices have been constructed for Spanish, Russian and Farsi. NLP resources such as parser and lexicons for the languages other than English are not as robust or well rounded; therefore affect classification accuracy in those languages is impacted.

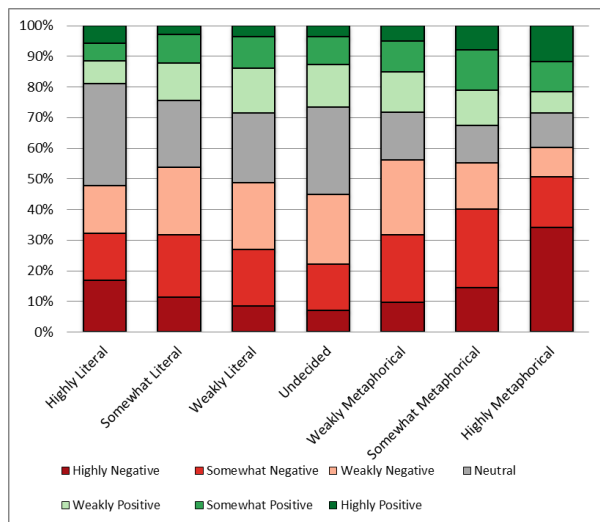


Figure 1: Distribution of affect polarity in human judgment of English literal and metaphorical expressions from the Governance domain. Metaphoricity of an expression (x-axis) is judged from highly literal (1) to highly metaphorical (7)

Table 5 shows the accuracy of affect detection for expressions that the system determined to be metaphors across all four languages under investigation. Evaluation set for numbers reported in Table 5 contains a total of 526 linguistic metaphors in these four languages.

English Affect Sample size = 220		System identified as		
		Positive	Negative	Neutral
Answer Key	Positive	40	16	3
	Negative	12	109	1
	Neutral	10	14	15

Table 4: Confusion matrix for affect classification in English linguistic metaphors in Governance and Economic Inequality Domain. Accuracy is 74.5%

	English	Spanish	Russian	Farsi
Accuracy	74.5%	71%	59%	64%

Table 5: Performance on affect classification for linguistic metaphors in four languages

## 10 Conclusion

In this paper we presented a new approach to automatic computing of affect in metaphors that exploits both lexical and semantic information in metaphorical expressions. Our method was evaluated through a series of rigorous experiments

where more than several dozen of qualified assessors judged hundreds of sentences (extracted from online sources) that contained metaphorical expressions. The objective was to capture affect associated with the metaphor itself. Our system can approximate human judgment with accuracy ranging from 59% for Russian to 74% for English. These results are quite promising. The differences are primarily due to varied robustness of the language processing tools (such as parsers and morphological analyzers) that are available for each language. We note that a direct comparison to lexical approaches such as described by Kozareva (2013) is not possible at this time due to differences in assessment methodology, although it remains one of our objectives.

Our next step is to demonstrate that the new way of calculating affect can lead to a reliable model of affective language use that correlates with other established measures of influence.

## Acknowledgements

Supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Defense US Army Research Laboratory contract number W911NF-12-C-0024. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoD/ARL, or the U.S. Government.

## References

- David W. Allbritton, Gail McKoon, and Richard J. Gerrig. 1995. Metaphor-based schemas and text Representations: making connections through conceptual metaphors, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(3):612-625.
- Eric P. S. Baumer, James P. White, and Bill Tomlinson. 2010. Comparing semantic role labeling with typed dependency parsing in computational metaphor identification. In *Proceedings of the NAACL HLT 2010 Second Workshop on Computational Approaches to Linguistic Creativity*, pages 14–22, Los Angeles, California.
- Margaret M. Bradley, and Peter Lang. 1999. Affective norms for English words (ANEW): Instruction manual and affective ratings. Technical Report C-2. University of Florida, Gainesville, FL.
- George Aaron Broadwell, Umit Boz, Ignacio Cases, Tomek Strzalkowski, Laurie Feldman, Sarah Tay-



- lor, Samira Shaikh, Ting Liu, Kit Cho, and Nick Webb. 2013. Using imageability and topic chaining to locate metaphors in linguistic corpora. In *Proceedings of International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction*, pages 102–109. Washington D.C.
- George Aaron Broadwell, Jennifer Stromer-Galley, Tomek Strzalkowski, Samira Shaikh, Sarah Taylor, Umit Boz, Alana Elia, Laura Jiao, Ting Liu and Nick Webb. 2012. Modeling socio-cultural phenomena in discourse. *Journal of Natural Language Engineering*, pages 1–45. Cambridge Press.
- Claudia Caffi, and Richard W. Janney. 1994. Towards a pragmatics of emotive communication. *Journal of Pragmatics*, 22:325–373.
- Jaime Carbonell. 1980. Metaphor: A key to extensible semantic analysis. In *Proceedings of the 18<sup>th</sup> Annual Meeting on Association for Computational Linguistics*.
- Jonathan, Charteris-Black. 2002. Second language figurative proficiency: A comparative study of Malay and English. *Applied Linguistics* 23(1):104–133.
- Dan, Fass. 1991. met\*: A Method for Discriminating Metonymy and Metaphor by Computer. *Computational Linguistics*, 17:49-90
- Jerome Feldman, and Srinivas Narayanan. 2004. Embodied meaning in a neural theory of language. *Brain and Language*, 89(2):385–392.
- Christiane D. Fellbaum. 1998. *WordNet: An electronic lexical database* (1<sup>st</sup> ed.). MIT Press.
- Matt Gedigian, John Bryant, Srini Narayanan and Branimir Ciric. 2006. Catching Metaphors. In *Proceedings of the Third Workshop on Scalable Natural Language Understanding ScaNaLU 2006*, pages 41–48. New York City: NY.
- Dirk Hovy, Shashank Shrivastava, Sujay Kumar Jauhar, Mrinmaya Sachan, Kartik Goyal, Huying Li, Whitney Sanders and Eduard Hovy. 2013. Identifying Metaphorical Word Use with Tree Kernels. In the *Proceedings of the First Workshop on Metaphor in NLP, (NAACL)*. Atlanta.
- Soo-Min Kim and Eduard Hovy. 2004. Determining the sentiment of opinions. In *Proceedings of the 20th international conference on Computational Linguistics*, COLING '04.
- 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 (ACL 2013)*
- 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. Rochester, NY.
- George Lakoff, and Mark Johnson. 1980. *Metaphors we live by*. University Of Chicago Press, Chicago, Illinois.
- George, Lakoff. 2001. *Moral politics: what conservatives know that liberals don't*. University of Chicago Press, Chicago, Illinois.
- Ting Liu, Kit Cho, George Aaron Broadwell, Samira Shaikh, Tomek Strzalkowski, John Lien, Sarah Taylor, Laurie Feldman, Boris Yamrom, Nick Webb, Umit Boz and Ignacio Cases. 2014. Automatic Expansion of the MRC Psycholinguistic Database Imageability Ratings. In *Proceedings of 9<sup>th</sup> Language Resources and Evaluation Conference, (LREC 2014)*Reykjavik, Iceland.
- Liisa, Malkki. 1992. National geographic: The rooting of people and the territorialization of national identity among scholars and refugees. *Society for Cultural Anthropology*, 7(1):24–44.
- James Martin. 1988. A computational theory of metaphor. *Ph.D. Dissertation*.
- Kenneth O. McGraw and S. P. Wong. 1996. Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1): 30–46.
- Mohammad, S.M., S. Kiritchenko, and X. Zhu. 2013. NRC-Canada: Building the state-of-the-art insentiment analysis of tweets. In *Proceedings of the Seventh International Workshop on Semantic Evaluation Exercises (SemEval-2013)*, Atlanta, Georgia, USA, June 2013.
- Michael Mohler, David Bracewell, David Hinote, and Marc Tomlinson. 2013. Semantic signatures for example-based linguistic metaphor detection. In *The Proceedings of the First Workshop on Metaphor in NLP, (NAACL)*, pages 46–54.
- Musolff, Andreas. 2008. What can critical metaphor analysis add to the understanding of racist ideology? Recent studies of Hitler's anti-semitic metaphors, critical approaches to discourse analysis across disciplines. *Critical Approaches to Discourse Analysis Across Disciplines*, 2(2):1–10.
- Kieran, O'Halloran. 2007. Critical discourse analysis and the corpus-informed interpretation of metaphor at the register level. *Oxford University Press*
- Charles E. Osgood. 1981. The cognitive dynamics of synaesthesia and metaphor. In *Proceedings of the National Symposium for Research in Art. Learning in Art: Representation and Metaphor*, pages 56-80. University of Illinois Press.
- Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135, January.
- Allan Pavio and Mary Walsh. 1993. Psychological processes in metaphor comprehension and memory. In Andrew Ortony, editor, *Metaphor and thought* (2<sup>nd</sup> ed.). Cambridge: Cambridge University Press.
- Patrick E Shrout and Joseph L Fleiss. 1979. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2):420–428.
- Ekaterina Shutova. 2010. Models of metaphors in NLP. In *Proceedings of ACL 2010. Uppsala, Sweden*.

- Ekaterina Shutova and Simone Teufel. 2010a. Metaphor corpus annotated for source - target domain mappings. In *Proceedings of Language Resources and Evaluation Conference 2010*. Malta.
- Ekaterina Shutova. 2010b. Models of metaphor in nlp. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ACL '10, pages 688–697.
- Ekaterina Shutova, Tim Van de Cruys, and Anna Korhonen. 2012. *Unsupervised metaphor paraphrasing using a vector space model* In *Proceedings of COLING 2012*, Mumbai, India
- Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Chris Manning, Andrew Ng and Chris Potts. 2013. In *Proceedings Conference on Empirical Methods in Natural Language Processing (EMNLP 2013)*. Seattle, USA.
- Sopory, P. and Dillard, J. P. (2002), The Persuasive Effects of Metaphor: A Meta-Analysis. *Human Communication Research*, 28: 382–419. doi: 10.1111/j.1468-2958.2002.tb00813.x
- Gerard Steen. 1994. Understanding metaphor in literature: An empirical approach. London: Longman.
- Carlo, Strapparava, and Rada Mihalcea. 2007. Semeval-2007 task 14: Affective text. In *Proceedings of the Fourth International Workshop on Semantic Evaluations*, pages 70–74. Association for Computational Linguistics.
- Tomek Strzalkowski, George Aaron Broadwell, Sarah Taylor, Laurie Feldman, Boris Yamrom, Samira Shaikh, Ting Liu, Kit Cho, Umit Boz, Ignacio Cases and Kyle Elliott. 2013. Robust extraction of metaphor from novel data. In *Proceedings of Workshop on Metaphor in NLP, NAACL*. Atlanta.
- Mike Thelwall, Kevan Buckley, and Georgios Patooglou. Sentiment in Twitter events. 2011. *Journal of the American Society for Information Science and Technology*, 62(2):406–418.
- Paul H. Thibodeau and Lera Boroditsky. 2011. Metaphors We Think With: The Role of Metaphor in Reasoning. *PLoS ONE* 6(2): e16782.
- Peter D, Turney. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, pages 417–424.
- Ielka van der Sluis, and C. Mellish 2008. Toward affective natural language generation: Empirical investigations. affective language in human and machine. AISB 2008 Proceedings Volume 2.
- Tony Veale and Guofu Li. 2012. Specifying viewpoint and information need with affective metaphors: a system demonstration of the metaphor magnet web app/service. In *Proceedings of the ACL 2012 System Demonstrations*, ACL '12, pages 7–12.
- Janyce, Wiebe and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. In *Language Resources and Evaluation*.
- Yorick, Wilks. 1975. Preference semantics. *Formal Semantics of Natural Language*, E. L. Keenan, Ed. Cambridge University Press, Cambridge, U.K., 329–348.
- Yorick Wilks, Lucian Galescu, James Allen, Adam Dalton. 2013. Automatic Metaphor Detection using Large-Scale Lexical Resources and Conventional Metaphor Extraction. In the *Proceedings of the First Workshop on Metaphor in NLP, (NAACL)*. Atlanta.
- Wiebe, J., Wilson, T., and Cardie, C.: Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2-3), pp. 165-210 (2005).
- Li Zhang and John Barnden. 2010. Affect and metaphor sensing in virtual drama. *International Journal of Computer Games Technology*. Vol. 2010.