On the Difficulty of Automatically Detecting Irony: Beyond a Simple Case of Negation

Antonio Reyes[1,2] and Paolo Rosso[2]

[1]Laboratorio de Tecnologías Lingüísticas
Instituto Superior de Intérpretes y Traductores, Mexico
[2]Natural Language Engineering Lab - ELIRF
DSIC, Universitat Politècnica de València, Spain
antonioreyes@isit.edu.mx
prosso@dsic.upv.es

Abstract. It is well-known that irony is one of the most subtle devices used to, in a refined way and without a negation marker, deny what is literally said. As such, its automatic detection would represent valuable knowledge regarding tasks as diverse as sentiment analysis, information extraction, or decision making. The research described in this article is focused on identifying key values of components to represent underlying characteristics of this linguistic phenomenon. In absence of a negation marker, we focus on representing the core of irony by means of three conceptual layers. These layers involve 8 different textual features. By representing four available data sets with these features, we try to find hints about how to deal with this unexplored task from a computational point of view. Our findings are assessed by human annotators in two strata: isolated sentences and entire documents. The results show how complex and subjective the task of automatically detecting irony could be.

1 Introduction

The problem of negation cuts through every aspect of language, from pronunciation to lexical choice, syntactic structure, semantics and conceptualization. Several approaches have been developed in order to throw light on the linguistic and cognitive processes that underlie this defining characteristics of humans [20]. For instance, from a psycholinguists perspective, the experiments reported by [17]. Here, the authors assessed whether or not the information introduced via a negation marker is retained or suppressed when mentally representing the negated concepts. In a same vein, considering a reading paradigm, [21] investigated the amount of milliseconds to process affirmative and negative sentences in which a target entity and a contradictory predicate were mentioned. Such research works have provided valuable clues to unveil subjective aspects of meaning, which are profiled by the presence of negation markers or by using linguistic strategies. However, most of them rely on literal language. In this type of language, of course, there is a formal linguistic constituent to indicate that something is being negated. The problem of negation becomes even more difficult to tackle when

dealing with figurative language which, unlike literal language, masks its real meaning by exploiting different linguistic devices to veil its underlying and real meaning. One of these devices is irony.

It is well-known that irony is one of the most subtle devices used to, in a refined way and without a negation marker, deny what is literally said. This fact represents a real challenge, from a theoretical point of view, as well as from a practical one. In this respect, most people agree that irony relies on *opposition*. However, unlike literal language, this opposition is not formally marked; i.e. there is not an explicit negation marker underlying ironic statements.

In this article we investigate how to represent, by means of textual features, the conceptual core of this figurative device. We think that it is unrealistic to seek a computational silver bullet for irony, and a general solution will not be found in any single technique or algorithm. Rather, we must try to identify specific aspects and forms of irony that are susceptible to computational analysis, and from these individual treatments, attempt to synthesize a gradually broader solution.

The rest of the paper is organized as follows: in Section 2 we describe the theoretical challenges which underpin any computational approach of irony, then our definition of this concept is introduced. In Section 3 we define the specific objective of this current work. Then, related work as well as data sets are described. In Section 4 our new linguistic model is introduced. In Section 5 we describe the experiments and discuss the results. In Section 6 we then present the evaluation and analyze its implications. Finally, in Section 7 we conclude with some final remarks and address some pointers to future works.

2   Figurative Language

The analysis of figurative language is one of the most arduous topics that natural language processing has to face. Unlike literal language, the former takes advantage of linguistic devices, such as metaphor, analogy, satire, irony, and so on, in order to communicate more complex meanings which, usually, represent a real challenge, not only for computers, but for humans beings as well. This kind of communication entails cognitive capabilities to abstractly represent meanings beyond simple words, or beyond syntax and semantics. In this layer, a communication act entails more than sharing a common code, it entails information not grammatically expressed (e.g. cultural or social referents, contextual knowledge, as well as linguistic competence) in order to be able to uncover the real meaning: if this information is not unveiled, the underlying and *real* meaning is lost and the figurative effect is not accomplished. One of the linguistic devices that most clearly represent the complexity of figurative language is irony. This device, apart from taking advantage of different linguistic strategies to be able to produce its effect (e.g. similes [39], or usage of satirical or sarcastic utterances to express a negative attitude [24, 3]), is a form of negation that does not make use of an explicit negation marker [16].

In the following subsection we give a brief overview of the concept of irony, and then we establish how it is interpreted within the framework of this research.

2.1 Irony

Like most figurative devices, irony is difficult to define in formal terms, and no single definition ever seems entirely satisfactory. So to begin with, let us consider three obvious examples of irony in everyday situations:

1. A man goes through the entrance to a building but fails to hold the door for the woman right behind him, even though she is visibly struggling with a heavy box. She says "Thank You! anyway".
2. A professor explains and re-explains a complicated algorithm to his class of undergraduates. After showing an image to clarify the concept, he asks: "Is it clear now?". "Clear as mud", a student replies.
3. After seeing a stereotyped romantic movie, the guy says: "I never believed love at first site was possible until I saw this film".

These examples suggest that pretense plays a key role in irony: speakers craft utterances in spite of what has just happened, not because of it. The pretense in each case alludes to, or echoes, an expectation that has been violated (cf. [8, 36]). This pretense may seem roundabout and illogical, but it offers a sharply effective and concise mode of communication. In this context, irony allows a speaker to highlight the expectation that has been violated while simultaneously poking fun at, and often rebuking, the violator (e.g. "clear as mud"). Additionally, an underlying sensation of false message (or $negation$ of what is expressed) permeates the conclusion of these examples (e.g. "thank you!" instead of "fuck you").

The experts point out that irony is essentially a communicative act which expresses an opposite meaning of what was literally said [42]. However, this is only one kind of irony. In the specialized literature we found two primaries kinds of irony: $verbal$ and $situational$. On the one hand, verbal irony conveys an opposite meaning; i.e. a speaker says something that seems to be the opposite of what s/he really means [10] (like in the previous examples). On the other hand, situational irony is a state of the world which is perceived as ironical [3]; i.e. situations that should not exist [27]. Moreover, others authors distinguish fine-grained types of ironies: dramatic irony [3]; discourse irony [24]; tragic irony [9]; etc.

Our work is focused on modeling features regarding verbal irony. This kind of irony is defined as a way of intentionally denying what it is literally expressed [11]; i.e. a kind of *indirect negation* [16], but as we previously pointed out, with no explicit negation markers. Moreover, and according to some pragmatic frameworks, the set of elements to truly decide whether or not an utterance is ironic varies depending on the author. For instance, Grice [19] considers that an utterance is ironic if it intentionally violates certain conversational maxims. Wilson and Sperber [42] assume that one of the most important elements to detect verbal irony should rely on understanding an utterance as echoic. Utsumi [38], in turn, suggests a ground of negative emotional attitudes to create an ironic environment. All these characteristics make irony a fuzzy device to be computationally, and even linguistically, modeled. For instance, fine-grained concepts, such as conversational maxim, imply ground knowledge that cannot be directly mapped from theory to $praxis$, due largely to the idealized communicative scenarios which they entail. Most times, such scenarios do not completely match with the examples found in

everyday situations (e.g. a detailed and manual analysis of the implicatures of what is literally said). In addition to such complexity, people have their own concept of irony, which rarely satisfies all the characteristics suggested by the experts, but instead, is mixed with other concepts. Let us consider the following examples:

4. "I feel so miserable without you, it's almost like having you here."
5. "Don't worry about what people think. They don't do it very often."
6. "Of course I'm in shape. Isn't round a shape?"

According to some user-generated tags, these three examples could be either ironic, or sarcastic, or even satiric. However, beyond the fact of what tag should be the right for every expression, we want to focus on the fact that, for many people, there is not a clear distinction with respect to the boundaries for differentiating between irony and sarcasm, or between irony and satire. This gets worse when the experts do not clearly define the boundaries among these concepts. For instance, in [9] and [12], sarcasm is a term commonly used to describe an expression of verbal irony; whereas in [14], sarcasm along with jocularity, hyperbole, rhetorical questions, and understatement, are only types of irony. In contrast, in [24], the authors consider a type of sarcastic irony which is opposed to the non sarcastic; while in [3], sarcasm is an overtly aggressive type of irony. Furthermore, for [6], satirical texts, specifically news articles, tend to incorporating irony and non sequitur in an attempt to provide a humorous effect; whereas for [15], irony is often compared to satire and parody.

Taking into consideration these statements, it is obvious how the limits among these figurative devices are not clearly differentiable. They share more similarities than differences. For instance, with respect to the similarities, it seems to be an underlying negative attitude which permeates such concepts; regarding the differences, they rely indeed on matters of usage, tone, and obviousness, which are not so evident in ordinary communication acts. Therefore, in this article and according to our purposes, we begin by defining irony as *a verbal expression whose formal constituents, i.e. words, attempt to communicate an underlying meaning which is opposite to the one expressed*. In addition, we differentiate between *aim* and *effect*. The aim of irony, according to our definition, is to communicate the opposite of what is literally said; whereas the effect may be a sarcastic, satiric, or even a funny sense that undoubtedly profiles negative connotations. In this context, sarcasm, satire, and figurative devices such as the ones suggested in [14] (jocularity, hyperbole, rhetorical questions, and understatement), are only specific extensions of a general and broad concept of irony[1].

3 Towards Irony Detection

According to [20], *negation* and its correlates (truth-values, false messages, contradiction, and *irony*), are defining characteristics of the human race. For this powerful

---

[1] We do not make a fine-grained distinction among these figurative devices due to the difficulty to clearly separate their uses, in particular when the focus is on user-generated contents, and due to the general point of views (cited previously) which take them as subtypes of irony. Nevertheless, there are others approaches which are focused on specific figurative devices (cf. the research described in [37, 12] with respect to sarcasm, as well as the one described in [6] regarding satire).

reason, any attempt to model these phenomena faces serious problems when setting up the objective as well as the scope and applicability of the results. In the following section, we describe our objective on the basis of a specif task: sentiment analysis. In addition, the related work and the data sets employed in our experiments (Section 5), will be introduced.

3.1 Objective

One of the most difficult problems when assigning either positive or negative polarity in sentiment analysis tasks is to determine what is the truth value of a certain statement. Like negation, irony allows to change the truth value of any statement. In case of literal language (e.g. "this movie is crap") the existent techniques achieve good results; instead, when the meaning in ground is totally different to the meaning in figure[2], the result may be a consequence of simply finding out what kinds of words prevail in the surface of such statement (e.g. "It's not that there isn't anything positive to say about the film. There is. After 92 minutes, it ends"). In such cases, the same automatic techniques lose effectiveness because the profiled and real meaning is in ground, or *veiled*, by the use of figurative devices. Such veiled meaning may be evident for humans; i.e. after processing the information of the last example, we easily realize that a negative polarity permeates the statement. However, how do we do to define a computational model capable to recognize the veiled meaning in which irony relies? Furthermore, how to differentiate between such examples if both may express either literal or figurative language?

The questions seem to be nearly impossible to be computationally solved. Nonetheless, in this research we attempt to investigate and provide insights into the figurative uses of textual elements to communicate ironic statements. Our objective thus is to propose a model capable of representing the most obvious attributes of verbal irony in a text (or at least what speakers believe to be irony), in order to be able to automatically detect possible ironic statements in user-generated contents (opinions, comments, reviews). The expected result is to provide information to the experts, either at sentence level or at document level, who will decide whether or not such information is really ironic[3].

Such information might represent fine-grained knowledge, which may be applied in tasks as diverse as sentiment analysis (cf. [32] about the importance of determining the presence of irony in order to set a fine-grained polarity), opinion mining (cf. [35], where the authors note the role of irony for minimizing the error when discriminating negative from positive opinions), or even advertising (cf. [23], about the function of irony to increase message effectiveness).

---

[2] Cf. [25] for a detailed explanation about the concepts employed in Cognitive Grammar.

[3] Like most tasks that involve information beyond grammar; i.e. subjective, social or cultural knowledge (e.g. machine translation), we believe that the irony detection task implies a human assessment to validate the results as well as to learn from errors. Further improvements of the model would suppose a less human involvement (such as in the research work reported in [5]).

3.2 Related Work

As far as we know, very few attempts have been carried out in order to integrate irony in a computational framework. One of the first computational approaches to formalize irony was described by Utsumi [38]. However, his model is too abstract to represent irony beyond an idealized hearer-listener interaction. More recently, from the perspective of computational creativity, Veale and Hao [39] have attempted to throw light on the cognitive processes that underlie verbal irony. By analyzing a large quantity of humorous similes of the form "as X as Y" collected from the web, they noted how web users often use figurative comparisons as a mean to express ironic opinions. Another recent approach is described in [7], here the authors presented some clues for automatically identifying ironic sentences by first recognizing emoticons, onomatopoeic expressions, and specific punctuation and quotation marks. Furthermore, Veale and Hao [40] have recently presented a linguistic approach to separating irony from non-irony in figurative comparisons. Authors noted how the presence of ironic markers like "about" can make rule-based categorization of ironic statements a practical reality, at least in the case of similes, and describe a system of linguistically-coded heuristics for performing this categorization. Finally, Reyes and Rosso [33] have proposed a model to integrate different linguistic layers (from simple n-grams to affective content) to represent irony in customer reviews.

In addition, there are others approaches which are focused on specific devices related to irony. This is the case of sarcasm and satire (cf. Section 2.1). In such approaches, authors have directly focused on such particular linguistic devices rather than on the whole concept of irony. For instance, Tsur et al. [37] address the problem of finding linguistic elements to mark the use of sarcasm in online product reviews. Based on a semi-supervised approach, they suggest that specific surface features, such as words that convey information about a product, its maker, its name, and so on, as well as very frequent words, or punctuation marks, can be used to identify sarcastic elements in reviews. In turn, in [12], authors reported high scores of precision, recall and F-measure when applying their algorithm to recognize sarcasm on texts collected from Twitter and Amazon. Finally, Burfoot and Baldwin [6] explore the task of automatic satire detection by evaluating features related to headline elements, offensive language or slang on a corpus of newswire documents and satire news articles.

3.3 Data Sets

Due to the lack of resources for irony detection[4], we decided to use four different data sets (already employed in tasks related to sentiment analysis) in order to evaluate our model. These are:

1. The polarity dataset v2.0 described in [30]. Hereafter $movies2$. This data set contains 1.000 positive and 1.000 negative processed reviews[5].
2. The polarity dataset v1.1 described in [29]. Hereafter $movies1$. This is a cleaned version that integrates 700 positive and 700 negative processed reviews[6].

---

[4] This is probably due to the difficulty and subjectivity to obtain *ad hoc* corpora.

[5] Available at http://www.cs.cornell.edu/People/pabo/movie-review-data.

[6] Ibid.

3. The English book review corpus. Hereafter *books*. This corpus is described in [43]. It contains 750 positive and 750 negative reviews[7].
4. The newswire documents and satire news articles described in [6]. Hereafter *articles*. This corpus is integrated with 4.000 real and 233 satire news articles[8].

All the tags with regard to documents polarity were kept as they were, and no further processing was applied, except by removing the stopwords. The average length per document, in terms of total words, is given in Table 1.

Table 1. Average length per document.

	Positive polarity	Negative polarity
*Movies*2	787.07	705.73
*Movies*1	690.71	712.40
Books	57.32	57.67

	Real	Satire
Articles	255.43	353.35

Finally, it is worth stressing that each data set was treated separately; i.e. the results (to be described in Section 5) are linked to every particular data set; thus, they cannot be generalized to all data sets.

4 Model

We are proposing a new model that is organized according to three conceptual layers: *signatures, emotional scenarios* and *unexpectedness*.

Unlike the existent models that consider surface features such as onomatopoeic expressions (cf. [7]), discursive markers such as "about" (cf. [40]), very frequent words (cf. [37]), or offensive language (cf. [6]), our model is intended to capture both low-level and high-level properties of irony on the basis of conceptual descriptions found in the specialized literature (for instance, opposition or incongruity); i.e. we intend to extract the core of the most defining characteristics of verbal irony (according to several formal studies such as the ones cited in Section 3.2), and then, transfer such core to our model by mapping it through *textual features*.

The textual features to represent each layer are listed and discussed below:

i. Signatures, concerning three textual features: *pointedness*, *counter-factuality*, and *temporal compression*.

---

[7] Available at http://www.informatics.sussex.ac.uk/users/tz21.
[8] Available at http://www.csse.unimelb.edu.au/research/lt/resources/satire.

ii. Emotional scenarios, concerning three textual features: *activation*, *imagery*, and *pleasantness*.

iii. Unexpectedness, concerning two textual features: *temporal imbalance* and *contextual imbalance*.

4.1 Signatures

This layer is focused on representing irony in terms of specific textual markers or signatures. It is largely characterized by typographical elements such as punctuation marks and emoticons, as well as by discursive elements that suggest opposition within a text. Formally, we consider signatures to be textual elements that throw focus onto certain aspects of a text. For instance, from a shallow perspective, the use of quotes or capitals which are often used to highlight a concept or an attribute (e.g." I HATE to admit it but, I LOVE admitting things"); while from a deeper perspective, the use of adverbs which communicate contradiction (or negation) in a text (e.g. "Saying we will destroy terrorism is *about* as meaningful as saying we shall annihilate mocking").

This layer is represented by three textual features: pointedness, counter-factuality, and temporal compression.

Pointedness is focused on detecting explicit marks which, according to the most relevant properties of irony (cf. Section 2.1), should reflect a sharp distinction in the information that is communicated. The set of textual elements that we considered here are punctuation marks (such as **.**, **. . .**, **;**, **?**, **!**, **:**, **,**), emoticons (such as :), ;), :-), :-o, ;-), :P, :¡)), as well as the use of quotes and capitalized words.

Counter-factuality is focused on implicit marks; i.e. discursive terms that hint at opposition or contradiction in a text, such as *about*, *nevertheless*, *nonetheless* or *yet*. In order to get a list with the main adverbs related to negation, as well as their synonyms, WordNet [28] resource was employed.

Temporal compression, is focused on identifying elements related to opposition in time; i.e. terms that suggest an abrupt change in the narrative sequence. These elements are represented by a set of temporal adverbs such as *suddenly*, *now*, or *abruptly*. The complete list of elements that we used to represent this layer can be downloaded from http://users.dsic.upv.es/grupos/nle.

4.2 Emotional Scenarios

Language, in all its forms, is one of our most natural and important means of conveying information about emotional states. Textual language provides specific tools on its own, such as the use of emoticons in web-based content to communicate information about moods, feelings, and our sentiments toward others. User-generated contents, for instance, often use such markers to accurately convey their communicative effects (e.g. "To err is human. To forgive for no good reason is plain stupid :P"). According to [13], the emotion of a sentence of text should be derived by composition of the emotions of the words in the sentence. In this respect, this layer is intended to capture information that goes beyond grammar, and beyond the positive or negative polarity of individual words. Rather, this layer attempts to characterize irony in terms of elements which symbolize abstractions such as overall sentiments, attitudes, feelings and moods, in order

to define a schema of favorable and unfavorable contexts for the expression of irony. Adopting a psychological perspective, we represent emotional scenarios in terms of the categories described in the Dictionary of Affect in Language [41][9]. They are activation, imagery, and pleasantness. Their respective descriptions are given below:

Activation refers to the degree of response, either passive or active, that humans exhibit in an emotional state (e.g. $burning$ is more active than $basic$).

Imagery tries to quantify how easy or difficult is to form a mental picture for a given word (e.g. it is more difficult to mentally depict $never$ than $alcoholic$).

Pleasantness measures the degree of pleasure suggested by a word (e.g. $love$ is more pleasant than $money$).

According to Whissell, such categories attempt to quantify the emotional content of words on the basis of scores obtained from human raters, who took into consideration their use in natural language scenarios.

4.3 Unexpectedness

The third layer is based on the premise that irony often exploits incongruity, unexpectedness and ridiculousness to ensure that an insincere text is not taken literally by a listener. Lucariello [27] suggests the term *unexpectedness* to represent the "imbalances in which opposition is a critical feature". She notes that surprise is a key component of irony, and even goes as far as to claim that unexpectedness underlies all ironic situations. Considering previous assumptions, we propose the *unexpectedness* layer as a mean to capture both temporal and contextual imbalances in an ironic text. According to Lucariello, these imbalances are defined in terms of oppositions or inconsistencies within contexts or situations, or between roles, or even across time-frames (e.g. "The wimp who grows up to be a lion tamer", or "A kiss that signifies betrayal"; cf. [27]). Assuming unexpectedness as an underlying element that permeates most of the ironic contents, we determined two features to represent this layer: temporal and contextual imbalance.

The former, **temporal imbalance** is used to reflect the degree of temporal opposition in a text with respect to the present and past tenses. Unlike temporal compression (see 4.1), here we are focusing on analyzing divergences related to verbs only (e.g. "I $hate$ that when you get a girlfriend most of the girls that $didn't$ want you all of a sudden $want$ you!"). To the end of obtaining the verbs as well as their tense, we used a public tool to label the documents in terms of POS tags. This tool is described in [2].

Contextual imbalance is intended to capture inconsistencies within a context. According to cognitive perspectives, context is often considered a crucial notion for understanding human communication. Moreover, in [18] it is pointed out that irony aptness should be sensitive to the amount of the disparity involved in its interpretation; i.e. context is essential to correctly interpret almost any statement, either literal or figurative. In

---

[9] This dictionary scores over 8,000 English words along its three categories. The range of scores goes from 1.0 (most passive, or most difficult to form a mental picture, or most unpleasant) to 3.0 (most active, or easiest to form a mental picture, or most pleasant). For instance, the item $flower$ is passive (activation = 1.0), easily representable (imagery = 3.0), and generally produces a pleasant affect (pleasantness = 2.75); whereas, $crazy$ is more active (1.33), moderately representable (2.16), and quite unpleasant (1.6).

order to represent this feature, we decided to estimate the semantic similarity of a text. If the semantic similarity is low, we consider that such context likely contains more imbalances in its narrative sequence[10]. Therefore, the probability of finding ironic content increases. Based on this hypothesis, we used a common semantic measure to estimate the similarity of concepts. The Resnik measure, implemented in WordNet::Similarity module [31], was then used to calculate the pair-wise semantic relatedness of all terms in a text. The contextual imbalance of a document is then calculated as the reciprocal of its semantic relatedness (that is, 1 divided by its semantic relatedness). The driving intuition here is: the smaller the semantic inter-relatedness of a text, the greater its contextual imbalance (suggesting an ironic text); the greater the semantic inter-relatedness of a text, the lesser its contextual imbalance (suggesting a non-ironic text).

5 Experiments and Results

All the experiments listed in this section were performed over the data described in Section 3.3. Apart from removing the stopwords, no other preprocessing was performed to the documents.

5.1 Experimental Set-up

The first phase consisted of representing the documents by means of the features previously described. To this end, we converted every single document into a vector of frequencies by applying Formula 1:

$$\delta(d_k) = \frac{\sum_{i,j} fdf_{i,j}}{|d|} \tag{1}$$

where i is the i-th conceptual layer ($i = 1\ldots3$); j is the j-th textual feature of i ($j = 1\ldots2$ for unexpectedness, and $1\ldots3$ otherwise); $fdf_{i,j}$ (*feature dimension frequency*) is the frequency of textual features j of layer i; and $|d|$ is the length of the k-th document d_k. For instance, the text "I love ugly people LIKE you :)" contains the elements $LIKE$ and :) which belong to pointedness; $love$ and $people$ which belong to pleasantness; and its contextual imbalance is 0.63. After applying Formula 1, we obtain a score of 4.63, which is then normalized relative to the length of the text (i.e. 7). Its δ, thus, is 0.66.

The following phase was focused on obtaining the **documents** with greater probability to have ironic content[11]. These documents were obtained by applying Formula 2:

$$\gamma(d_k) = \frac{\delta(d_k)}{tdf} \tag{2}$$

---

[10] A recent interesting approach considering semantic information is reported by [26], in which authors are focused on mining fine-grained information from queries by suggesting a semantic similarity-based clustering approach.

[11] This content must be understood on the basis of the definition given in Section 2.1.

where tdf (*total dimension of features*) is the number of textual features; i.e. $fdf = 8$. The underlying hypothesis is: the greater γ of document d_k, the greater is the probability to have ironic contents along the whole document[12].

According to the formula, the document with highest γ value per set was: document *cv270_6079.txt* (set movies2); document *cv173_tok-11316.txt* (set movies1); document 233 (set books), and document *training-1581.satire* (set articles).

The final phase consisted of obtaining the **sentences** more likely to be ironic. In this case, we reduced our scope to the 50 documents per set with highest γ value; i.e. 200 documents in total. To this end, we firstly split the 200 documents in isolated sentences. Then, after modifying one parameter, Formula 1 was applied. The modification lay on eliminating the highest and lowest values of i in order to avoid biased δ values. Finally, in order to identify the sentences with greater probability to be ironic[13], Formula 2 was applied. The 100 sentences with highest γ value were then considered to be ironic; i.e. 400 sentences in total.
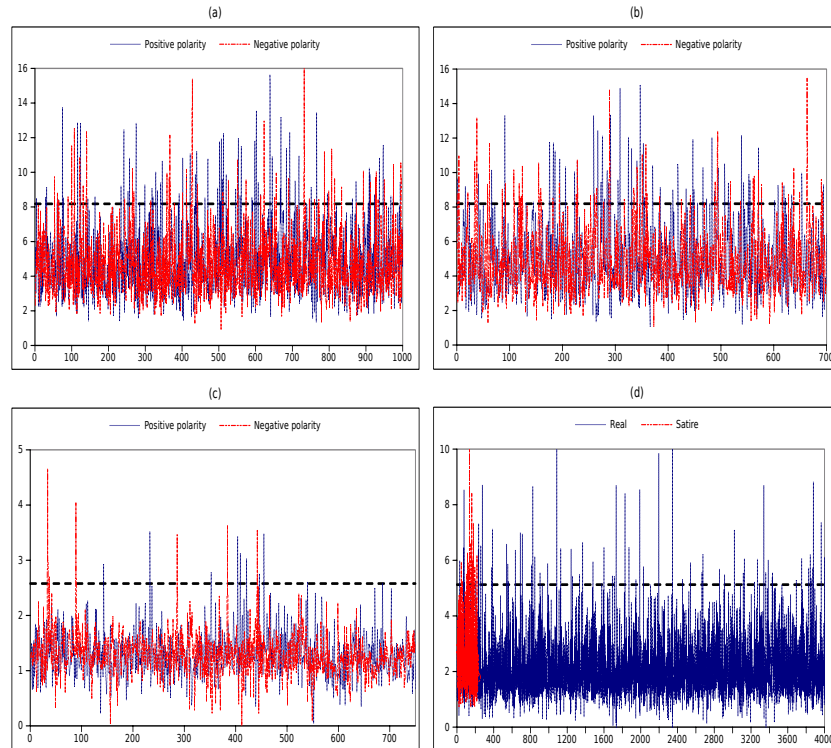

Fig. 1. γ values per set: $movies2$ (a); $movies1$ (b); $books$ (c); $articles$ (d).

---

[12] Note that we do not consider the whole document to be completely ironic. Instead, we highlight the possibility to have fragments or sentences that can be considered to be ironic.

[13] No matter if two or more sentences belong to a same document.

5.2 Results

Figure 1 shows the results after applying Formula 2. X axis represents every single document within its respective set. Y axis represents its γ value. The dotted line represents the minimum γ value after which a document is considered as potentially ironic. Such minimum γ value was determined by obtaining the mean between the highest and the lowest value of each set.

According to this figure, there are four facts to be highlighted:

i. The highest γ values are centered on the sets $movies2$ and $movies1$, then the set $articles$, and finally, the set $books$. This fact is correlated to the information given in Table 1. The amount of words per document drastically varies across the sets: from an average length of 787 words in the set $movies2$ to an average length of 57 words in the set $books$. Nonetheless, this variation does not affect the quality of the documents candidates to have ironic content due to the documents are normalized according to their length (cf. Formula 1).

ii. The more complex documents, discursively talking, are the ones of the sets movies. This is given by their length, which implies to follow more elaborate narrative sequences. Now then, focusing only on these sets, it is observable how the documents labeled with positive polarity are the ones in which the ironic content tends to often exceeds the minimum γ value. This behavior could provide some evidence about the presence of figurative content (either by using irony, sarcasm, satire, or even humor) when the speaker tries to consciously negate his/her literal words. According to this argument, the positive documents might be projecting a negative meaning in ground, completely different to the positive one profiled in the surface.

iii. Most documents, regardless of the set they belong to, do not exceed the minimum γ value. About 90% of documents, or even more (see graphic (c)), are far away from the minimum. This fact indicates that only very few documents might have ironic content. This is the expected situation because figurative content does not appear constantly; i.e. there is not a balanced distribution between literal and figurative content. For instance, for every 10 literal web comments, one might expect that 1 or 2 had figurative content. This is clearer when analyzing graphic (d). Despite there are only 233 satiric articles, most of them are close to the minimum γ values. Contrary situation with the real articles: they are 4,000 documents, but most are far away from the minimum.

iv. According to the last argument, and considering the set $articles$ (graphic (d)) a kind of gold standard because its data is labeled as satiric or real[14], it is evident that the model is representing underlying information linked to the way in which figurative content is expressed by people. Despite not all 233 documents exceed the minimum γ value, most of them steadily appear close to it (unlike the real documents). This is clearer when considering the 50 most ironic documents of this set: 34 documents belonged to the documents labeled with the tag satire; whereas the remaining 16 documents belonged to the ones labeled with the tag real. Thus, we might remark that the model is identifying elements that are commonly used to verbalize ironic content.

---

[14] Unlike the others sets which are only labeled with positive or negative polarity.

Finally, according to the experiments performed in 5.1, we transcribe in Appendix A some examples regarding the sentences with greater γ value.

6 Evaluation and Discussion

In the previous section we described the appropriateness or representativeness of different textual features to automatically detect ironic texts; i.e. find insights with respect to the ways in which users employ words and visual elements when speaking in a mode they consider to be ironic. In this section, in turn, we look for obtaining empirical judgments regarding the insights previously described. To this end, two external evaluations were performed. The first one consisted of assessing the 400 sentences with highest γ value by two human annotators[15]. They were asked to evaluate whether or not those sentences might profile an ironic sense. Apart from their own concept of irony, no theoretical background was requested or offered. All the sentences should be evaluated in isolation; i.e. their contexts were not provided. Each annotator evaluated 200 sentences (50 sentences per set). Furthermore, in order to estimate the degree of agreement between the annotators, the Krippendorff α coefficient was calculated. According to Artstein and Poesio [1], this coefficient calculates the expected agreement by looking at the overall distribution of judgments without regard to which annotator produces such judgments. Table 2 presents their evaluation, for which Krippendorff α coefficient of 0.490 was noted. The percentages of *approved* sentences were obtained by dividing the amount of sentences marked as ironic by the annotators, by the total of sentences evaluated (50 per set).

Table 2. Evaluation in terms of isolated sentences.

	Annotator 1	Percentage	Annotator 2	Percentage
$Movies2$	13	26%	18	36%
$Movies1$	12	24%	18	36%
$Books$	8	16%	6	12%
$Articles$	17	34%	24	48%

The second evaluation consisted of assessing the same sentences alongside the whole document they belong to. Thus, each annotator had to evaluate 25 documents per set. After reading the whole documents, they had to decide whether or not: i) the document was completely ironic; ii) the document contained any fragment (sentence or phrase) which may be considered as ironic. In this case, apart from their own concept of irony, we provided our definition of irony stated in Section 2.1. Table 3 presents their evaluation, for which Krippendorff α coefficient of 0.717 was noted. The percentages of *approved* documents were now obtained by dividing the amount of documents marked as ironic by the total of documents evaluated (25 per set).

---

[15] Both annotators are bilingual and they work as English-Spanish translators.

Table 3. Evaluation in terms of whole documents.

	Annotator 1			Annotator 2		
	i) Document	ii) Fragment	Percentage	i) Document	ii) Fragment	Percentage
*Movies*2	Not	16	64%	Not	14	56%
*Movies*1	Not	22	88%	Not	20	80%
Books	Yes	2	8%	Not	1	4%
Articles	Not	24	96%	Not	22	88%

According to the information depicted in both tables, we can infer the following facts:

  i. The results given in Table 2 are quite poor. Each annotator evaluated 50 sentences per set, and the highest value achieved is 48%; i.e. less than half of them would be ironic. Results show that the problem of automatically classifying sentences as ironic is very challenging. For instance, it is completely senseless that only 6 of 50 sentences (the worst result) may come to be regarded as ironic when the purpose is just the contrary. Considering the sentences that are supposed to be more likely regarded as ironic (due to they come from the documents labeled as satiric of the *articles* set), the evaluation evidences that the model has difficulty in identifying sentences which leave no doubt with respect to their ironic ground to any human.

 ii. Based on annotators' comments, it is also evident that, except in very clear cases, an isolated sentence is not sufficient to correctly decide whether or not that sentence is ironic. After manually analyzing some of these sentences, we could realize how hard is to figure out what is their ground meaning, especially, because of the lack of context. In absence of elements to map the information provided by the sentence, the fact of considering a sentence as ironic is almost a random process. For instance, the sentence *"I never believed love at first site was possible until I saw this film"* could project both an ironic as a positive meaning. Similarly, the sentence *"The plot, with its annoying twists, is completely inane"* could be profiling both a negative as an ironic meaning. This is a conceptual problem that points to the question stated in Section 3.1 about the difficulty of automatically differentiating between literal and figurative language. If the context is not accessible to the annotator, this will hardly have elements to appreciate the existence of ironic content on the basis of an isolated sentence. Therefore, his/her evaluation will mostly depend on grammatical issues, which leave no room to figurative interpretations.

iii. The second evaluation was performed on the basis of these issues: if isolated sentences are not sufficient to determine the existence of ironic content, then we should try with entire documents. In this case, the results given in Table 3 show a clear improvement. Despite the results are not excellent (consider that only 1 document of 200 was regarded to be completely ironic[16], as well as the very low percentage of ironic content with respect to the documents belonging to the set *books*), it is evident how, when considering the whole document instead of isolated sentences, the

---

[16] It is unlikely to expect more ironic documents in these data sets because they were not compiled with the purpose of finding irony.

spectrum to really appreciate irony clearly increased: 96% and 88% in the documents belonging to the set *articles*, as well as 88% and 80% in the documents of the set *movies*1. This fact shows the need of considering context and information beyond grammar for tasks such as this one. By examining the entire documents, the annotators are able to access to very valuable information, which makes sense as a whole, thereby achieving a complete overview of the meanings profiled. The consequence: annotators now have elements to adequately judge whether or not ironic content exists in the documents suggested by the model. Perhaps the participation of the experts to evaluate the results will increase: it is quite different to evaluate just a few sentences than entire documents, but it is also different to evaluate only some documents guided by the presence of such sentences than evaluating a complete data set.

iv. By providing our definition of irony to the annotators, the scope of documents with ironic content substantially increased. This directly impacts on the scenario of applicability: a sentiment analysis task (cf. Section 3.1). According to the arguments given in Section 2.1, except in prototypical examples, the boundaries to correctly separate figurative phenomena are quite fuzzy. This is clearer when dealing with user-generated contents in which people mix ironic remarks with observations about ironic, sarcastic or even funny situations[17]; i.e. polarity depends on factors beyond the semantic of the words. If we intend to find out the underlying polarity of any document, we must spread the spectrum of phenomena related to the topic we are interested. By considering phenomena related to irony (e.g. sarcasm and satire, which in many cases are considered part of it, or subclasses), the annotators had more elements, besides the context, to correctly make their decision.

v. Going deeper into this point, the results depicted in Table 3 show some very interesting facts: the amount of documents with ironic content is 60.5% (121 of 200). 69 of them belong to documents labeled with the positive polarity tag (documents labeled with the satiric tag are also considered here); whereas the remaining 52 belong to the ones labeled with the negative polarity tag (documents labeled with the real tag are considered here). This means that ironic content does not always occur in the documents in which it is supposed to; i.e. irony should occur quite often in the documents labeled with the positive polarity tag due to their main underlying aim is to produce an effect that denies their surface information. Now, when considering others kinds of effects (funny, disrespectful, sarcastic, etc.), the spectrum of sources to find ironic content increases. In this case, the definition provided to the annotators allowed them to access to others sources in which the figurative content profiled **negative connotations**, regardless of such content appears in a document labeled as positive or negative. On the basis of a sentiment analysis task, this approach might be useful to provide categorized information based on the effects produced by irony.

vi. We have already noticed that the results of the evaluation are not as we have expected, in particular, with respect to the set *books*. Apart from the arguments given previously, there are many other reasons that may explain them. For instance, there is not ironic content at all; most documents are concise, thus irony is seldom used;

---

[17] Cf. [34] regarding the presence of irony to produce funny effects in user-generated contents.

isolated sentences may evidence an underlying ironic meaning, but when they are putted into their contexts, this meaning does not correspond with the general meaning profiled by the author[18]; irony is a subjective phenomenon which varies according to people; perhaps not all the annotators are capable to find out pragmatic phenomena such as inferences, assumptions, implicatures and so on. Thus, they were not capable to find out the trigger of the ironic effect.

vii. Finally, as described in Section 2.1, irony implies the negation of polarity. Therefore, its detection is not a trivial task, but a real challenge. The proposed model should thus be tested in the near future on *ad hoc* data sets for irony detection. Unfortunately, compiling such a data set will be a challenge itself because of the subjectivity of truly determining, beyond prototypical examples, what an ironic instance is.

7   Conclusions and Further Work

Negation is a grammatical category that allows changing the truth value of a proposition. Its automatic processing is important not only for sentiment analysis and opinion mining, but also for many other natural language processing tasks such as question answering, textual entailment, or even for analyzing collective and social behavior ([4, 22]). If automatic negation processing is already quite complex when dealing with literal language, it becomes even more difficult and challenging when dealing with figurative language. In this respect, we have shown that irony is a sophisticated, subtle and ambiguous way of communication, whose main characteristic points to negate the surface meaning of what is communicated. Due to the presence of different narrative strategies, such as tone, obviousness, or funniness, as well as to the absence of a negation marker, its computational treatment seems to be nearly impossible. Moreover, note the difficulty of detecting irony at textual level when valuable information, such as the tone employed to trigger the ironic effect, is not available.

In this article we have suggested a new model which attempts to identify salient characteristics of irony. According to our definition of this phenomenon (Section 2.1), we have established a model to represent verbal irony in terms of three conceptual layers: signatures, emotional scenarios, and unexpectedness. These layers comprise eight different textual features that intend to symbolize low-level and high-level properties of irony. No single textual feature captures the essence of irony, but all eight kinds together provide a valuable linguistic inventory for this task. Due to the lack of data sets compiled specifically for irony detection, we used four data sets already employed in tasks related to sentiment analysis in order to evaluate or model. Two kinds of results were obtained: isolated sentences and entire documents. Such results were assessed by two annotators on two key strata: i) determining whether or not the sentences could be regarded as ironic only on the basis of the information provided by the sentence itself; ii) determining whether or not, by considering also the context of each sentence, the documents to which they belong could be regarded as being completely ironic or

---

[18] Specifically, note that most sentences of the set *books*, which were considered to be ironic by the annotators, when re-analyzing within their contexts acquire their real meaning: they are only positive or negative utterances without any figurative sense.

having ironic content. Despite the two evaluations showed some model weakness, in particular with respect to the first stratum (it is quite hard to perceive irony on the basis only of a sentence which belongs to a whole narrative). It is necessary to stress, however, that according to the evaluations obtained in the second stratum (when taking into consideration the context), the capabilities to correctly determine the presence of irony in the documents substantially increased. Finally, the results provide interesting insights into the figurative issues regarding tasks in which underlying knowledge like this could represent valuable information.

Further work consists of improving the quality of textual features, as well as providing new ones (mainly on the basis of pragmatic phenomena), in order to come up with an improved model capable to detect better ironic patterns in different kinds of texts. Last, but not least, the new model should be finally tested on a specific data set for irony detection, whose compilation implies a challenge itself because of the subjectivity of determining irony at textual level.

Appendix A: Sample of the most ironic sentences

According to our model predictions, here are presented some of the most ironic sentences. Each sentence has a document identifier. Such identifiers were kept as in the original data sets in order to facilitate their location.

1. $Movies2$
   - "Expecting them to give the viewer insights into the human condition is like expecting your car to vacuum your house " (doc. id. *cv116_28942.txt*).
   - "That degree of complexity combined with those very realistic looking dinosaur effects is just about as much as I require" (doc. id. *cv116_28942.txt*).
   - "Moulin Rogue is an original, and an original, even a flawed one, is a thing to be cherished" (doc. id. *cv275_28887.txt*).
   - "In some respects, Rush Hour is the ultimate exercise in cliched filmmaking. The hero is the renegade cop that prefers to work alone. The cop in question cannot solve the case until he gets in trouble. All chinese people are somehow involved in the criminal element. The duo must always be completely mismatched. The hero has to say some smart-assed comment before (and after) shooting someone. However, that doesn't necessarily make for a bad film" (doc. id. *cv402_14425.txt*).
   - "Making her dramatic debut, after appearing in over 300 triple X adult films, porn star Nina Hartley takes command of her role with considerable assurance and a screen presence which puts many other contemporary 'straight' actresses to shame" (doc. id. *cv422_9381.txt*).
   - "If I'd laughed any more, I might have needed an iron lung" (doc. id. *cv507_9220.txt*).
   - "I never believed love at first site was possible until I saw this film" (doc. id. *cv513_6923.txt*).
   - "Usually a movie is about something more than a soiled rug" (doc. id. *cv718_11434.txt*).
   - "I remember really enjoying this movie when I saw it years ago. I guess my memory really sucks" (doc. id. *cv982_22209.txt*).

- "It's not that there isn't anything positive to say about the film. There is. After 92 minutes, it ends". (doc. id. *cv123_12165.txt*).
- "There's an enormous woman (played by transvestite porn star)" (doc. id. *cv142_23657.txt*).
- "However, isn't bad at all. The actors do the best they can with the bad material" (doc. id. *cv733_9891.txt*).

2. *Movies*1
    - "The only actor in the movie with any demonstrable talent is a cute little prairie dog named Petey" (doc. id. *cv039_tok-11790.txt*).
    - "This film needed that whole theatre-shaking: they needed to wake everybody up because they were so bored" (doc. id. *cv229_tok-9484.txt*).
    - "Appreciate this movie for the few weeks it will be in theaters folks" (doc. id. *cv342_tok-24681.txt*).
    - "I hated this movie for every second that I sat watching it, and I actively hate it now, days later, with the simpering, superficial, nauseatingly sentimental images forever plaguing my memories" (doc. id. *cv352_tok-15921.txt*).
    - "It's too trashy to be good drama, but too dramatic to be good trash" (doc. id. *cv494_tok-11693.txt*).
    - "I only wish that I could make that one hour and forty-five minutes of my life re-appear" (doc. id. *cv495_tok-18551.txt*).
    - "In order to make the film a success, all they had to do was cast two extremely popular and attractive stars, have them share the screen for about two hours and then collect the profits" (doc. id. *cv176_tok-15918.txt*).
    - "(Why, oh why, couldn't Lucas use computers to substitute better performers in the lead roles?)" (doc. id. *cv228_tok-8817.txt*).
    - "Nostalgia appears to have a great appeal, but don't you think we could have more than 14 years before we yearn for the past?" (doc. id. *cv173_tok-11316.txt*).
    - "The weak scenes could have been cut, but then there wouldn't have been much left" (doc. id. *cv198_tok-11090.txt*).
    - "It's not a silent movie; there is lots of atmospheric music, occasional screams and weird sound effects, but nobody ever utters an audible word; unfortunately, is so bad that it's really bad" (doc. id. *cv524_tok-20616.txt*).
    - "It seems that comedy is the main motive, and the violence is only intended to punctuate the laughs. Unfortunately, there are no laughs" (doc. id. *cv680_tok-12227.txt*).

3. *Books*
    - "Essentially the entire plot can be summarised in a sentence of two, girl falls in love with boy, girl becomes damsel in distress, boy saves girl, end of....." (doc. id. *document 017 Negative*).
    - "Yes that literally is the entire plot, but far worse than this is the complete lack of intelligent character design" (doc. id. *document 017 Negative*).
    - "Harry goes to Hogwarts, bad guys try to kill Harry, battle with the bad guys, Harry triumphs - hurrah!" (doc. id. *document 043 Negative*).
    - "In fact I could see myself possibly enjoying this book ten years ago. Than again, maybe not" (doc. id. *document 108 Negative*).

4. *Articles*

- "As we examine the passengers' cell-phone calls and flight recordings, we get a sense of the incredible courage displayed by these ordinary men and women" (doc. id. *014-test-0153.satire*).
- "Despite years of diplomatic stalemate in the Mideast crisis, Syrian officials appeared eager to mend troubled Arab-Israeli relations this week by participating in a second round of U.S.-led peace talks, which feature representatives from every country in the region, as well as a complimentary continental breakfast in the hotel lobby" (doc. id. *016-test-0165.satire*).
- "Unfortunately, most of the men and women who passed by seemed to speak only a bizarre Asian dialect unknown to me, and those who could communicate were more interested in selling me exotic cologne out of a duffel bag" (doc. id. *022-test-0294.satire*).
- "This is merely about improving liquidity, said King" (doc. id. *095-test-1483.satire*).
- "Virtually free, except for digging, pumping, processing, storage, by-product-disposal and shipping costs" (doc. id. *179-training-1407.satire*).
- "Maybe the one person who allowed Bush to ignore the opinions of 45 percent of America has a busy schedule" (doc. id. *144-training-0769.satire*).

References

1. R. Artstein and M. Poesio. Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4):555–596, December 2008.

2. J. Atserias, B. Casas, E. Comelles, M. González, L. Padró, and M Padró. Freeling 1.3: Syntactic and semantic services in an open-source nlp library. In *Proceedings of the 5th International Conference on Language Resources and Evaluation*, pages 48–55, 2006.

3. S. Attardo. Irony as relevant inappropriateness. In R. Gibbs and H. Colston, editors, *Irony in Language and Thought*, pages 135–174. Taylor and Francis Group, 2007.

4. S. Banerjee and N. Agarwal. Analyzing collective behavior from blogs using swarm intelligence. *Knowledge and Information Systems*, 2012. DOI: DOI: 10.1007/s10115-012-0512-y.

5. G. Beydoun and A. Hoffmann. Dynamic evaluation of the development process of knowledge-based information systems. *Knowledge and Information Systems*, 2012. DOI:10.1007/s10115-012-0491-z.

6. C. Burfoot and T. Baldwin. Automatic satire detection: Are you having a laugh? In *ACL-IJCNLP '09: Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pages 161–164, 2009.

7. P. Carvalho, L. Sarmento, M. Silva, and E. de Oliveira. Clues for detecting irony in user-generated contents: oh...!! It's "so easy" ;-). In *TSA '09: Proceeding of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*, pages 53–56, Hong Kong, China, November 2009. ACM.

8. H. Clark and R. Gerrig. On the pretense theory of irony. *Journal of experimental psychology: General*, 113(1):121–126, 1984.

9. H. Colston. On necessary conditions for verbal irony comprehension. In R. Gibbs and H. Colston, editors, *Irony in Language and Thought*, pages 97–134. Taylor and Francis Group, 2007.

10. H. Colston and R. Gibbs. A brief history of irony. In R. Gibbs and H. Colston, editors, *Irony in Language and Thought*, pages 3–24. Taylor and Francis Group, 2007.

11. C. Curcó. Irony: Negation, echo, and metarepresentation. In R. Gibbs and H. Colston, editors, *Irony in Language and Thought*, pages 269–296. Taylor and Francis Group, 2007.

12. D. Davidov, O. Tsur, and A. Rappoport. Semi-supervised recognition of sarcastic sentences in Twitter and Amazon. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, CoNLL '10, pages 107–116, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics.

13. V. Francisco, P. Gervás, and F. Peinado. Ontological reasoning for improving the treatment of emotions in text. *Knowledge and Information Systems*, 24(2):23, 2010.

14. R. Gibbs. Irony in talk among friends. In R. Gibbs and H. Colston, editors, *Irony in Language and Thought*, pages 339–360. Taylor and Francis Group, 2007.

15. R. Gibbs and H. Colston. The future of irony studies. In R. Gibbs and H. Colston, editors, *Irony in Language and Thought*, pages 339–360. Taylor and Francis Group, 2007.

16. R. Giora. On irony and negation. *Discourse Processes*, 19(2):239–264, 1995.

17. R. Giora, N. Balaban, O. Fein, and I. Alkabets. Negation as positivity in disguise. In H. Colston and A. Katz, editors, *Figurative language comprehension: Social and cultural influences*, pages 233–258. Hillsdale, NJ: Erlbaum, 2005.

18. R. Giora, S. Federman, A. Kehat, O. Fein, and H. Sabah. Irony aptness. *Humor*, 18:23–39, 2005.

19. H. Grice. Logic and conversation. In Peter Cole and Jerry L. Morgan, editors, *Syntax and semantics*, volume 3, pages 41–58. New York: Academic Press, 1975.

20. L. Horn and Y. Kato. Introduction: Negation and polarity at the millennium. In Horn and Kato, editors, *Studies in Negation and Polarity*, pages 1–19. Oxford University Press, 2000.

21. B. Kaup, J. Lüdtke, and R. Zwaan. Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics*, 38:1033–1050, 2006.

22. S. Kisilevich, C. S. Ang, and M. Last. Large-scale analysis of self-disclosure patterns among online social networks users: A Russian context. *Knowledge and Information Systems*, 2011. DOI: 10.1007/s10115-011-0443-z.

23. R. Kreuz. Using figurative language to increase advertising effectiveness. In *Office of Naval Research Military Personnel Research Science Workshop*, Memphis, TN, 2001. University of Memphis.

24. S. Kumon-Nakamura, S. Glucksberg, and M. Brown. How about another piece of pie: The allusional pretense theory of discourse irony. In R. Gibbs and H. Colston, editors, *Irony in Language and Thought*, pages 57–96. Taylor and Francis Group, 2007.

25. R. Langacker. *Concept, Image and Symbol. The Cognitive Basis of Grammar*. Mounton de Gruyter, 1991.

26. J. Liu and K. Wang. Anonymizing bag-valued sparse data by semantic similarity-based clustering. *Knowledge and Information Systems*, 2012. DOI: 10.1007/s10115-012-0515-8.

27. J. Lucariello. Situational irony: A concept of events gone away. In R. Gibbs and H. Colston, editors, *Irony in Language and Thought*, pages 467–498. Taylor and Francis Group, 2007.

28. G. Miller. Wordnet: A lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.

29. B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 79–86, Morristown, NJ, USA, 2002. Association for Computational Linguistics.

30. Bo Pang and Lillian Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the ACL*, pages 271–278, 2004.

31. T. Pedersen, S. Patwardhan, and J. Michelizzi. Wordnet::similarity - measuring the relatedness of concepts. In *Proceeding of the 9th National Conference on Artificial Intelligence*

(AAAI-04), pages 1024–1025, Morristown, NJ, USA, 2004. Association for Computational Linguistics.

32. A. Reyes and P. Rosso. Linking humour to blogs analysis: Affective traits in posts. In *Proceedings of the 1st Workshop on Opinion Mining and Sentiment Analysis (WOMSA), CAEPIA-TTIA Conference*, pages 100–109, Sevilla, Spain, 13 November 2009.

33. A. Reyes and P. Rosso. Making objective decisions from subjective data: Detecting irony in customers reviews. *Decision Support Systems*, In press. DOI: 10.1016/j.dss.2012.05.027 `http://dx.doi.org/10.1016/j.dss.2012.05.027`.

34. A. Reyes, P. Rosso, and D. Buscaldi. Humor in the blogosphere: First clues for a verbal humor taxonomy. *Journal of Intelligent Systems*, 18(4):311–331, 2009.

35. L. Sarmento, P. Carvalho, M. Silva, and E. de Oliveira. Automatic creation of a reference corpus for political opinion mining in user-generated content. In *TSA '09: Proceeding of the 1st international CIKM workshop on Topic-Sentiment Analysis for mass opinion*, pages 29–36, Hong Kong, China, November 2009. ACM.

36. D. Sperber and D. Wilson. On verbal irony. *Lingua*, 87:53–76, 1992.

37. O. Tsur, D. Davidov, and A. Rappoport. ICWSM – A great catchy name: Semi-supervised recognition of sarcastic sentences in online product reviews. In William W. Cohen and Samuel Gosling, editors, *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, pages 162–169, Washington, D.C., 23-26 May 2010. The AAAI Press.

38. A. Utsumi. A unified theory of irony and its computational formalization. In *Proceedings of the 16th conference on Computational Linguistics*, pages 962–967, Morristown, NJ, USA, 1996. Association for Computational Linguistics.

39. T. Veale and Y. Hao. Support structures for linguistic creativity: A computational analysis of creative irony in similes. In *Proceedings of CogSci 2009, the 31st Annual Meeting of the Cognitive Science Society*, pages 1376–1381, 2009.

40. T. Veale and Y. Hao. Detecting ironic intent in creative comparisons. In *Proceedings of 19th European Conference on Artificial Intelligence - ECAI 2010*, pages 765–770, Amsterdam, The Netherlands, 2010. IOS Press.

41. C. Whissell. Using the revised dictionary of affect in language to quantify the emotional undertones of samples of natural language. *Psychological Reports*, 105(2):509–521, 2009.

42. D. Wilson and D. Sperber. On verbal irony. In R. Gibbs and H. Colston, editors, *Irony in Language and Thought*, pages 35–56. Taylor and Francis Group, 2007.

43. T. Zagibalov, K. Belyatskaya, and J. Carroll. Comparable English-Russian book review corpora for sentiment analysis. In *Proceedings of the 1st Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, pages 67–72, Lisbon, Portugal, 2010.