# COMPUTATIONAL METAPHOR IDENTIFICATION: A METHOD FOR IDENTIFYING CONCEPTUAL METAPHORS IN WRITTEN TEXT


**Eric Baumer**[1]
ebaumer@ics.uci.edu
**Bill Tomlinson**[1]
wmt@uci.edu
**Lindsey Richland**[2]
lerich@uci.edu
[1] Department of Informatics, Donald Bren School of Information and Computer Sciences, 5029 Bren Hall
[2] Department of Education, Berkeley Place 2046
University of California, Irvine
Irvine, CA 92697



## ABSTRACT

While much research has pursued the development of computational models of metaphorical and analogical reasoning in controlled contexts, less work has been conducted on model capturing less structured, everyday relational reasoning. This paper describes computational metaphor identification (CMI), which uses computational linguistic techniques to identify patterns in written text indicative of potential conceptual metaphors. The paper presents an overview of CMI, followed by sample results from two different corpora: middle school students' science writing and political blogs from during the 2008 US election. These results demonstrate CMI's capacity to identify linguistic patterns potentially indicative of deep conceptual metaphors that could subtly yet powerfully influence reasoning.


## Introduction

In order to understand the processes of human analogical reasoning, researchers have created numerous models and processual accounts thereof, e.g., (Gentner, 1983; Holyoak & Thagard, 1989), some of which have been implemented in computational systems, e.g., (Falkenhainer, Forbus, & Gentner, 1989; Hall, 1989; Hummel & Holyoak, 2003). Most such models take as input encoded knowledge about two or more distinct representations, and generate as output analogical correspondences mapping attributes of the source onto the target (though see Doumas, Hummel, & Sandhofer, 2008). However, as noted previously (Falkenhainer et al., 1989), such an approach requires encoding knowledge into a representation amenable to such processing. It may be beneficial to explore alternatives based on deriving such mappings by analyzing the form in which many human ideas are conveyed: written natural-language.

In a separate body of research, computational linguists have grappled with issues related to metaphor and analogy, e.g., (Fass, 1991; Lu & Feldman, 2007; Martin, 1990), specifically as related to text, but from a slightly different perspective. Rather than attempting to determine what analogical mappings or processes might be at work, these techniques have been designed for natural language processing (NLP) systems, such as extracting and codifying knowledge from large bodies of text. In this context, the goal is often to determine whether a given phrase is literal or figurative, and then to perform special additional processing on figurative text to determine its literal meaning. Such systems generally approach metaphorical language on a case-by-case basis, determining if each individual phrase they encounter is literal or figurative, rather than searching for overarching patterns that may indicate underlying conceptual mappings.

This paper represents a departure from previous approaches to metaphorical language and relational reasoning. Specifically, we describe computational metaphor identification (CMI), a technique for identifying linguistic patterns in large textual corpora potentially indicative of conceptual metaphors. This approach differs from previous computational linguistic work in that rather than attempting to discern whether individual phrases are metaphorical or literal, it focuses on overall linguistic patterns that suffuse a body of text. It also differs from most previous approaches to modeling relational reasoning in that, rather than using abstract knowledge representations constructed by a researcher, this approach focuses solely on mapping correspondences between linguistic patterns within different corpora of written natural language.

This work also represents an important new direction for AI research. The goal of CMI is not to state definitively the metaphors that are present in a corpus, but rather to identify for consideration potential metaphors, encouraging critical reflection on what identified linguistic patterns might imply. This approach is something of an inversion of more traditional artificial intelligence research. Classical AI is often associated with asking, "Can people make computers think?" That is, can we develop computational systems that behave in a manner we might call intelligent. Instead, the work described here asks, "Can computers make people think?" That is, can we design and implement computational technologies that encourage human users to think in new, different ways or to approach familiar concepts and ideas from novel, alternative perspectives. Thus, from this perspective, CMI should be evaluated not on its accuracy at identifying cross-domain mappings, but rather on its ability to find patterns of language that can promote critical thinking and creativity.

## RELATED WORK

Relational reasoning is an integral part of humans' ability to interact generatively and creatively in the world, and enables reasoners to draw inferences from one domain, or representation, to better conceptualize another (Penn, Holyoak, & Povinelli, 2008). The work described here draws on previous linguistics work on conceptual metaphor and computational linguistics approaches to develop a new system for analyzing cross-domain relational reasoning in everyday thought. This section briefly reviews the theory of conceptual metaphor and describes a prior attempt to study them computationally as a foundation to the current system.

### *Conceptual Metaphor*

The work presented in this paper is informed largely by the Lakoffian perspective of conceptual metaphor (Lakoff, 1993; Lakoff & Johnson, 1980) and views metaphor not as a literary or poetic device, but rather as a fundamental aspect of human cognition. For example, when discussing money, one might say "he *poured* money *into* his savings account," "they *froze* my assets," or "capital freely *flowed* between investors." Lakoff and Johnson (1980) claim that such linguistic patterns evidence the conceptual metaphor MONEY IS A LIQUID[1], that we understand the abstract concept of money in terms of our concrete physical experiences with liquids. We use words from our knowledge of liquids to talk about money because the cognitive structure of the metaphor "sanctions the use of source domain language and inference patterns for the target domain" (Lakoff, 1993, p. 208). This is not to say that conceptual metaphor is a primarily a linguistic phenomenon. Rather, the linguistic patterns serve as evidence for the cognitive phenomenon.

1 This paper uses SMALL CAPS for metaphors, *italics* for concepts, ALL CAPS for domains, and "quotes" for example phrases.

One important aspect of conceptual metaphor theory is that many different metaphors may be used to frame the same concept, a phenomenon referred to as metaphorical pluralism. For example, a cornucopia of metaphors can be used for the concept of *love*, such as LOVE IS A JOURNEY: “this relationship is[n't] *going anywhere*;” LOVE IS MADNESS: “I'm just *wild* about Harry;” or LOVE IS MAGIC: “she is *bewitching*” (Lakoff & Johnson, 1980, pp. 44,49). Each metaphor simultaneously highlights certain aspects of a situation while downplaying others. Lakoff and Johnson argue that “successful functioning in our daily lives seems to require a constant shifting of [many] metaphors ... that are inconsistent with one another ... to comprehend details of our daily existence” (Lakoff & Johnson, 1980, p. 221). Moreover, suggestion of an alternative, novel metaphor can provide a reconceptualization that draws attention to different aspects of the situation, can “cause us to try to understand how [the novel metaphor] could be true, [and] makes possible a new understanding of our lives” (Lakoff & Johnson, 1980, p. 175). The purpose of CMI, then, is to promote such critical reflection by identifying particular linguistic patterns and presenting a reader with the metaphors those patterns might imply.

### *Metaphor in Computational Linguistics*

As described above, most previous computational linguistics research on metaphor has focused on discerning figurative from literal phrases and then determining the literal meaning of figurative language (Fass, 1991; Lu & Feldman, 2007; Martin, 1990). For example, one application, designed as a help system for the UNIX command line, would determine that statements such as “how do I get into Lisp?” meant the user wanted to know how to invoke Lisp programming environment (Martin, 1990). One exception to this trend is CorMet (Mason, 2004), a system designed to extract known conceptual metaphors from domain-specific textual corpora. For example, CorMet was used to extract the metaphor MONEY IS A LIQUID by using documents acquired from web searches about the LABORATORY domain and the FINANCE domain, based on the way that verbs such as “pour,” “flow,” “freeze,” and “evaporate” are associated with words for *liquid* in the LABORATORY corpus and with words for *money* in the FINANCE corpus. The technique presented here draws largely on CorMet but differs in two important ways. First, CorMet was designed to extract known conventional metaphors, whereas this work involves identifying potential metaphors in arbitrary corpora. Second, little work has explored using such computationally identified metaphors to promote critical thinking about, and creative generation of, metaphors.

## COMPUTATIONAL METAPHOR IDENTIFICATION

While space limits preclude a fully detailed description, this section provides an overview of computational metaphor identification (CMI), a technique for identifying linguistic patterns in textual corpora indicative of potential conceptual metaphors. This work draws on and extends previous computational linguistics research (Mason, 2004).

Since metaphors map from a source domain to a target domain, CMI begins by gathering corpora for the sources and target, where the target corpus is the text in which to identify metaphors, and the source corpora represent source domains for the metaphors one might wish to identify. The researcher selects the source domain based on a theoretical prediction that this may be a likely source for the identified target domain. For example, the domain LIQUID might be identified as a potential source for the target concept *money*. This approach treats domain representations as less formally structured than the structure-mapping (Gentner, 1983) or multi-constraint satisfaction (Holyoak & Thagard, 1989) models of analogy. This flexibility allows for consideration of more dispersed, subtle infiltration of cross-domain mappings into everyday thought.

This paper describes results from analyzing two target corpora: a collection of middle-school students' writing about cellular biology, and a set of political blogs. The implementation described here uses Wikipedia articles for source corpora, as they provide a readily available, categorically sorted, large source of text on a wide variety of topics. A source corpus for a given domain consists of all the Wikipedia articles in the category for that domain, as well as all articles in its subcategory. All documents in the source and target corpora are parsed to extract grammatical relationships (de Marneffe, MacCartney, & Manning, 2006).

The crux of CMI is selectional preference learning (Resnik, 1993), which identifies association between different classes of words through specific grammatical relationships. For example, words for the concept of *food* are often the direct object of the verb "eat." Selectional preferences are often calculated in terms of verbs' preferences for nouns, but they can just as readily be calculated in terms of nouns' preferences for verbs (Light & Greiff, 2002). Using the parsed documents, CMI calculates selectional preferences of the characteristic nouns in a corpus, where characteristic means that the noun is highly frequent in the corpus relative to its frequency in general English. Selectional preference is quantified as the relative entropy of the posterior distribution conditioned on a specific noun and grammatical case slot with respect to the prior distribution of verbs in general English:

$$S(c) = \sum_{v} P(v|c) \log \frac{P(v|c)}{P(v)}$$

where $c$ is a class of nouns (e.g., nouns representing the concept *food)* and a case slot, and $v$ ranges over all the verbs for which $c$ appears in the given case slot. Selectional preference captures the "choosiness" of a particular grammatical relationship; selectional association measures the degree to which that grammatical relationship is associated with a particular verb:

$$\lambda(c,v) = \frac{1}{S(c)} P(v|c) \log \frac{P(v|c)}{P(v)}$$

While selectional associations are calculated for classes of words, corpora consist of words that may represent many possible classes of nouns. Thus, individual nouns count as partial observations of each class of words that they might represent, using the ontological dictionary WordNet (Fellbaum, 1998). For example, the words "water," "liquor," and "ammonia" can all represent the concept of *liquid*, as *liquid* is a parent, or hypernym, of each in the WordNet hierarchy. WordNet uses synsets (sets of synonyms) to represent classes of words. For example, the synonyms "liquid," "liquidness," "liquidity," and "liquid state" comprise the synset for the liquid state of matter. These synsets are then clustered using two-nearest-neighbor clustering based on the verbs for which they select. Each cluster represents a coherent concept in the corpus.

This approach of clustering hypernyms resonates with Lakoff's argument that metaphorical mappings occur not at the level of situational specifics, but at the superordinate level. For example, in the metaphor LOVE IS A JOURNEY, the relationship between the lovers is the *vehicle* in which they travel. Although specific instantiations of the metaphor may frame that vehicle variously as a train ("off the track"), a car ("long, bumpy road"), or a plane ("just taking off"), "the categories mapped will tend to be at the superordinate level rather than the basic level" (Lakoff, 1993, p. 212). The method described here causes observations at the basic level to accumulate in the superordinate levels they collectively represent.

To identify metaphors, CMI looks for selectional correspondences between conceptual clusters in the source and target corpora. For example, in the LIQUID domain, the cluster for the concept *container* would have strong selectional associations for the object of the preposition "into" with the verb "pour," the object of the preposition "from" with the verb "flow," the subject of the verb "hold," and so on. In documents about banking or finance, the cluster for *institution* also selects for those same verbs in the same grammatical relationships. Based on the degree of correspondence between those selectional associations, each cluster-to-cluster mapping is given a confidence score indicating how likely the linguistic patterns are to evidence a conceptual metaphor. One of CMI's strengths is that it works in the aggregate. While individual phrases such as “flowed from the Federal Reserve” and “poured money into my IRA” may not at first glance appear metaphorical, it is the systematicity of these patterns that becomes compelling evidence for the existence of a metaphor.

This mapping of selectional associations is reminiscent of the selectional restrictions violations view of metaphor (cf. Fass, 1991), also called the “anomaly view” (Tourangeau & Sternberg, 1982), wherein a metaphor can be identified when semantic constraint expectations are violated. For example, in the phrase “my car drinks gasoline,” cars do not usually drink, and gasoline is not usually drunk; thus, two selectional restriction violations indicate a potential metaphor. However, as Ortony (Ortony, 1980) and others have pointed out, such violations are highly contextually dependent. Thus, rather than attempting to derive general-purpose selectional associations, CMI maps selectional associations from a specific source corpus, the original context of use, to a specific target corpus, in which constraints derived from the original context of use may be seen as violated.

An important aspect of CMI is that it identifies only linguistic patterns potentially indicative of conceptual metaphors, not the metaphors themselves. As Lakoff (1993) emphasizes, metaphor is primarily a cognitive phenomenon, and metaphorical language serves as evidence for the cognitive phenomenon. CMI leverages computational power to search through large bodies of text in order to identify patterns of potential interest, then presents those patterns to a human user along with the potential metaphors they might imply. This approach places the job of finding patterns in the hands of the computer, and the job of interpreting those patterns in the hands of the human user.

## SAMPLE RESULTS

This section presents sample portions of results from applying CMI to two different corpora. The first is a collection of science students' written answers to questions about cellular reproduction. The second is a set of posts from political blogs during the 2008 US election. The results from these two rather different corpora demonstrate CMI's ability to identify potential metaphors in a wide variety of content. Both corpora are part of larger research projects that use CMI to foster critical thinking and reflection.

### *Science Students' Writing*

The first corpus comes from 7$^{th}$ grade (ages 12-13) science students' written answers to questions about cellular reproduction. Open-ended questions were asked before and during the model, such as: “What are some differences between mitosis and meiosis?”; or, “Do you think offspring of ALL organisms are always different from their parents? Why or why not?.” Answers to these questions were collected and analyzed using CMI to determine what metaphors students might be invoking. The corpus consisted of approximately 20,000 words, about 2,000 sentences, and 15,000 grammatical relations.

Students' answers were analyzed for metaphors from the ARCHITECTURE domain. This source domain choice was informed partly by prior work describing the metaphor BODIES ARE BUILDINGS (Lakoff, Espenson, & Schwartz, 1991) and partly by the common instructional metaphor CELLS ARE BUILDINGS BLOCKS in the body. Table 1 shows computationally identified metaphors for the concept *cell*. The “Conf” column indicates the confidence score assigned by CMI to the potential metaphor. These scores typically fall in the range 0 to 10 and follow a roughly exponential distribution, with a few high-confidence mappings and many low-confidence mappings. The metaphors shown here are in the upper one percentile in terms of confidence. The “Cell is Like” column indicates the source concept from the ARCHITECTURE domain mapping to *cell*, i.e., a cell “is like” the contents of this column. To reiterate, the source and target concepts are represented by the automatically identified clusters of synsets. The “Target Example Fragments” and “Source Example Fragments” columns show example sentence fragments (with errors unaltered) that serve as evidence for the metaphor in the target and source corpora, respectively, matched by verb-case slot. While the examples may not seem individually compelling, in the aggregate they can suggest potential conceptual framings. The “#” column shows the total number of sentences from each corpus acting as evidence for the metaphor with each verb-case slot. These mappings are each mediated by six to nine verb-case slots; in the interest of space, only three are shown for each mapping.

The example sentences demonstrate CMI's adherence to the idea that mappings

| Conf | Cell is Like | Target Example Fragments | # | Source Example Fragments | # |
|---|---|---|---|---|---|
| 5.157 | material | “... **created from** an already existing **cell** ...” | 2 | “... **creating** electricity **from hydrogen** ...” | 5 |
| | | “... **mixing dna** with other **organism** ...” | 14 | “... **mixed with** an **aggregate** ...” | 15 |
| | | “... **producing** more identical **cells** ...” | 13 | “... **produces** pig **iron** ...” | 93 |
| 5.023 | style | “... **cell** is **created** ...” | 2 | “... **style** was **created** ...” | 48 |
| | | “... **combine with** the original daughter **cell** ...” | 10 | “... **combine** it **with** other **styles** ...” | 47 |
| | | “... **leads to** new **organisms** ...” | 78 | “... **led to** the Baroque **style** ...” | 105 |
| 4.838 | building | “... **creating** one **cell** ...” | 16 | “... **created** blob-like **architecture** ...” | 106 |
| | | “... **produces** completely alike **organisms** ...” | 13 | “... **produced** Beaux-Arts **architecture** ...” | 51 |
| | | “... **created from** an already existing **cell** ...” | 2 | “... **creates** isovists **from building** ...” | 2 |
| 4.482 | piece | “... **replicated in** each new **cell** ...” | 1 | “... **replicated in** red brick **tile** ...” | 1 |
| | | “... **cell** is **created** ...” | 2 | “... **tiles** were **created** ...” | 1 |
| | | “... **made** up **of** mostly multiple **cells** ...” | 3 | “... **made of** two **pieces** ...” | 3 |
| 4.152 | structure | “... **creating** one **cell** ...” | 16 | “... **creating** spacious **aisles** ...” | 164 |
| | | “... **producing** more identical **cells** ...” | 13 | “... **produced architecture** ...” | 108 |
| | | “... **divides into** two **cells** ...” | 4 | “... **divided into** two **arches** ...” | 28 |
| 3.567 | design | “... **organisms** are **created** ...” | 2 | “... **designs** were **created** ...” | 49 |
| | | “... **produce** four daughter **cells** ...” | 13 | “... **producing** a **design** ...” | 153 |
| | | “... **made from** two mixed **cells** ...” | 1 | “... **made from** the **design** ...” | 34 |

*Table 1: Computationally identified metaphors from the CITY domain mapping to the concept of a* cell *in students' writing.*

occur at the level of superordinate concepts but individual instances occur with respect to situational specifics. For example, in the A CELL IS A MATERIAL mapping, the three example fragments shown here do not involve the word "material" but rather specific instances of materials: "hydrogen," "aggregate," and "iron."

The strongest metaphor, A CELL IS A MATERIAL, helps confirm that CMI behaves as one might expect, resonating with the common instructional metaphor CELLS ARE BUILDING BLOCKS, a metaphor implicit in the educational module. The identified metaphor A CELL IS A PIECE also aligns with this overall conceptual framing that cells are component pieces of a larger whole. On the other hand, the identified metaphors A CELL IS A BUILDING and A CELL IS A STRUCTURE suggest a somewhat different conceptualization, wherein a cell is not a part of a larger system but is a system in its own right, possibly indicating an alternative way in which students are invoking the BODIES ARE BUILDINGS (Lakoff et al., 1991) metaphor.

Two more identified metaphors, A CELL IS A STYLE and A CELL IS A DESIGN, suggest a number of novel, creative relational comparisons. First, rather than a component in a construction process, a cell and the DNA it contains could be seen as a specific style or design for how to build an organism, where one cell may "combine with" another to form the design of a slightly different organism. Second, since all living things are "made from" cells, the very notion of a cell could be seen as a style or design pattern from which nature is constructed. Third, this metaphor draws attention to a potential misconception about a cell's agency. Styles and designs imply stylers and designers with intentionality and agency, both of which have the potential to produce misconceptions when applied to understanding cellular reproduction. Thus, CMI can provide opportunities to examine critically potential metaphors and to determine ways in which they might *not* fit the target concept.This approach has been applied to the design of another computer-based educational module. Students' written answers were analyzed for metaphors, and some identified metaphors were then incorporated back into latter parts of the module. For three such metaphors, students were asked if the metaphor made sense and if it was similar to their thinking. They were also asked how the metaphor might *not* fit, e.g., how a cell might not be like a building or a structure. Responses to these questions are being analyzed to determine not only the amount but also the kinds of critical thinking that CMI can support. Do students focus more on aspects of the source or the target concept. Do they focus on surface features or structural aspects? Surface features arguably play a prominent role as a retrieval cue for generating metaphors (e.g., Holyoak & Koh, 1987), but what role do they play in assessing potential metaphors?

### *Political Blogs*

The second corpus consists of posts from 11 political blogs, all with an express Republican bias, collected during the 2008 US election. These posts totaled approximately 100,000 words, 5,400 sentences, and 16,600 grammatical relations.

These blog posts were analyzed for metaphors from the SCIENCE domain. While political metaphors often draw on the domains of WAR or SPORTS (Howe, 1988), SCIENCE was chosen to explore CMI's ability to find metaphors from non-obvious source domains. Table 2 lists the strongest identified metaphors in this corpus, those for the concept of *candidate,* using the same format as Table 1.

These metaphors provide two novel and potentially compelling framings of a political candidate in an election. First, the metaphor A CANDIDATE IS A SCIENTIST provides a new potential framing to an election, e.g., scientists may be criticized for their ideas, just as candidates might be criticized for policy suggestions. Second, and somewhat more compelling, are the other four metaphors shown here: A CANDIDATE IS A THEORY, A CANDIDATE IS AN IDEA, A CANDIDATE IS A STUDY, and A CANDIDATE IS A HYPOTHESIS. These indicate the way that a candidate represents a potentiality, a theory or an idea to be "tested," something that is not yet fully known, is not yet proven. Some theories work well in some situations but not in others, just as candidates may have strengths and weaknesses in their campaigns. A voter might "support" one candidate while "criticizing" another, just as a scientist might "support" one hypothesis while "criticizing" another. Ultimately, one candidate is proven in an election just as one or another theory might be proven through experimental validation. Rather than a war, a battle, or a contest, these four metaphors provide a new way of seeing an election, framing it as a scientific process of experimentation wherein candidates do not try to defeat one another but rather try to prove themselves through various tests. These mappings demonstrate CMI's ability to identify linguistic patterns indicative of potential metaphors that might not be readily apparent but can provide compelling, novel reconceptualizations of familiar ideas or situations.

These and similar results are being used in a tool designed for readers of political blogs (cf. Baumer, Sueyoshi, & Tomlinson, 2008). metaViz [http://metaviz.ics.uci.edu] visually presents potential metaphors in a variety of blogs and allows users to comments on the metaphors. Rather than focusing on individual posts, metaViz draws readers' attention to patterns that span many posts or multiple blogs. This tool encourages readers to consider not only what is being said by the words themselves, but between and behind the words.

The metaViz system is currently being evaluated from two perspectives. First, a controlled experimental study was recently con-

| Conf | Candidate Is Like | Target Example Fragments | # | Source Example Fragments | # |
|---|---|---|---|---|---|
| 2.156 | theory | "... **Obama** is **tested** ..." | 4 | "... **theory** is **tested** ..." | 14 |
| | | "... **supporting McCain** ..." | 15 | "... **support** the causal **theory** ..." | 72 |
| | | "... **tests** Barack **Obama** ..." | 3 | "... **testing theories** ..." | 40 |
| 1.819 | scientist | "... **Obama** were **elected** ..." | 7 | "... **fellows** are **elected** ..." | 5 |
| | | "... **criticizing Obama** ..." | 2 | "... **criticize scientists** ..." | 9 |
| | | "... **McCain** could be **considered** ..." | 1 | "... **scientists** and engineers may both be **considered** ..." | 34 |
| 1.787 | idea | "... **supporting Obama** ..." | 15 | "... **supported** his later **theory** ..." | 77 |
| | | "... **tests** Barack **Obama** ..." | 3 | "... **test** a causal **model** ..." | 60 |
| | | "... **linked to** the **candidate** ..." | 1 | "... **linked to** another **concept** ..." | 14 |
| 1.767 | study | "... **devoted to** Barack **Obama** ..." | 1 | "... **devoted to** the **sciences** ..." | 32 |
| | | "... **McCain** could be **considered** ..." | 1 | "... **physics** is **considered** ..." | 104 |
| | | "... **identify with** Sarah **Palin** ..." | 2 | "... **identified with** scientific **knowledge** ..." | 6 |
| 1.424 | hypothesis | "... **McCain** has been **tested** ..." | 4 | "... **hypotheses** can be **tested** ..." | 3 |
| | | "... **support** a **candidate** ..." | 15 | "... **supporting** the **hypothesis** ..." | 14 |
| | | "... **described** herself **as** a Legacy **Democrat** ..." | 1 | "... **described as** the chronology projection **conjecture** ..." | 2 |

*Table 2: Computationally identified metaphors from the SCIENCE domain mapping to the concept of a* candidate *in political blogs.*

ducted comparing use of metaViz to reading blogs alone to determine differences in critical thinking and creativity. These results indicate that subjects who used metaViz did not exhibit a greater *amount* of critical thinking, but they did demonstrate wider *variety* in the ways they critically examined potential metaphors (Baumer, Sinclair, Hubin, & Tomlinson, 2009). The second evaluation involves a long-term, *in situ* study with a small pilot group of political blog readers, the goal of which is to understand how metaViz integrates with daily blog-reading practices, how it changes readers' perceptions of language and blogs, and how it might enable new ways of reading.

### CONCLUSION

This paper described computational metaphor identification (CMI), a technique for identifying potential conceptual metaphors in written text. Unlike other computational models of analogical thinking, CMI does not use encoded representations of knowledge structures, but rather identifies potential mappings based on cross-corpus linguistic correspondences. The sample results presented here demonstrate the potential of using CMI to promote critical and creative thinking about conceptual metaphors.

The work described here represents a new opportunity for research on relational thinking. Lakoff argues that “issues [about metaphor] are not matters for definitions; they are empirical questions” (1993, p. 202). However, one common critique of such cognitive linguistics work is its lack of clarity with respect to methods, especially its empirical grounding (Gibbs, 2007). By examining systemic patterns in large bodies of text, computational metaphor identification has the potential to help fill that gap, providing a new method for empirical investigation of relational thinking in everyday language and thought.

## Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. IIS-0757646, the Donald Bren School of Information and Computer Sciences, and the California Institute for Telecommunication and Information Technology (Calit2). The authors also thank the Social Code Group and The Learning and Cognition Lab for their feedback and advice.

## References

Baumer, E., Sinclair, J., Hubin, D., & Tomlinson, B. (2009). metaViz: Visualizing Computationally Identified Metaphors in Political Blogs. In *IEEE International Symposium on Social Computing (SocialCom)*. Vancouver, BC, Canada: IEEE Computer Society.

Baumer, E., Sueyoshi, M., & Tomlinson, B. (2008). Exploring the Role of the Reader in the Activity of Blogging. In *ACM SIGCHI Conf* (pp. 1111-1120). Florence, Italy: ACM.

Doumas, L. A. A., Hummel, J. E., & Sandhofer, C. M. (2008). A Theory of the Discovery and Predication of Relational Concepts. *Psychological Review*, *115*(1), 1-43.

Falkenhainer, B., Forbus, K., & Gentner, D. (1989). The Structure-Mapping Engine: Algorithm and Examples. *Artificial Intelligence*, *41*, 1-63.

Fass, D. (1991). Met*: A method for discriminating metonymy and metaphor by computer. *Comp Ling*, *17*(1), 49-90.

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*, 155-170.

Hall, R. P. (1989). Computational approaches to analogical reasoning: A comparative analysis. *Artificial Intelligence*, *39*(1), 39-120. doi: 10.1016/0004-3702(89)90003-9.

Holyoak, K. J., & Koh, K. (1987). Surface and Structural Similarity in Analogical Transfer. *Memory and Cognition*, *15*(4), 332-340.

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science*, *13*(3), 295-355. doi: 10.1016/0364-0213(89)90016-5.

Howe, N. (1988). Metaphor in Contemporary American Political Discourse. *Metaphor and Symbol*, *3*(2), 87. doi: 10.1207/s15327868ms0302_2.

Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review*, *110*(2), 220-264.

Lakoff, G. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), *Metaphor and thought, 2nd. ed.* (pp. 202-251). New York: Cambridge Univ Press.

Lakoff, G., Espenson, J., & Schwartz, A. (1991). *The Master Metaphor List*. Berkeley: Cognitive Linguistics Group, University of California, Berkeley.

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By* (2003rd ed.). Chicago, IL: University of Chicago Press.

Light, M., & Greiff, W. (2002). Statistical Models for the Induction and Use of Selectional Preferences. *Cognitive Science*, *26*(3), 269-281.

Lu, X., & Feldman, A. (2007). Workshop on Computational Approaches to Figurative Language. In *Human Lang Tech / Conf of the North Amer Chapt of the Assoc for Comp Ling (HLT/NAACL 2007)*. Rochester, NY. Retrieved from http://chss.montclair.edu/~feldmana/FigLang2007/.

de Marneffe, M., MacCartney, B., & Manning, C. (2006). Generating Typed Dependency Parses from Phrase Structure Parses. In *Lang Res and Eval (LREC)*. Genoa, Italy. Retrieved from http://nlp.stanford.edu/pubs/LREC06_dependencies.pdf.

Martin, J. (1990). *A Computational Model of Metaphor Interpretation*. San Diego, CA: Academic Press.

Mason, Z. (2004). CorMet: a computational, corpus-based conventional metaphor extraction system. *Comp Ling*, *30*(1), 23-44.

Ortony, A. (1980). Some Psycholinguistic Aspects of Metaphor. In *Cognition and Figurative Language* (pp. 69-83). Hillsdale, NJ: Erlbaum Associates.

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's Mistake: Explaining the Discontinuity Between Human and Non-human Minds. *Behavioral and Brain Sciences*, *31*(02), 109-130. doi: 10.1017/S0140525X08003543.

Resnik, P. (1993). *Selection and Information: A Class-Based Approach to Lexical Relationships*. University of Pennsylvania, Department of Computer and Information Science.

Tourangeau, R., & Sternberg, R. (1982). Understanding and Appreciating Metaphors. *Cognition*, *11*, 203-244.