The on-line processing of written irony

Ruth Filik [a,*], Linda M. Moxey [b]

[a] School of Psychology, University of Nottingham, UK
[b] Department of Psychology, University of Glasgow, UK

A B S T R A C T

We report an eye-tracking study in which we investigate the on-line processing of written irony. Specifically, participants' eye movements were recorded while they read sentences which were either intended ironically, or non-ironically, and subsequent text which contained pronominal reference to the ironic (or non-ironic) phrase. Results showed longer reading times for ironic comments compared to a non-ironic baseline, suggesting that additional processing was required in ironic compared to non-ironic conditions. Reading times for subsequent pronominal reference indicated that for ironic materials, both the ironic and literal interpretations of the text were equally accessible during on-line language comprehension. This finding is most in-line with predictions of the graded salience hypothesis, which, in conjunction with the retention hypothesis, states that readers represent both the literal and ironic interpretation of an ironic utterance.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

The use of irony is ubiquitous in everyday utterances. For example, Gibbs (2000) reported that ironic remarks occurred on 8% of conversational turns in talk among friends, while Kreuz, Roberts, Johnson, and Bertus (1996) showed that readers of contemporary American literature encounter approximately one instance of irony every four pages. Because ironic language is so prevalent, its processing and interpretation must be accounted for in any theory of language use and understanding. Despite this, the majority of current research in psycholinguistics has focused on literal language.

Most empirical research on figurative language has concentrated on metaphor, with very little work investigating the underlying mechanisms involved in understanding ironic utterances. However, there has been a relatively large amount of theoretical work concerned with the communicative function of irony (see Gibbs & Colston, 2007, for an overview). From this, a number of contemporary theories regarding how irony is processed and understood have

emerged. In the current paper we report an eye-tracking study in which we contrast the predictions of contemporary theories regarding the processing of written irony.

1.1. The standard pragmatic view

According to the standard pragmatic view (Grice, 1975; Searle, 1979, 1993), irony is a figure of speech that communicates the opposite of what is said. For example, by saying "What lovely weather!" in the middle of a storm, the speaker actually communicates "What terrible weather". Under this view, the comprehension of non-literal language takes place in a number of stages. A reader or listener must firstly compute the utterance's context-independent, literal interpretation, before deciding whether the literal interpretation is the speaker's intended interpretation. If a mismatch with context indicates that the literal interpretation is inappropriate, it is then necessary to cancel the surface-literal interpretation, and compute the non-literal interpretation by assuming the opposite of the literal interpretation. During this process, the context-incompatible literal meaning is suppressed as being irrelevant or disruptive to the intended interpretation. It is clear from this that non-literal language requires further processing effort than

* Corresponding author. Tel.: +44 (0) 115 951 5402.
  E-mail address: ruth.filik@nottingham.ac.uk (R. Filik).

literal language, in the form of extra inferential processes resulting from a mismatch with context. Thus, from a processing perspective, non-literal language should result in a higher processing cost than the same utterance that is intended literally. In addition, since the literal meaning is suppressed, it should then be relatively inaccessible to readers.

1.2. The direct access view

The standard pragmatic model has been challenged by the direct access view, which assumes that contextual information interacts with lexical processes very early on (see e.g. Gibbs, 1986, 1994; see also Clark & Gerrig, 1984; Sperber & Wilson, 1995, for similar assumptions). The basic premise is that similar underlying mechanisms are involved in the initial processing of both literal and figurative language (Gibbs, 1994). That is, "understanding irony does not necessarily require special cognitive processes beyond those used to comprehend literal speech" (Gibbs, 1994, p. 437).

Thus, if context supports an ironic interpretation of the statement, it can be directly accessed (or constructed, in the case of unfamiliar ironies) without the need to access (or construct) the literal interpretation first. In processing terms, no extra steps would be required for non-literal statements, resulting in no additional processing cost. In terms of which interpretations of the input are maintained in the reader's mental representation, it is clear from this that since "people need not first analyse the literal, pragmatic-free meaning of an utterance before determining its figurative, implicated interpretation" (Gibbs, 1994, p. 421), only the ironic interpretation would be retained, as it would be the only interpretation ever calculated.

2. The graded salience hypothesis

An alternative to these accounts is the graded salience hypothesis (Giora, 1997), according to which the most salient interpretation of a statement is always accessed first. In order to be salient, utterances have to be coded in the mental lexicon. In addition, they must be prominent due to their conventionality, frequency of exposure, experiential familiarity, or prototypicality. Salient interpretations are assumed to be accessed from the mental lexicon immediately on encountering the linguistic input. Non-salient interpretations require extra inferential processes and strong contextual support. In processing terms, if an ironic utterance is familiar (encoded in the lexicon) then the ironic interpretation would be computed first, and would not require additional processing (Giora & Fein, 1999). In contrast, if an ironic utterance is novel or unfamiliar, then the more salient literal interpretation would be computed first, leading to an additional processing cost when a mismatch with context means that the utterance must be re-interpreted as being ironic.

Underlying the graded salience hypothesis is the assumption that irony is a form of indirect negation of which the function is to draw attention to a failed expectation (Giora, 1995). For example, by saying "What lovely weather!" in the middle of a rainstorm, the speaker points out that the observed state of affairs does not live up to expectations, and the weather is in fact far from being lovely. Based on this view, interpreting irony does not involve cancelling the literal interpretation and replacing it with the opposite interpretation, as suggested by the standard pragmatic view. Rather, both the literal and intended interpretations are maintained so that the dissimilarity between them may be computed. More recently, this process has been explicitly formulated as a functional principle referred to as the 'retention hypothesis' (e.g. Giora, 2003; Giora & Fein, 1999), which specifies that both the literal and ironic interpretations are "integrated into the current representation of the discourse" (Giora, 2003, p. 72), making both available for further processes.

3. Empirical evidence

To test the predictions of these accounts, a number of empirical studies have investigated the processing of ironic vs. non-ironic statements. For example, Giora, Fein, and Schwartz (1998, Experiments 2 and 3) presented participants with a probe word related either to the literal or ironic interpretation of a target phrase. After delays of 150 ms and 1000 ms, reaction times to probe words related to the literal interpretation were faster than those related to the ironic interpretation. This difference disappeared at 2000 ms, suggesting that the ironic interpretation became as available as the literal interpretation relatively late. Giora et al. (1998, Experiment 1) report further evidence of delayed processing for ironic statements, with longer sentence reading times for target sentences presented in irony-biasing than literal-biasing contexts (see also Dews & Winner, 1999; Schwoebel, Dews, Winner, & Srinivas, 2000).

In contrast, Gibbs (1986, Experiment 1) reported *shorter* sentence reading times for ironic comments (e.g. "You are a fine friend") than non-ironic counterparts ("You are a bad friend"). In addition, when reading times for sentences with the same surface form ("You are a fine friend") were compared in ironic (somebody *not* being a good friend) and non-ironic (somebody being a good friend) contexts, there were no differences. However, simple non-ironic acknowledgements like "You are a good friend" were read more quickly than ironic comments (see Giora, 1995, for further discussion of Gibbs' results).

Evidence from ERPs (Katz, Blasko, & Kazmerski, 2004) suggests that although there is a processing cost related to the comprehension of irony, it occurs early, specifically, before the reader has finished reading the last word in the phrase. Furthermore, in a word-by-word self-paced reading study, Ivanko and Pexman (2003, Experiment 3) showed that context could modulate processing difficulty for ironic utterances. Importantly, in some contexts, reading times for ironic statements were faster or equivalent to those for non-ironic statements.

It can be seen from this that empirical evidence regarding the time-course of irony processing is somewhat mixed. The research reported above used a variety of methodologies, such as probe word, word-by-word reading time, whole sentence reading time, and ERPs. While providing interesting results, these methodologies have a