Using Imageability and Topic Chaining to Locate Metaphors in Linguistic Corpora

George Aaron BROADWELL[1], Umit BOZ[1], Ignacio CASES[1], Tomek STRZALKOWSKI[1], Laurie FELDMAN[1], Sarah TAYLOR[2], Samira SHAIKH[1], Ting LIU[1], Kit CHO[1] and Nick WEBB[3]

(1) State University of New York – University at Albany, NY 12222 USA
(2) Sarah M. Taylor Consulting, LLC
(3) Union College, Schenectady, New York

Abstract. The reliable automated identification of metaphors still remains a challenge in metaphor research due to the ambiguity between the semantic and contextual interpretation of individual lexical items. In this article, we describe a novel approach to metaphor identification which is based on three intersecting methods: *imageability*, *topic chaining*, and *semantic clustering*. Our hypothesis is that metaphors are likely to use highly imageable words that do not generally have a topical or semantic association with the surrounding context. Our method is thus the following: (1) identify the highly imageable portions of a paragraph, using psycholinguistic measures of imageability, (2) exclude imageability peaks that are part of a topic chain, and (3) exclude imageability peaks that show a semantic relationship to the main topics. We are currently working towards fully automating this method for a number of languages.

Keywords: automated metaphor identification, imageability, topic chaining, semantic clustering, linguistic corpora, MRC psycholinguistic database, WordNet.

1 Introduction

Humans can reliably distinguish literal from metaphorical interpretations of words in discourse in a seemingly effortless way. For example, if we consider the word *minefield*, in some cases, it refers literally to a piece of ground in which mines have been placed. In others cases, it refers to some dangerous area or territory. A simple Google search for the word *minefield* will return over four million hits; millions of them will have literal interpretations and millions will have metaphorical interpretations. A human user will be able to distinguish them. But how could a computational system distinguish between the two readings to return only results relevant to mines? Or only results relevant to metaphors about dangerous situations?

Although the development of large linguistic corpora over the last two decades has greatly improved many aspects of the computational linguistic search processes, reliable automated identification of metaphors still remains a challenge in metaphor research. The pioneering work of Martin (1988) and Cameron (1999) in this field has led to a number of techniques, but such approaches also have their limitations. One is that metaphor identification in such methods relies largely on the micro-level discourse analysis of the data with extensive interpretation of the semantic and contextual attributes of individual lexical items (Crisp et al., 2007; Steen et al., 2010).

In this paper, we describe a novel approach to metaphor identification which is based on three intersecting methods: imageability, topic chaining, and a semantic clustering analysis.

2 Metaphor Identification via Imageability

Our first method is based on the imageability of words with a sequence, a property of words which is well-established in the psycholinguistic literature. A word is more imageable to the extent that it is possible to form a mental picture of its meaning. The measure is based on rating data and values range from 100 to 700 (Paivio, Yuille, & Madigan, 1968; Gilhooly & Logie, 1980). Imageability has been shown to play an important role in memory word recall and lexical decision tasks (Bleasdale, 1987; Nelson & Schreiber, 1992; Winnick and Kressel, 1965; Paivio and O'Neill, 1970; Reilly and Kean, 2007) and in experimental tasks of reading (Strain, Patterson, & Seidenberg, 1995). In our metaphor identification method, we use the imageability ratings in percentages based on the original ratings determined via human subjects and available in the MRC psycholinguistic database (MRCPD) (Coltheart, 1981, Wilson 1988). Our hypothesis is that metaphors are likely to use highly imageable words, and words that are generally more imageable than the surrounding context. Our method graphs the words in a paragraph according to their imageability and looks for peaks as potential metaphors.

For instance, Fig.1 shows the imageability ratings of the words in a portion of a paragraph (i.e., "…by simply cutting out all the fat in government") on government and bureaucracy [18]. In this example, the word "fat" has the highest imageability percentage, which, based on our approach, suggests a metaphorical usage in this particular context. As a first step in our metaphor identification method, imageability analysis allows us to identify such highly imageable words as potential metaphors in a given passage. However, considering that such words may be used literally, we further analyze the passage via topic chaining and semantic clustering to identify and eliminate the highly imageable words with literal meanings. Other metaphorical expressions that commonly employ imageable words can be listed as "navigating through the *maze (79%)* of bureaucracy", "being caught in a *web(86%)* of state and local government bureaucracy", "*wheels(%82)* of bureaucracy" etc.

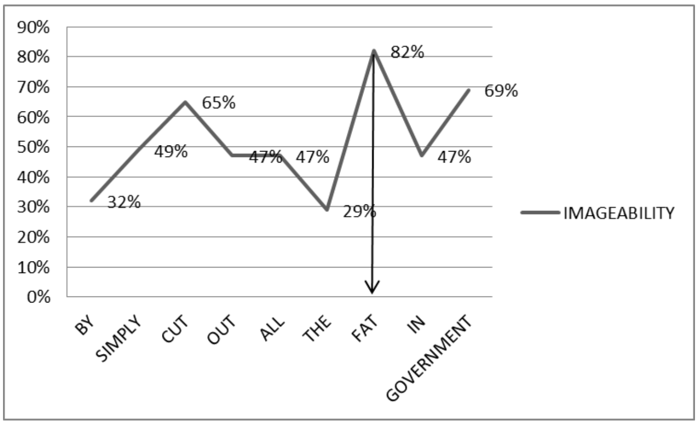

Fig. 1. Imageability values (%) of words in a sample metaphorical expression.

2.1 MRCPD Expansion

Although the MRCPD contains data for over 150,000 words, a major limitation of the database is that not all words have ratings for all 26 variables. This finding is not particularly surprising because the MRCPD is composed from four separate sources of data, and not all sources collected values for all 26 variables. Of interest in the present study is the variable of *imageability* (i.e., how easily and quickly the word evokes a mental image) for which the MRCPD has ratings for only 9,240 (6%) of the total words in its database. We expanded the MRCPD database by adding imagery ratings for an additional 59,989 words (Cho et al., 2013). This was done by taking the words for which the MRCPD database has an imageability rating and using that word as an index to synsets as determined using the WordNet (Miller, 1995) database. New lexical items linked to those synsets were then added to the MRCPD data. The results were validated using regression analyses to compare how well imageability ratings and word frequency of words predict subjects' reaction time in identifying the word, as previous research reported a strong relationship among these variables (Baayen, Feldman, & Schreuder, 2006; Green & Brock, 2002).

For example, the imageability rating for the word *gerbil* is not present in the MRC database, but the word *rat* is present. Since *gerbil* and *rat* share a WordNet hypernym, we developed a method in which (1) nouns and adjectives inherit the imageability scores of sister terms, where sister terms are those with a shared direct hypernym (e.g. where *rodent* is the shared direct hypernym of *gerbil* and *rat*) and (2) nominal and adjectival hyponyms inherit imageability scores from their hypernyms. By the logic of the second step, if we have imageability scores for *dog*, we can allow these scores to be inherited by words such as *terrier* and *poodle.* Because a single English word may appear in multiple WordNet synsets, we also face a question of which synset to select for inheritance purposes. The word *dog*, for example appears in the synset corresponding to the familiar meaning 'domestic dog; canis familiaris', but also in synsets where it serves as a synonym for 'unattractive girl or woman', 'morally reprehensible person', 'andiron', and 'sausage served in a bun', among others. WordNet generally structures lexical relations in such a way that the first synset corresponds to the most literal use of word, therefore we pursued a methodology in which we selected only the first synset for expansion.

A third step proved to be useful for verbs, in which we used verbal imageability scores to approximate the imageability of their hypernyms. Thus from the imageability rating of the verb *to dog*, we approximate the imageability of hyponyms such as *chase,* and *hunt.*

By using a method of imageability rating expansion via first WordNet synset for nouns and adjective and hypernyms for verbs, we were able to construct an expanded imageability lexicon of 32,505 distinct words for English (or 59,989 if words which appear as multiple parts of speech are counted in each occurrence).

To validate the appropriateness of such an expansion, we used the previously-established negative relationship between lexical properties (e.g., lexical frequency, imageability) and reaction time (RT) in a lexical decision task where participants must judge whether letter strings are real words and indicate either "yes" or 'no" by a button press (Baayen, Feldman, and Schreuder 2006; Hargreaves and Pexman, 2012). Specifically, for words that appear infrequently in the literature or are less imageable, subjects require a longer period of time to correctly identify whether the string of letters is an Engish word. The English Lexicon Project (Balota, et al. 2007) (http://elexicon.wustl.edu/) provides RT data as well as lexical measures for over 40,000 words. We used simultaneous multiple regression methods to examine the how well word frequency and imagebility, individually and collectively, predict RT by selecting samples from the original and expanded concreteness database and comparing the predictive relation between RT and concreteness ratings. If the relationships of these variables are similar for both sources of words, i.e., the original and expanded lexicon, we would have confidence that our expansion method is indeed valid. For words in the original MRC database, the frequency and imageability variables are negatively related to RT, -.291, and -.624, respectively.[1] Conjunctively, these two variables accounted for 49.2% of the total variance of RT, which is statistically significant by conventional standards, i.e., $p < .05$, two-tailed.[2] These findings replicate those in the previous literature (e.g., Hargreaves and Pexman, 2012). Of current interest is whether words whose imageability rating is estimated would show a similar relationship. The results indicated the affirmative: frequency and imageability were both negatively related to RT, -.197 and -.747, respectively, and conjunctively they accounted for 58.4% of the variance of RT, which was statistically significant. These findings show that words in the original MRC database and words in the expansion via first WordNet synset (for nouns/adjectives) and hypernym (for verbs) show highly similar predictive values with respect to RT. This demonstrates the validity of our tool.

---

[1] These values represent Standardized β values which indicates the number of standard deviations that the outcome variable changes as a result of one standard deviation change in the predictor.

[2] The adjusted R^2 value ranges from 0-100%. The value is derived from squaring the correlation coefficient.

We take these results to be a strong indication that the method of imageability rating expansion via WordNet produces a lexicon in which words with inherited imageability ratings show very similar psycholinguistic properties to words where imageability rating were derived by experiments with human subjects. With the larger imageability lexicon, we are thus able to approach the problem of linguistic metaphor identification with a more robust tool which is able to plot imageability values for the great majority of English text. We should also note that the imageability analysis along with other two methods reported in the following sections is currently being applied to metaphor identification in Spanish. Comparable imageability results are available in the Spanish psycholinguistic literature at Davis and Perea (2005) and other sources. We have used a expansion technique parallel to that just described for English with the Spanish imageability results to make them usable for a larger portion of the Spanish lexicon.

3 Topic Chaining

We combine the search via imageability with a second method which seeks to establish the topic chains in a discourse. We follow the definition of topic chain in Broadwell et al. (2012) where a topic chain is a noun phrase along with all subsequent noun phrases which refer to the first noun phrase (by pronominal mention, repetition, or synonym). According to Broadwell et al., each noun phrase in a topic chain can be considered as a *local topic*, whereas a noun phrase outside of any topic chain (a word that is mentioned only once in data) is conceptualized as a non-local topic. Drawing on this definition, our hypothesis is that metaphorical words do not tend to have a topical association with the surrounding context, thus standing out from the topical structure of discourse like islands on the ocean.

For instance, we want to exclude highly imageable phrases (e.g. *the fat*, as in *cut the fat*) which occur in a paragraph on grilling meat, since the word *fat* is likely to be topically related to the previous mentions of *fat,* thus being used with its basic meaning. However the same phrase, *cut the fat*, should be counted as metaphorical in a paragraph on economics (e.g., there really was no *fat* left to *cut* from the budget), since our hypothesis is that neither imageable word is likely to appear outside a metaphorical context.

Table 1 illustrates our topic chaining method by indicating the majority of the topic chains extracted from a text on government and economy. In that particular text, the distribution of noun phrases by topicality was as follows: topic chains 68% (103 counts), non-local topics 32% (48 counts). Upon filtering out the topic chains, our imageability analysis indicates the non-local topics *armies*, *cancer*, and *fat* as having the highest imageability scores (83%, 81%, 82% respectively) in the passage, which are all in fact used metaphorically.

Table 1. Longest topic chains in a passage on government and economy.

Topic Chains	Total Number of Local Topics
bureaucracy	22
government	19
tax	7
conservatives	6
city	5
waste	5
budgets	3
agencies	3
programs	3
man	3

One limitation of this method is that in certain cases, a metaphorical word may be repeated more than once, thus forming a topic chain with its subsequent mentions. In that case, our method would disregard such instances for further consideration as a metaphor. While this phenomenon may be highly infrequent, we plan to address this limitation in our future efforts.

4 Semantic Clustering

The third method is a semantic clustering technique, in which we identify and cluster words in the data that are semantically related to the main topics of the paragraph. As with topically related words, we exclude semantically related words from the set of likely metaphors. In order to implement this particular analysis, we are developing an automated technique which incorporates search processes from the Corpus of Contemporary American English (COCA) and Princeton WordNet.

As a sample analysis, let's consider the following passage that includes a metaphorical phrase:

> Seven Morgan County transportation enhancement projects awarded over the past two years remain in the planning stages, *seemingly ensnared in a web of state and local government bureaucracy*.

In this example, *ensnared* and *web* are imageability peaks with their scores 73% and 86% respectively; *award* (img:75%) is also high in imageability, but is excluded as a potential metaphor because of its semantic association with *project* (img:65%). The words *transportation* (img:61%), *state* (img:73%), *local* (img:69%), and *government* (img:69%) are also filtered out as they are part of a topic chain in the greater context.

As previously discussed, topic chaining and semantic clustering methods act essentially as filters on the first method to exclude highly imageable lexical items where contextual clues point toward a literal reading. In this respect, we believe our method results in fewer false positives that the other methods of semantic annotation (e.g. Koller et al 2008) which do not incorporate a contextual component.

Our method is thus the following: (1) identify the highly imageable portions of a paragraph, using both existing psycholinguistic measures of imageability and an expanded lexical database of imageability, (2) exclude imageability peaks that are part of a topic chain, and (3) exclude imageability peaks that show a semantic relationship to the main topics.

5 Evaluation

In order to create a ground truth for the accurate retrieval of metaphorical examples, we collected human assessments both from expert judgments as well as crowd-sourcing interfaces like Amazon Mechanical Turk. We first presented participants with a brief set of instructions on how to identify a number of metaphors in a given context, which was immediately followed by the actual assessment task. For the actual task, participants were given a set of text fragments for which they were asked to judge the degree to which “metaphorical language is present or not”. The validation task involved a set of other questions such as how imagable and common the expression was. Once we had enough human assessment collected, we could establish a ground truth against which system performance could be judged. Additionally, crowd-sourced data was validated using a grammar test score as well as ability to correctly classify known metaphorical and literal examples.

Once ground truth was established in this manner, we could adjust the performance of system in correctly classifying metaphorical vs. literal language. Our prototype system performs at 71% accuracy in detecting metaphors in English language data and 80% for Spanish language data. It should be noted that this result is obtained by applying a series of approx. 100 examples where only about 50% are determined to be metaphorical by our assessors. We anticipate that this initial performance can be substantially improved by optimizing the automated underlying processes, including dependency parsing, local topic co-reference tracking, topical clustering, and word sense disambiguation. We have not yet attempted these optimizations, focusing instead on feasibility of the overall method. Part of our future work is also to compare the contribution of each of the three methods of metaphor identification described here on the overall performance of our system.

6 Conclusion

Using automated tools to identify metaphorical language is a challenging task. However, we believe that our approach addresses this challenge by incorporating a multilayered analysis where lexical, semantic, and contextual properties in a given text are captured and formulated towards making a reliable distinction between literal and metaphorical readings. Additionally, unlike traditional machine learning approaches, our system is not reliant on large amounts of training data. The logic of our approach draws on an intuition found in much other work on metaphor (e.g. Steen et al, 2010) that metaphors are used to express ideas in a more concrete form. The use of *imageability* scores allows us to operationalize concreteness, and the *topic chain* and *semantic cluster* filter out false positives to improve the precision of the final system.

Acknowledgment: This research is 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

1. Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., …Treiman, R: The English lexicon project. Behavior Research Methods, 39, 445– 459 (2007)
2. Baayen, R.H., Feldman, L.F., Schreuder, R.: Morphological influences on the recognition of monosyllabic monomorphemic words. Journal of Memory and Language. 53, 496-512 (2006)
3. Bleasdale, F. A: Concreteness-dependent associative priming:Separate lexical organization for concrete and abstract words. Journal of Experimental Psychology: Learning, Memory, & Cognition, 13, 582-594 (1987)
4. Broadwell, G., Stromer-Galley, J., Strzalkowski, T., Shaikh, S., Talor, S., Liu T., Boz, U., Elia, A., Jiao, L., Webb, N.: Modeling SocioCultural Phenomena in Discourse. Journal of Natural Language Engineering. Cambridge University Press, Cambridge (2012)
5. Cameron, L.: Operationalising ‘metaphor’ for applied linguistic research. In: Cameron, L., Low, G. (eds.) Researching and Applying Metaphor, pp. 3-28. Cambridge University Press, Cambridge (1999)
6. Cho, K.W., Webb, N., Feldman, L.B., Strzalkowski, T., Shaikh, S., Broadwell, G.A., Liu, T., Boz, U., Taylor, S.: Automatic Expansion of the MRC Psycholinguistic Database Imageability Ratings. (forthcoming)
7. Coltheart, M.: The *MRC* Psycholinguistic Database. Quarterly Journal of Experimental Psychology. 33A, 497-505 (1981)
8. Crisp, P., Gibbs, R., Deignan, A., Low, G., Steen, G., Cameron, L., Semino, E., Grady, J., Cienki, A., Kövecses, Z., The Pragglejaz Group: MIP: A method for identifying metaphorically used words in discourse. Metaphor and Symbol. 22, 1–39 (2007)
9. Douglas, A.J. The case for bureaucracy. Government is Good: An Un apologetic Defense of a Vital Institution. http://www.governmentisgood.com/articles.php?aid=20&print=1 (2007)
10. Green, M.C., Brock, T.C.: In the mind’s eye: Transportation-imagery model of narrative persuasion. In: Green, M.C., Strange, J.J., Brock, T.C. (eds.)

Narrative impact: Social and cognitive foundations, pp. 315-341. Lawrence Erlbaum Associates Publishers, New Jersey (2002)

11. Gilhooly, K.J. and Logie, R.H: Age of acquisition, imagery, concreteness, familiarity and ambiguity measures for 1944 words. Behaviour Research Methods and Instrumentation, 12, 395-427 (1980).
12. Hargreaves, I. S., & Pexman, P. M: Does richness lose its luster? Effects of extensive practice on semantic richness in visual word recognition. Frontiers in Human Neuroscience, 6: 234, doi: 10.3389/fnhum.2012.00234 (2012)
13. Koller, V., Hardie, A., Rayson, P., Semino, E.: Using a semantic annotation tool for the analysis of metaphor in discourse. Metaphorik.de. 15, 141-160 (2008)
14. Martin, J.H.: A computational theory of metaphor. University of California, Berkeley, Computer Science Division. Rep. No. UCB/CSD, 88-465 (1988)
15. Miller, G.A.: WordNet: A Lexical Database for English. Communications of the ACM. 38, 39-41 (1995)
16. Nelson, D. L., & Schreiber, T. A: Word concreteness and word structure as independent determinants of recall. Journal of Memory and Language, 31, 237-260 (1992)
17. Paivio, A., Yuille, J.C., & Madigan, S.A: Concreteness, imagery, and meaningfulness: Values for 925 nouns. Journal of Experimental Psychology, 76, 1–25 (1968)
18. Paivio, A., O'Neill, B.J.: Visual recognition thresholds and dimensions of word meaning. Perception and Psychophysics 8, 273-75 (1970)
19. Reilly, J. & Kean, J.: Formal Distinctiveness of High- and Low-Imageability Nouns: Analyses and Theoretical Implications.Cognitive Science 31(1), 157-168 (2007)
20. Steen, G.J., Dorst, A.G., Herrmann, J.B., Kaal, A.A., Krennmayr, T., Pasma, T.: A Method for Linguistic Metaphor Identification: From MIP to MIPVU. John Benjamins Publishing, Amsterdam/Philadelphia (2010)
21. Strain, E., Patterson, K., & Seidenberg, M. S: Semantic effects in single-word naming. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21, 1140-1154 (1995)
22. Wilson, M.D.: The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2. Behavioural Research Methods, Instruments and Computers. 20, 6-11 (1988)
23. Winnick WA, Kressel K.: Tachistoscopic recognition thresholds, paired-associate learning, and immediate recall as a function of abstractness-concreteness and word frequency. Journal of Experimental Psychology. 70, 163–168 (1965)