Can Cognitive Scientists Help Computers Recognize Irony?

Byron Wallace (byron_wallace@brown.edu)
Public Health-Health Services Policy & Practice, Box G-S121-8
Providence, RI 02912-G

Laura Kertz (laura_kertz@brown.edu)
Cognitive, Linguistic, & Psychological Sciences, Box 1821
Providence, RI 02912-1821

Abstract

This workshop aims to bring together researchers in cognitive science and computer science with a shared interest in irony. Irony detection is an especially difficult problem for natural language processing. Unlike other types of classification tasks, the difficulty of identifying irony is not alleviated by 'throwing more data' at the problem. Rather, it seems a different *kind* of data is needed: contextual data. We also need new *models* that can exploit this data. In this workshop, which capitalizes on the collocation of the AAAI and CogSci conferences, we invite cognitive scientists and computer scientists to engage in a dialogue around new machine learning models for irony detection and new methods and tools for testing predictions of cognitive theories of irony against large-scale data sets.

Keywords: irony, metaphor, pragmatics, natural language processing, machine learning

Irony in the Age of the Internet

Irony is a rhetorical device that takes many forms, e.g. *hyperbole* or *sarcasm*. When successful, an ironic utterance manages to communicate something other than what is literally said. This device may be exploited in interpersonal communication for a variety of reasons, e.g., to inject humor into a situation or to soften criticism. Irony may also serve complicated social functions like supporting group cohesion or reinforcing group norms. (See Gibbs, 2000, for a review).

While irony has been a subject of scholarly interest for millennia—especially within philosophy and literary studies—the contemporary study of irony sits squarely within the domain of cognitive science. Our understanding of what irony is and how it is comprehended has been informed by research carried out over the last 30 years by researchers working at the intersection of language and social cognition.

More recently, as the Internet has taken on an increasingly prominent role in mediating human communication, the task of automatically detecting ironic intent has captured the interest of computer scientists in the sub-disciplines of machine learning (ML) and natural language processing (NLP). There are two reasons for this. First, there is a practical need. Increasingly, companies have begun to leverage automated methods for inferring consumer sentiment from online reviews and other sources. Irony hinders the performance of these systems, because ironic utterances convey sentiment that is the opposite of (or at least not equivalent to) what the writer/speaker actually believes. Although the prevalence of irony online has yet to be estimated with any precision, there are reasons to suspect that it is non-negligible. (Our own preliminary data suggests that irony is present in 10-15% of comments posted to the social news aggregation site Reddit[1].) Irony thus poses a real and growing challenge for the task of accurate sentiment detection.

The second reason that the problem of irony detection is attracting attention is the sheer quantity of data available on the Internet. Many problems long thought to be intractable have yielded, of late, to relatively simple machine learning models *trained* on massive datasets. Machine translation is an instructive example of this: Google translations are now passable due primarily to their exploitation of massive parallel corpora.

The standard machine learning approach to discerning ironic utterances found online (e.g., in tweets or comment posts) has been to transform utterances into vectorized representations that capture counts of *tokens*, which may include words and punctuation (for example). This is called the *bag-of-words* representation. A classification model is then trained to predict whether a given utterance is *ironic*, based on its bag-of-words representation. With a sufficient amount of manually categorized examples (i.e., *training data*), this approach has proven extremely successful for many classification tasks, such as spam filtering. But irony detection has proved to be much harder.

The reason irony poses such a challenge is because it is difficult to detect irony using the simplistic bag-of-words representation described above. Consider, for example, that a single utterance—"Obama is a socialist", for example—can be intended either literally or ironically. Some people genuinely believe this statement, while others would be tacitly mocking the literal proposition. There is simply no way to know which is the case from the tokens comprising the utterance alone. Rather, what seems to be needed is information about the speaker and/or the speaker's communicative intent.

Insights from Cognitive Science

A common theme in theories of irony within linguistics, psychology, and philosophy, is a notion of distance between the speaker who issues an ironic utterance and the content of that utterance itself. It has been suggested, for example that an ironic stance is a form of role-play (Clark and Gerrig,

[1] http://www.reddit.com

1984) or 'echoing' another's belief (Wilson and Sperber, 1992). This suggests that part of understanding irony is assessing the 'goodness of fit' between the content of an utterance and the speaker who has uttered it. This link between speaker and message is nowhere to be found in existing computational models of irony.

Furthermore, research into the psycholinguistics of irony has confirmed the importance of context in supporting successful interpretation. Context may include the novelty (or familiarity) of the utterance itself (Giora & Fein, 1999), information about the speaker, and whether the utterance is consistent with prior discourse (Pexman & Ferretti, 2000). Each of these factors influences the likelihood of interpreting a given utterance as ironic and may also affect the ease of processing, as well as subsequent recall. These findings underscore the relevance of considering an utterance with respect to its speaker for the purpose of detecting irony and furthermore indicate that consistency (or disparity) between an utterance and its discourse context might serve as useful information for detecting ironic intent.

Meeting of the Minds

The proposed workshop will capitalize on the co-location of CogSci with AAAI to bring together cognitive scientists, linguists, and artificial intelligence researchers in a dialogue about modeling and detecting irony in text. Insights from cognitive science may inform computational approaches to irony detection by suggesting novel sources of information that can be exploited in a machine learning model. In particular, we believe that developing representations of speakers and contexts and building models that factor these representations into judgments of utterances may drastically improve automated irony detection.

And we hope that feedback will also run in the opposite direction: operationalizing and then experimenting with models that incorporate insights from linguistics and cognitive science is likely to shed empirical light on the validity of theories of irony usage. Indeed we suspect that bringing computer scientists, linguists, and cognitive scientists together will provide exciting new opportunities for empirical testing of cognitive models of irony through the use of large-scale data sets.

Workshop Overview

Organizers Wallace (computer scientist with expertise in machine learning and natural language processing) and Kertz (linguist with expertise in discourse and language processing) are currently co-PIs on a project that seeks to improve irony detection in the ways described above. As part of our research, we are developing a corpus of hand-annotated ironic utterances culled from Internet posts. This corpus will serve as a valuable resource to members of the community targeted for this workshop.

Program The workshop includes a mix of invited talks and selections made from an open call. The invited and accepted talks cover three topics: the social/narrative function of irony, human comprehension of ironic content, and the state of the art in computational irony detection. Invited speakers are Greg Bryant (Communication Studies, UCLA), Ellen Riloff (School of Computing, University of Utah), and Vera Tobin (Cognitive Science, Case Western).

References

Clark, H. H., & Gerrig, R. J. (1984). On the pretense theory of irony. Journal of Experimental Psychology: General, 113(1),121–126.

Gibbs, R. (2007). A brief history of irony. In H. Colston & R. Gibbs (Eds.), Irony in Language and Thought. A Cognitive Science Reader (pp. 3–24).

Giora, R., & Fein, O. (1999). On understanding familiar and less-familiar figurative language. Journal of Pragmatics, 31(12), 1601–1618.

Pexman, P. M., Ferretti, T. R., & Katz, A. N. (2000). Discourse factors that influence online reading of metaphor and irony. Discourse Processes, 29(3), 201–222.

Wilson, D., & Sperber, D. (1992). On verbal irony. Lingua, 87(1), 53–76.