

Document downloaded from:

<http://hdl.handle.net/10251/35314>

This paper must be cited as:

Reyes Pérez, A.; Rosso, P.; Buscaldi, D. (2012). From humor recognition to Irony detection: The figurative language of social media. *Data and Knowledge Engineering*. 74:1-12. doi:10.1016/j.datak.2012.02.005.



The final publication is available at

<http://dx.doi.org/10.1016/j.datak.2012.02.005>

Copyright Elsevier

# From Humor Recognition to Irony Detection: The Figurative Language of Social Media

Antonio Reyes<sup>a,1,\*</sup>, Paolo Rosso<sup>a</sup>, Davide Buscaldi<sup>b</sup>

<sup>a</sup> *Natural Language Engineering Lab - ELiRF  
Departamento de Sistemas Informáticos y Computación  
Universidad Politécnica de Valencia  
Camino de Vera, s/n 46022, Valencia, Spain*

<sup>b</sup> *Institut de Recherche en Informatique de Toulouse (IRIT)  
Université Paul Sabatier Campus, 118 Route de Narbonne  
F-31062 Toulouse, France*

---

## Abstract

The research described in this paper focuses on analyzing two playful domains of language: humor and irony, in order to identify key values components for their automatic processing. In particular, we focus on describing a model for recognizing these phenomena in social media, such as “tweets”. Our experiments are centered on five data sets retrieved from Twitter taking advantage of user-generated tags, such as “#humor” and “#irony”. The model, which is based on textual features, is assessed on two dimensions: representativeness and relevance. The results, apart from providing some valuable insights into the creative and figurative usages of language, are positive regarding humor, and encouraging regarding irony.

**Keywords:** Humor Recognition, Irony Detection, Natural Language Processing, Web Text Analysis

---

---

\*Corresponding author

*Email addresses:* areyes@dsic.upv.es (Antonio Reyes), proso@dsic.upv.es (Paolo Rosso), dbuscaldi@acm.org (Davide Buscaldi)

<sup>1</sup>Permanent address: Esfuerzo 72. Depto. 4501. C.P. 07530, Mexico, D.F. Telephone: +52 55 59121923

## 1. Introduction

Figurative language is one of the most arduous topics facing natural language processing (NLP). Unlike literal language, the former takes advantage of linguistic devices, such as metaphor, analogy, ambiguity, irony, and so on, in order to project more complex meanings which, usually, represent a real challenge, not only for computers, but for humans as well. This is the case of humor and irony. Each one exploits different linguistic strategies to be able to produce an effect (e.g., ambiguity and alliteration regarding humor [29, 26]; similes regarding irony [40]). Sometimes the strategies are similar (e.g., usage of satirical or sarcastic utterances to express a negative attitude [22, 3]).

These devices entail cognitive capabilities to abstract and meta-represent meanings beyond “physical” words. In this communicative layer, communication is more than sharing a common code, but being capable of inferring information beyond syntax or semantics; i.e. figurative language implies information not grammatically expressed to be able to decode its underlying meaning: if this information is not unveiled, the real meaning is not accomplished and the figurative effect is lost. Let us consider a joke. The amusing effect sometimes relies on not given information. If such information is not filled, the result is a bad, or better said, a misunderstood joke. This information entails a great challenge because it points to social and cognitive layers that are quite difficult to be computationally represented. However, despite the inconveniences that figurative language entails, the approaches to automatically process figurative devices, such as humor, irony or sarcasm, seem to be largely encouraging. For instance, the research works focused on automatic humor generation [7, 36] and humor recognition [26, 25, 31]; as well as the ones focused on irony detection [39, 40, 10], satire detection [9], and sarcasm detection [38], have shown the feasibility of computationally dealing with figurative language. In addition, it is important to highlight the relevance of taking into consideration the scopes that such investigations might represent facing scenarios, such as edutainment, advertising, sentiment analysis, trends discovery, and so on.

In this framework, this paper aims at showing how two specific domains of figurative language —humor and irony, may be automatically handled by means of considering linguistic devices, such as ambiguity and incongruity, and meta-linguistic devices, such as polarity and emotional scenarios. We especially focus on discussing how underlying knowledge, which relies on shallow and deep linguistic layers, may represent relevant information to automatically identify figurative uses of language. In particular, and contrary to most of researches that deal

with figurative language, we aim at identifying figurative uses regarding language in social media. This means that we are not focused on analyzing prototypical jokes nor literary examples of irony, rather, we try to find patterns in texts whose intrinsic characteristics are quite different to the characteristics described in the specialized literature. For instance, a joke which exploits phonetic devices to produce a funny effect, and a tweet in which humor is self-contained in the situation. Considering this scenario, we suggest a set of features which work together as a system: no single feature is particularly humorous or ironic, but all together provide a useful linguistic inventory for detecting humor and irony at textual level.

The paper is organized as follows. In Section 2 we underline the theoretical issues which underlies humor and irony. In Section 3 we describe the related work on the role of automatic figurative language processing. In Section 4 we detail the set of features. In Section 5 we evaluate the features effectiveness. In Section 6 the results and their implications are discussed. Finally, in Section 7 we draw some final remarks and address the future work.

## **2. Theoretical Issues**

### *2.1. Humor*

One of the characteristics which defines us as human beings and social entities is a very complex, as well as very common concept: humor. This concept, which we could simply define by the presence of amusing effects, such as laughter or well-being sensations, plays a relevant role in our lives. Humor's main function is to release emotions, sentiments or feelings, positively impacts on human health. In a social context, humor's cathartic properties make most people react to a humorous stimulus regardless of their beliefs, social status or cultural differences. Moreover, humor provides valuable information related to linguistic, psychological, neurological and sociological phenomena. However, given its complexity, humor is still an undefined phenomenon. Partly, because the stimuli that make people laugh can hardly be generalized or formalized. For instance, cognitive aspects as well as cultural knowledge, are some of the multi-factorial variables that should be analyzed in order to understand humor's properties. Despite these inconveniences, different disciplines such as philosophy [19], linguistics [1], psychology [32], or sociology [20], have attempted to study humor in order to provide formal insights to explain better its basic features. On one hand, from a psychological point of view, Ruch [32] has analyzed the relationship between personality and humor appreciation, providing interesting observations about this perspective, and about the type of necessary stimuli required to produce a response. On the

other hand, some linguistic studies have tried to explain humor by means of semantic and pragmatic patterns. Attardo [1, 2] tries to explain verbal humor<sup>2</sup> as a phenomenon that suggests the presence of some *knowledge resources*, such as language, narrative strategies, target, situation, logical mechanisms or opposition, in order to produce a funny effect. Finally, from a sociological point of view, the most studied features regarding humor appreciation are cultural patterns. Hertzler [20] stresses the importance of analyzing the cultural background in order to categorize humor as an entire phenomenon.

## 2.2. Irony

Like most figurative devices, irony is difficult to be defined in formal terms. According to Wilson and Sperber [43], irony is basically a communicative act that expresses the opposite of what is literally said. Experts consider two types of irony: verbal irony and situational irony. Most theories concur that verbal irony communicates an opposite meaning; i.e. a speaker says something that seems to be the opposite of what he/she means [12]. Situational irony, in contrast, is an unexpected or incongruous property in a situation or event; i.e. situations that are just not meant to be [24]. In fine-grained approaches, there are authors that distinguish other types of ironies: dramatic irony [3]; discourse irony [22]; tragic irony [11]; etc. Our main interests in this work is focused on verbal irony. We think that there are some features given in situations where humor and irony are implied that are worth analyzing.

Most studies about verbal and situational irony have a linguistic approach, unlike humor. Concerning verbal irony, literature suggests a prototypical characteristic, it intentionally denies what is literally expressed [13], i.e. an indirect negation [17]. Based on some pragmatic contexts, Grice [18] considers that an expression is ironic, if and only if, it intentionally violates any conversational maxim. Wilson and Sperber [43] assume that verbal irony must be understood as echoic, i.e. as a distinction between use and mention. Utsumi [39] proposes an ironic environment that causes a negative emotional attitude as a requirement to consider an expression as ironic. According to the above mentioned perspectives, these suggest different forms of explaining the underlying concept of opposition in order to conceive as verbal expression as ironic. At this point, it is important to highlight that most of ironic expressions used in real life confirm the mean-

---

<sup>2</sup>Verbal humor refers to the kind of humor that is generated by linguistic strategies, i.e. by language. Our work is focused on this particular type of humor.

ing of irony as the assumption of an opposite meaning from what is said, without considering any pragmatic rules.

Another important characteristic of verbal irony is related to the slight boundary in meaning from irony, sarcasm, and satire. On one hand, Colston [11] considers sarcasm as a term that is commonly used to describe an expression of verbal irony. Gibbs [15] points out that sarcasm combined with devices such as jocularity, hyperbole, rhetorical questions, and understatement, are types of irony; whereas in [22], authors argue about a type of sarcastic irony that is opposed to the non sarcastic one. On the other hand, Gibbs and Colston [16] suggest that irony is often compared to satire and parody. According to this point of view, the only characteristic that seems to be recurrent, is an underlying negative meaning that varies in intensity. Finally, Dews et al. [14] state that irony plays different social functions, considering humor as one of them. Based on their experiments, a funny meaning is quite often considered as fundamental for ironic expressions.

Considering previous investigations, it is our interest to deeply analyze the terms humor and irony, in order to detect if there are any elements in common that may allow us to represent the basic features of both phenomena.

### 3. Automatic Figurative Language Processing

The interest for automatically processing matters related to figurative language is not new in NLP. Particularly, considering the approaches related to automatic humor processing, we can divide them in two areas: generation and recognition. The aim of the former is to study lexical features which could be formalized in order to simulate their patterns and generate a funny effect. Researches in [6, 7] have shown the importance of these forms of patterns, especially on the basis of phonetic and syntactic information, for automatically generating humor. The example in 1 illustrates the type of linguistic elements that underlie humor:

1. “What do you use to talk to an elephant? An elly-phone” [6].

In this sentence we can see how structural features, codified through linguistic information, are used to automatically generate a text with funny connotations. Note how *elly – phone* has a phonological similarity with telephone. Moreover, *elly – phone* is related, phonologically and “semantically”, to the word which gives its right meaning: elephant. This form of funny question answering structure —punning riddles, takes advantage of linguistic patterns in order to produce an amusing effect. Characteristics of a more complex nature have been also studied to represent and generate funny patterns. The findings reported in [36]

demonstrated how incongruity and opposite concepts are important elements for producing funny senses. By means of combining words, which in social terms represent opposite referents, authors have automatically produced new funny senses for acronyms such as MIT (Massachusetts Institute of Technology):

2. “Mythical Institute of Theology”.

With respect to humor recognition, most of research has focused on the analysis of particular funny structures: one-liners. These short structures are syntactically very simple, and the humorous effect relies on more complex features. Consider the example in 3:

3. “Infants don’t enjoy infancy like adults do adultery”.

This sentence contains phonological information which helps to produce humor, but this is not everything. There is also a pun that plays an oppositional role between the meanings of the words. Together, they produce a funny result. These types of surface elements have provided evidence for characterizing humor in terms of formal components (which may automatically be recognized). For instance, Mihalcea and Strapparava [26] applied machine-learning techniques to identify humorous patterns in one-liners. Some of the elements they reported are alliteration, antonymy or adult slang. In addition, they suggested semantic spaces which are triggers of humor: human centric vocabulary (example 4), negative orientation (example 5), and professional communities (example 6):

4. “Of all the things **I** lost, **I** miss **my** mind the most”.
5. “Money **can’t** buy your friends, but you do get a better class of **enemy**”.
6. “It was so cold last winter that I saw a **lawyer** with his hands in his own pockets”.

Moreover, the work of Sjöbergh and Araki [35] has focused on finding patterns in syntactic and semantic layers. According to their results, devices such as similarity, style or idiomatic expressions, are sources in which humor tends to appear. When considering bigger structures such as news articles or blogs, the research of Mihalcea and Pulman [25] has evidenced how negative polarity is a very important factor to recognize humor. In contrast, the research of Reyes et al. [31] has shown the role of semantic ambiguity and keyness to characterize funny blogs. Furthermore, the evaluation described in [30] has supported the relevance of some outstanding features, such as affective and emotional content, to automatically retrieve funny web comments.

Finally, according to the arguments stated in the previous section, some of the conclusions regarding the role of humor beyond surface features point the relevance of incongruity, idiomatic expressions, common sense knowledge, ambiguity, and irony as sources to investigate deeper features (see [26]).

On the other hand, the computational approaches which deal with more abstract uses of figurative language tend to be more restricted. However they are current hot topics in NLP due to the advances in areas such as sentiment analysis and opinion mining, trends discovery, or electronic commerce. For instance, regarding automatic irony processing, the research described by Utsumi [39] was one of the first attempts to computationally formalize irony. His model suggested an idealized hearer-listener interaction; therefore, it was too abstract to be implemented in a computational framework. Veale and Hao [40], in contrast, analyzed a large quantity of humorous similes of the form “as X as Y” to explain the cognitive processes that underlie irony. According to their research, they noted how people use these figurative comparisons as a mean to express ironic opinions with funny effects. Carvalho et al. [10] suggested some clues for automatically identifying ironic sentences. These clues are based on the fact of detecting emoticons, onomatopoeic expressions, punctuation and quotation marks. On the basis of this simple approach, they achieved interesting results in the task. Recently, Veale and Hao [41] presented a linguistic approach in order to automatically differentiate irony from non-irony in figurative comparisons. They noted how the presence of markers like “about” can produce a rule-based categorization between ironic and non-ironic texts.

In fine-grained approaches, Burfoot and Baldwin [9] have tried to determine whether or not newswire articles can automatically be classified as satirical. Using lexical and semantics features such as headlines, profanity —offensive language, or slang, authors could separate satirical from “true” (sic) newswire articles with interesting F-measure scores. From a different perspective, Tsur et al. [38] addressed their research in order to find elements to automatically detect sarcasm in online products reviews. On the basis of a semi-supervised approach, they suggested surface features such as content words —words regarding information about the product, company, title, etc., frequent words or punctuation marks, to represent sarcastic texts. According to their results, precision and recall scores are significantly positive.

Although these approaches have demonstrated that both humor and irony can be handled in terms of computational means, it is necessary to improve the mechanisms to represent their characteristics, and especially, to create a feature model capable to symbolize, the less theoretical as possible, both linguistic and social

knowledge in order to describe deeper and more general properties of these phenomena. For instance, most of results here described are text-specific, i.e. they are centered either on one-liners, punning riddles, similes, newswire articles, or products reviews; thus, their scope regarding different instances in which figurative language appears, are limited. Therefore, part of our objective is to identify salient components, both for humor and irony, by means of formal linguistic arguments, i.e. words and sequences of them, in order to gather a set of more general attributes to characterize these figurative devices.

#### 4. Feature Model

This section describes the set of features to firstly represent humor, and then, irony. Concerning humor, we focus on assessing intrinsic features on the basis of ambiguity. With respect to irony, we suggest more abstract features to represent favorable and unfavorable ironic contexts on the basis of profiled polarity, unexpectedness and emotional scenarios<sup>3</sup>.

##### 4.1. Preliminaries

Most of works related to humor stress out the role of ambiguity as a major mechanism to produce a funny effect. However, in this context of NLP, ambiguity is still work in progress. So far, the results are important for tasks in which the target is literal language, for instance, POS tagging or word sense disambiguation. However, the results are quite different regarding tasks in which the treatment of figurative language is involved (that is why ambiguity is regarded as future work in most of these tasks). Thus, the question is how to capitalize the advances in the treatment of ambiguity beyond literal language. In particular, taking into consideration that ambiguity in figurative language, and especially in humor, usually entails knowledge beyond the word or the sentence. Let us consider example 7 below to clarify this point.

7. “Jesus saves, and at today’s prices, that’s a miracle!”.

Unlike example 1 and 3, in which humor was given by phonological ambiguity (elly-phone vs. telephone), in this example humor is given by exploiting semantic and pragmatic ambiguity. The funny effect relies on different possible interpretations that are based on semantic meanings and cultural information. These facts,

---

<sup>3</sup>These features are not necessarily humor-dependent or irony-dependent. However, their relevance is analyzed when assessing the presence of these phenomena in different contexts.

according to cognitive grammar arguments —cf. Langacker [23], turn the *figure* of the sentence, i.e. “Jesus saves”, in the *ground*, thereby shifting the profiled sense of the whole sentence. These changes generate an ambiguous meaning and, consequently, a funny result. In other words, this example entails two interpretations: the first one is related to the figure, and logical sense about preserving someone from harm or loss. The second one shifts this meaning, from the logical sense related to a religious interpretation, to a ground sense related to economy. This sense is promoted as figure and then the meaning of the entire sentence becomes funny. This type of strategies, according to Mihalcea and Strapparava [26], leads surprise and create the humorous effect.

On the other hand, since irony cuts through different aspects of language — from pronunciation to lexical choice, syntactic structure, semantics and conceptualization, it is unrealistic to seek a general solution just in one single technique or algorithm. Moreover, the fuzzy boundaries among irony and concepts such as sarcasm or satire (cf. discussion in Section 2.2), make more difficult to establish patterns beyond punctuation marks or domain-specific words. Let us consider the following examples to illustrate the difficulty of this task:

8. “I feel so miserable without you, it’s almost like having you here.”
9. “Sometimes I need what only you can provide: your absence.”
10. “I thank God that you are unique!.”

According to the perception of the people, which is profiled by employing specific user-generated tags, these three examples could be either ironic, sarcastic or even satiric. This means there is not a clear distinction about their boundaries. Where does irony end, and where does sarcasm (or satire) begin? While irony courts ambiguity and often exhibits great subtlety, sarcasm is delivered with a cutting or withering tone that is rarely ambiguous. Regardless of these fine-grained differences, the final purpose of these examples is to communicate negative aspects. Furthermore, like in humor, these examples take advantage of unexpected situations to convey their meaning. This is clearer in examples 8 and 9, where the expected conclusion, given the initial chunks, would suggest a different and “sweeter” final. In addition, according to [34], irony evokes certain types of emotions. In these examples, we can cite aggressiveness, surprise, desire, and why not, zest and pleasure. Finally, we cannot obviate their funny effect.

Therefore, based on these preliminary arguments, the set of features we evaluate are:

- i. ambiguity, focusing on three layers: structural, morphosyntactic and semantic;

- ii. polarity, focusing on words which denote either positive or negative semantic orientations;
- iii. unexpectedness, focusing on contextual imbalances among the semantic meanings of the words;
- iv. emotional scenarios, focusing on psychological contexts regarding natural language concepts.

#### 4.2. *Structural Ambiguity*

As previously noted, humor uses some predictable features such as rhetoric devices in order to get humorous effects. On the basis of this property, we aim at investigating how much valuable information could be extracted by measuring, in terms of language models, the **perplexity** of a set of funny texts against that of non-funny ones. This measure, according to Jurafsky and Martin [21], predicts linguistic representation quality, given two probabilistic models. Therefore, since humor exploits ambiguity as a mechanism to generate its effect, our hypothesis is that humorous texts maximize the degree of perplexity by profiling a structural ambiguity. This structural ambiguity can be represented by the dispersion in the number of combinations among the words that constitute the humor examples. According to our hypothesis, structural ambiguity can be considered a trigger for potential funny situations and should thus appear quite often in humorous texts.

#### 4.3. *Morphosyntactic Ambiguity*

It is well-known that ambiguity covers all linguistic levels: from morphological up to discursive level. Hence, it was also considered important to study the impact that morphosyntactic ambiguity can have on processing figurative language, specifically humor. On this subject, we think that the number of POS tags that any word in context can have, it is a hint at the underlying mechanism of humor to produce its effect. For instance, the funny effect could rely on noting a meaning shift due to the use of, let us say, *lie* as a verb (to be postrated), rather than a noun (prevarication). Moreover, this ambiguity can be codified at syntactic level as well. Thus we aim at representing this property, in terms of a syntactic dependencies, in such a way to be able to determine the **complexity** of humorous and non-humorous texts.

#### 4.4. *Semantic Ambiguity*

Ambiguity is closely related to the different meanings that a word, phrase or sentence may produce. As we already mentioned, in example 7 the trigger which enables the funny interpretation is linked to semantic and pragmatic referents.

This represents a great challenge for NLP research. On this subject, both semantic and pragmatic levels constitute an important source of ambiguity triggers and, consequently, we aim at studying how valuable information can be obtained from these layers. We defined a measure to statistically estimate the range of **semantic dispersion** profiled by a text in order to determine how ambiguous this text is. The measure is based on the hypernym distance between synsets, calculated with respect to WordNet. Our hypothesis relies on the fact discussed in Section 4.1: humor is often the result of a shift in the ground meaning. If ground meaning is profiled, then logical meaning is broken and humor is produced.

#### 4.5. *Polarity*

One of the most common properties, both in humor and irony, relies on conveying the opposite meaning by profiling positive qualities over negative ones (cf. examples 5 and 10). This property, as we previously discussed, sometimes profiles an aggressive meaning, sometimes only yields a funny one. With this category, we intend to obtain an indicator about the correlation between words which semantically profile positive and negative meanings in a text; i.e. to determine the **polarity**. Our hypothesis is to find a greater weight regarding the presence of positive elements, regardless of the aggressive or funny meaning profiled in the texts. To this end, we employed the Macquarie Semantic Orientation Lexicon (MSOL) [33] to label our data sets. This resource contains 76,400 entries (30,458 positive and 45,942 negative ones).

#### 4.6. *Unexpectedness*

Irony often relies on situational phenomena related to incongruity, non expected situations, senseless, absurd, and so on. Lucariello [24] suggests the term unexpectedness to represent “imbalances in which opposition is a critical feature”. According to her arguments, a surprising factor is a key component, not only concerning irony but concerning humor as well. Unexpectedness therefore is intended to be a mechanism for representing contextual imbalances both in funny and ironic texts. This property, in accordance with her research, is an event related to oppositions or inconsistencies in contexts, situations, roles, or time. In order to measure **contextual imbalance**, we estimated the similarity of concepts taking into consideration their semantic relatedness. As noted by Oliva et al. [27], the fact of estimating semantic similarity, especially regarding short texts, is very important when facing natural language processing tasks. Therefore, our underlying hypothesis is: the lesser semantic relatedness, the greater contextual imbalance

(funny/ironic texts); the greater semantic relatedness, the lesser contextual imbalance (non funny/ironic texts).

#### 4.7. *Emotional Scenarios*

According to Shelley [34], an emotional ground is necessary to increase "the sense that [a] situation is ironic". This information, which is profiled by selecting certain linguistic elements, represents valuable knowledge regarding our task because many figurative expressions rely on these contents to produce their effect (cf. example 8)<sup>4</sup>. Emotional scenarios is thus a way of representing information regarding contents beyond grammar, and beyond positive or negative polarity. In others words, this feature attempts to especially characterize irony in terms of elements that symbolize abstract contents such as sentiments, attitudes, feelings, moods, and so on, in order to define a schema of favorable and unfavorable emotional contexts. On the basis of a psychological perspective, we intend to represent these contexts in terms of the categories described in [42]. These categories quantify emotional words in terms of scores obtained from human ratings regarding natural language. They are **activation** (degree of response, either passive or active, that humans have under an emotional state), **imagery** (how difficult it is to form a mental picture of a given word), and **pleasantness** (degree of pleasure produced by the words).

### 5. Evaluation

Several experiments were performed in order to evaluate the capabilities to automatically discriminate both humorous and ironic texts. The following schema summarizes the evaluation phase:

- i. feature representativeness. Phase focused on representing our evaluation corpus by means of the features previously discussed.
- ii. feature relevance. Phase focused on assessing the features by means of a classification task.

#### 5.1. *Evaluation Corpus*

Defining whether or not a text is funny (or ironic) is extremely subjective. In general, the property of being funny or ironic not only depends on the source (i.e.

---

<sup>4</sup>It is worth noting the importance of emotional content in NLP tasks. For instance, consider the recent advances in building knowledge bases for emotion detection described in [4].

the texts), but also on the target (i.e. the reader). If the latter is not capable of decoding the underlying meaning, it does not mean that such source is not funny or ironic. Personal factors such as mood, stress, or even linguistic competence, impact on the final interpretation. Therefore, in order to avoid the subjectivity of collecting a corpus based on personal judgments, we decided to collect examples a-priori considered either funny or ironic. Thus, we centered on user-generated tags. In particular, we used the ones posted on one of the current trendy enterprises regarding social media: Twitter. We collected a corpus of 50,000 texts. It is divided in five sets, each contains 10,000 texts. A mandatory requirement, except for one set, was determined in order to retrieve the remaining four: they should be labeled with a *hashtag*; i.e. a user-generated tag provided by the users themselves when posting their texts to focus their contributions on particular subjects. The hashtags were *#humor*, *#irony*, *#politics*, and *#technology*. The set *general* was retrieved without considering particular requirements<sup>5</sup>.

By considering this approach, apart from avoiding personal judgments, we obtained two adjacent benefits: i) it is unnecessary neither a manual annotation nor an inter-annotator agreement regarding the positive texts; and ii) according to our objective, we can extend the scopes of this research to others types of texts which contain figurative language.

## 5.2. Feature Representativeness

In order to estimate the perplexity in our data sets, we used the SRILM Toolkit [37]. Five particular language models were created (one for each data set), as well as a more representative one obtained by using the Google N-grams [8]. All the language models were trained with trigrams, employing interpolation and Kneser-Ney discount as smoothing methods. The perplexity for each data set was determined by comparing its language model against the Google language model. Subsequently, every text was represented with a perplexity ratio. This ratio was obtained by dividing the perplexity of the set it belongs to (for instance humor) by the size of the data set (i.e. 10,000), and finally multiplying this result for the length of each text.

With respect to the sentence complexity, we employed Formula 1, introduced in [5], to estimate how complex the syntactic structure is:

$$sentence\ complexity = \forall t_n \frac{\sum v_l + \sum n_l}{\sum cl} \quad (1)$$

---

<sup>5</sup>The whole corpus is available by contacting authors.

where  $v_l$  and  $n_l$  are the number of verbal and nominal links respectively; and  $cl$  is the number of clauses for text  $t_n$ . The process consisted in firstly labeling all texts with POS tags, and then, a syntactic parser was carried out. It is important to stress that, in order to reduce the ambiguity due to POS tags, we enable a POS disambiguation module before running the syntactic parser.

The last experiment concerning ambiguity was addressed to represent semantic dispersion among the words of a text. Thus, Formula 2 was employed:

$$\delta(w_s) = \frac{1}{P(|S|, 2)} \sum_{s_i, s_j \in S} d(s_i, s_j) \quad (2)$$

where  $S$  is the set of synsets  $(s_0, \dots, s_n)$  for word  $w$ ;  $P(n, k)$  is the number of permutations of  $n$  objects in  $k$  slots; and  $d(s_i, s_j)$  is the length of the hypernym path between synsets  $(s_i, s_j)$  according to WordNet. This formula is a way of quantifying the difference among the senses of a word. For instance, the noun *killer* has four synsets<sup>6</sup>. Taking into account only the synsets  $s_0$  and  $s_1$  we obtain as first common hypernym *physical entity*. The number of nodes to reach this hypernym is 6 and 2, respectively. Thus, its semantic dispersion is the sum of those distances divided by 2. Now, considering all its synsets, we obtain six possible combinations whose distances to their first common hypernym generates a dispersion of 6,83. This process was carried out for all texts in every data set, summing the semantic dispersion of all the words in a text, and dividing by its length.

The following experiment was focused on determining the polarity in texts. Firstly, all texts were stemmed and stopwords were removed. Then, on the basis of the MSOL database, Formula 3 was applied in order to represent, in terms of a sentiment analysis perspective, the underlying polarity:

$$Polarity(d_k) = \frac{\sum_{s_p} + \sum_{s_n}}{|d|} \quad (3)$$

where  $s_p$  is the set of positive words;  $s_n$  is the set of negative words; and  $|d|$  is the length of  $d_k$ .

Contextual imbalance was determined by measuring the semantic similarity among the words. As conducted on the previous experiment, texts were stemmed and stopwords removed. The semantic similarity was estimated by applying the Resnik measure, implemented in WordNet::Similarity [28], in order to obtain the

---

<sup>6</sup>cf. WordNet v. 3.0

semantic relatedness in all the texts. The context was delimited in a 3-word window. A backoff method to assign the most frequent sense was enabled as well.

Finally, regarding the emotional scenarios, texts were represented by the categories described in Section 4.7. In order to represent activation, imagery and pleasantness ratios, we used Whissell's Dictionary of Affect in Language [42]. This dictionary scores around 9,000 English words regarding the three scenarios. The score ranges goes from 1 (passive, difficult to form a mental picture, unpleasant) to 3 (active, easy to form a mental picture, pleasant). For instance, the term *flower* is passive (score = 1), easily representable (score = 3), and tends to be pleasant (score = 2.75).

### 5.3. Feature Relevance

This phase consisted of assessing the capabilities of automatically classifying texts into the data set they belong to. Each one of the 50,000 documents was converted in a frequency-weighted term vector<sup>7</sup> according to the ratios obtained in the representativeness phase. Then, a decision tree was used to classify the texts. Four classifiers were trained considering a binary scenario; i.e. a positive set was always compared against a negative one. All of them were performed considering 70% for training, and 30% for test. The following schema illustrates the features evaluated for each classifier respectively:

- i. features regarding ambiguity (perplexity, sentence complexity, and semantic dispersion), considering the set humor *vs.* the sets irony, politics, technology, and general;
- ii. features regarding polarity (positive and negative), unexpectedness (contextual imbalance), and emotional scenarios (activation, imagery, pleasantness), considering the set irony *vs.* the sets humor, politics, technology, and general;
- iii. features regarding ambiguity, polarity, unexpectedness, and emotional scenarios, considering the set humor *vs.* the sets irony, politics, technology, and general;
- iv. features regarding polarity, unexpectedness, emotional scenarios, and ambiguity, considering the set irony *vs.* the sets humor, politics, technology, and general.

---

<sup>7</sup>All documents were preprocessed removing hashtags and stopwords. Stemming was applied as well.

Results in terms of accuracy, precision, recall, and F-measure, are detailed in Tables 1- 4. Table 1 shows the results about ambiguity concerning humor (first classifier); Table 2 contains the results about polarity, unexpectedness, and emotional scenarios concerning irony (second classifier); Table 3, includes the results about ambiguity, polarity, unexpectedness, and emotional scenarios concerning humor (third classifier); finally, Table 4 illustrates the results about polarity, unexpectedness, emotional scenarios, and ambiguity, concerning irony (fourth classifier).

Implications are discussed in the following section.

Table 1: Results classifier i. (Features: ambiguity)

	Accuracy	Precision	Recall	F-measure
Humor <i>vs.</i> Irony	85.15	0.96	0.73	0.83
Humor <i>vs.</i> Politics	77.35	0.75	0.82	0.78
Humor <i>vs.</i> Technology	71.27	0.66	0.88	0.75
Humor <i>vs.</i> General	77.27	0.93	0.59	0.72

Table 2: Results classifier ii. (Features: polarity, unexpectedness, emotional scenarios)

	Accuracy	Precision	Recall	F-measure
Irony <i>vs.</i> Humor	62.30	0.62	0.61	0.62
Irony <i>vs.</i> Politics	67.73	0.68	0.67	0.68
Irony <i>vs.</i> Technology	59.58	0.59	0.65	0.61
Irony <i>vs.</i> General	55.78	0.56	0.57	0.56

Table 3: Results classifier iii. (Features: ambiguity, polarity, unexpectedness, emotional scenarios)

	Accuracy	Precision	Recall	F-measure
Humor <i>vs.</i> Irony	93.13	0.93	0.93	0.93
Humor <i>vs.</i> Politics	85.93	0.87	0.85	0.86
Humor <i>vs.</i> Technology	85.82	0.85	0.86	0.86
Humor <i>vs.</i> General	92.15	0.92	0.93	0.92

Table 4: Results classifier iv. (Features: polarity, unexpectedness, emotional scenarios, ambiguity)

	Accuracy	Precision	Recall	F-measure
Irony <i>vs.</i> Humor	84.33	0.80	0.91	0.85
Irony <i>vs.</i> Politics	91.97	0.90	0.95	0.92
Irony <i>vs.</i> Technology	88.97	0.87	0.91	0.89
Irony <i>vs.</i> General	70.12	0.78	0.56	0.65

## 6. Discussion

According to the results depicted in these tables, we can note important implications regarding the usefulness of these features to represent recurrent properties in figurative language. Taking into consideration the accuracy achieved in most of experiments, it is clear the effectiveness of these features when discriminating texts belonging to five distinct sets. Moreover, looking at precision, recall, and F-measure rates, we can corroborate their relevance. This means that, at least regarding the data sets here employed, the capabilities of representing two types of expression concerning figurative language (humor and irony) are satisfactory.

According to the results, it is clear that humor is more suitable than irony. For instance, considering only ambiguity (Table 1), humor achieves ratios which always exceed 70% of accuracy, whereas irony hardly achieves ratios higher than 60% (Table 2). In contrast, when considering the whole set of features, humor reaches up to 93% of accuracy (Table 3), whereas irony markedly improves its score, reaching up to 90% in its best result (Table 4). Despite these differences, it is worth noting to highlight the accuracy achieved when considering these devices: regardless the features evaluated by each classifier, when discriminating humor *vs.* irony, and vice versa, the results are (usually) better than when classifying the remaining sets.

Furthermore, in order to evaluate the features beyond a binary representation, a last classifier was built. In this one, all data sets were represented with the whole set of features, then, the five sets were classified in a multi-class classification. The results support the previous ones: 80% of accuracy, and F-measure = 0.79. All these results point to make evident the presence of underlying patterns in both figurative devices that are well-represented by these features.

With respect to this assumption, we decided to verify which features are the most relevant to represent either humor or irony. Thus, we applied to each classifier an information gain filter. According to the results obtained, we could ap-

precipitate how the relevance of every feature is related to the kind of information profiled by each data set. For instance, when classifying texts from humor and politics sets, the most informative features were perplexity, pleasantness, sentence complexity, and semantic dispersion; whereas when classifying texts from irony and politics sets, the most relevant ones were pleasantness, activation, perplexity, and contextual imbalance.

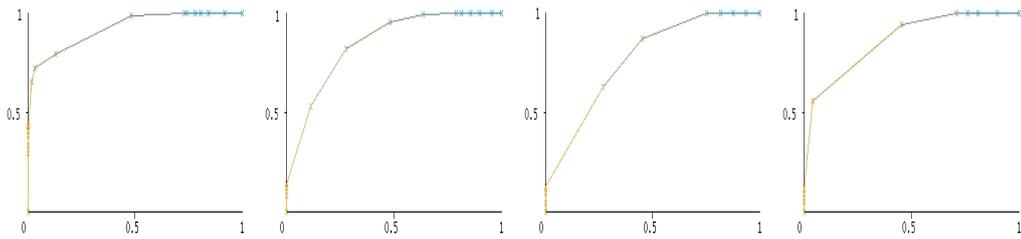


Figure 1: Learning curves regarding classifier i. humor *vs.* irony; politics, technology, general.

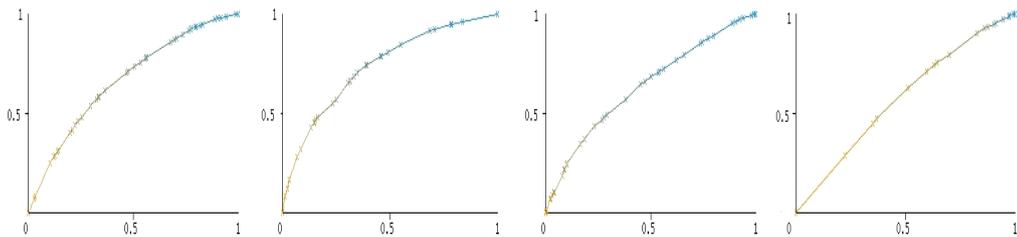


Figure 2: Learning curves regarding classifier ii. irony *vs.* humor; politics, technology, general.

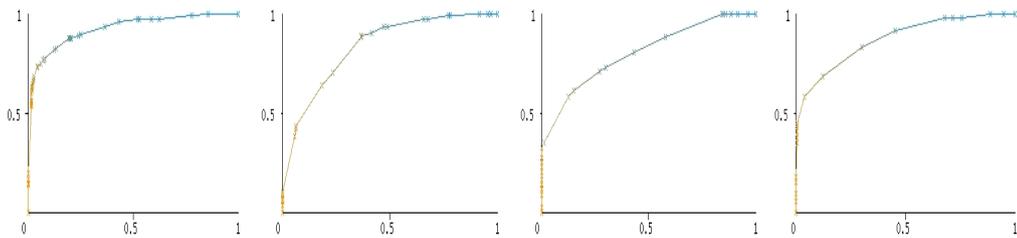


Figure 3: Learning curves regarding classifier iii. humor *vs.* irony; politics, technology, general.

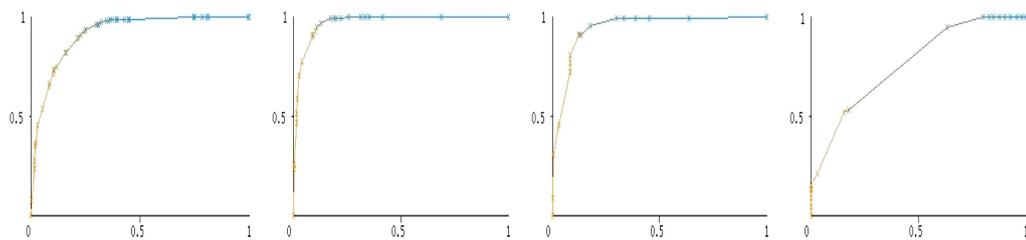


Figure 4: Learning curves regarding classifier iv. iron *vs.* humor; politics, technology, general.

This is clearer when analyzing the learning curves achieved in each classification. In Figures 1 - 4, we graphically show the performance of these features when classifying each one of the positive sets *vs.* the negative ones. Figure 1 shows the learning curves regarding the evaluations of the first classifier; Figure 2 depicts the ones regarding the second classifier; whereas Figure 3 and Figure 4, present the ones regarding the third and fourth classifier, respectively.

According to these figures, it is obvious that the learning curve, both for humor and irony data sets, is achieved with less instances when the whole features are considered; i.e. there is a noticeable improvement. However, the feature performance is not constant for all the cases. For instance, the convergence is easily reached when discriminating either humor or irony set from politic set. Just the contrary with respect to technology or general sets. This might suppose that every data set profiles specific linguistic information in order to efficiently convey its message. Thus, the feature effectiveness will be related to the types of negative data; i.e. they can represent a better solution for some data sets but a worse one for others. Despite these issues, we consider the feature performance to be satisfactory.

Finally, we would like to stress some remarks regarding every feature.

- i. The results obtained by estimating the perplexity demonstrated, according to our initial assumption, how the underlying structure in figurative language is less predictable and, probabilistically, more ambiguous than literal language. This means that, given two different distributional schemes, the structures that have a broader range of combinations are the ones concerning humorous and ironic discourses.
- ii. Morphosyntactic ambiguity seems to be another important feature to represent figurative language. By means of measuring syntactic complexity, we could note that both funny and ironic texts are less complex than texts in the

remaining sets. This behavior suggests well-formed structures in figurative devices, which exploit other types of strategies. Mainly, based on semantic and pragmatic layers.

- iii. The role of semantic layer as a trigger of ambiguous situations seems to be more relevant. On the basis of semantic dispersion results, it is possible to infer the relevance of semantic strategies in figurative language. By profiling, at least two possible interpretations, it is more likely to generate hollows of ambiguity which contribute to produce more complex meanings both in funny and ironic texts.
- iv. With respect to polarity, despite the greater number of negative words in the MSOL (more than 15,000 words of difference; cf. Section 4.5), it is worth noting how positive words are more representative concerning the funny texts. In contrast, the ironic texts concentrate most of negative words. This fact is contrary to results described in [25](which is focused only on one-liners). They suggest the relevance of negative information to generate humor. In addition, concerning irony, these results makes question our assumption about the use of positive information to express an underlying negative meaning.
- v. Regarding unexpectedness, our underlying assumption relies on the fact that a text whose constituents profile senses that significantly differ among them is more likely to be used in figurative language, than a text whose words project senses that slightly differ. Based on contextual imbalance results, we could appreciate the relevance of this feature when classifying the ironic texts. Opposite situation regarding humor: contextual imbalance was, usually, useless to represent funny texts. This behavior suggests, at least regarding these type of texts and this specific feature, different strategies to achieve either funny or ironic effects.
- vi. The role played by the last feature (emotional scenarios) on the classifications is significant. Considering the three categories (activation, imagery, pleasantness), it is remarkable the effectiveness of this feature for increasing the classification accuracy, both for funny and ironic texts. We can interpret the results achieved with this feature as a way of communicating ad hoc stimuli, through which, people easily produce favorable contexts to express figurative language.

## 7. Conclusions and Further Work

In this paper we have presented an approach to the representation of two important figurative devices in short online texts: humor and irony. The features we have considered represent different types of patterns from a text: ambiguity, polarity, unexpectedness, and emotional content. They intended to symbolize low and high level properties of figurative language on the basis of formal linguistic elements. An evaluation corpus of 50,000 texts automatically retrieved from Twitter was used to evaluate the patterns. Two goals were considered in the evaluation: representativeness and relevance. Some of the results, apart from being satisfactory in terms of classification accuracy, precision, recall, and F-measure, confirmed our initial assumptions about the usefulness of this kind of information to characterize these devices. According to the results, it is important to highlight that the set of features work together as a system; i.e. no single feature is distinctly humorous or ironic, but all of them together provide a useful linguistic inventory for detecting these types of figurative devices at textual level. Further work consists of improving the quality of the features, as well as in identifying new ones, especially regarding irony. In addition, we aim at assessing the scope of the features by verifying their performance with other types of data sets, and considering other types of figurative devices.

## Acknowledgments

The TEXT-ENTERPRISE 2.0 (TIN2009-13391-C04-03) research project has partially funded this work. The National Council for Science and Technology (CONA-CyT - Mexico) has funded the research work of Antonio Reyes.

## References

- [1] Attardo, S., 1994. *Linguistic Theories of Humor*. Mouton de Gruyter.
- [2] Attardo, S., 2001. *Humorous Texts: A semantic and pragmatic analysis*. Mouton de Gruyter.
- [3] Attardo, S., 2007. Irony as relevant inappropriateness. In: Gibbs, R., Colston, H. (Eds.), *Irony in Language and Thought*. Taylor and Francis Group, pp. 135–174.

- [4] Balahur, A., Hermida, J., Montoyo, A., Muñoz, R., 2011. Emotinet: A knowledge base for emotion detection in text built on the appraisal theories. In: Proceedings of the the 16th International Conference on Natural Language Processing and Information Systems (NLDB). pp. 27–39.
- [5] Basili, R., Zanzotto, F., 2002. Parsing engineering and empirical robustness. *Journal of Natural Language Engineering* 8 (3), 97–120.
- [6] Binsted, K., 1996. Machine humour: An implemented model of puns. Ph.D. thesis, University of Edinburgh, Edinburgh, Scotland.
- [7] Binsted, K., Ritchie, G., 1997. Computational rules for punning riddles. *Humour* 10, 25–75.
- [8] Brants, T., Franz, A., 2006. Web 1t 5-gram corpus version 1.
- [9] Burfoot, C., Baldwin, T., 2009. Automatic satire detection: Are you having a laugh? In: ACL-IJCNLP '09: Proceedings of the ACL-IJCNLP 2009 Conference Short Papers. pp. 161–164.
- [10] Carvalho, P., Sarmiento, L., Silva, M., de Oliveira, E., November 2009. 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. ACM, Hong Kong, China, pp. 53–56.
- [11] Colston, H., 2007. On necessary conditions for verbal irony comprehension. In: Gibbs, R., Colston, H. (Eds.), *Irony in Language and Thought*. Taylor and Francis Group, pp. 97–134.
- [12] Colston, H., Gibbs, R., 2007. A brief history of irony. In: Gibbs, R., Colston, H. (Eds.), *Irony in Language and Thought*. Taylor and Francis Group, pp. 3–24.
- [13] Curcó, C., 2007. Irony: Negation, echo, and metarepresentation. In: Gibbs, R., Colston, H. (Eds.), *Irony in Language and Thought*. Taylor and Francis Group, pp. 269–296.
- [14] Dews, S., Kaplan, J., Winner, E., 2007. Why not say it directly? the social functions of irony. In: Gibbs, R., Colston, H. (Eds.), *Irony in Language and Thought*. Taylor and Francis Group, pp. 297–317.

- [15] Gibbs, R., 2007. Irony in talk among friends. In: Gibbs, R., Colston, H. (Eds.), *Irony in Language and Thought*. Taylor and Francis Group, pp. 339–360.
- [16] Gibbs, R., Colston, H., 2007. The future of irony studies. In: Gibbs, R., Colston, H. (Eds.), *Irony in Language and Thought*. Taylor and Francis Group, pp. 339–360.
- [17] Giora, R., 1995. On irony and negation. *Discourse Processes* 19 (2), 239–264.
- [18] Grice, H., 1975. Logic and conversation. In: Cole, P., Morgan, J. L. (Eds.), *Syntax and semantics*. Vol. 3. New York: Academic Press, pp. 41–58.
- [19] Halliwell, S., 2008. *Greek Laughter. A Study of Cultural Psychology from Homer to Early Christianity*. Cambridge University Press, New York.
- [20] Hertzler, J., 1970. *Laughter: A social scientific analysis*. Exposition Press.
- [21] Jurafsky, D., Martin, J., 2007. *Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Prentice Hall.
- [22] Kumon-Nakamura, S., Glucksberg, S., Brown, M., 2007. How about another piece of pie: The allusional pretense theory of discourse irony. In: Gibbs, R., Colston, H. (Eds.), *Irony in Language and Thought*. Taylor and Francis Group, pp. 57–96.
- [23] Langacker, R., 1991. *Concept, Image and Symbol. The Cognitive Basis of Grammar*. Mouton de Gruyter.
- [24] Lucariello, J., 2007. Situational irony: A concept of events gone away. In: Gibbs, R., Colston, H. (Eds.), *Irony in Language and Thought*. Taylor and Francis Group, pp. 467–498.
- [25] Mihalcea, R., Pulman, S., 2007. Characterizing humour: An exploration of features in humorous texts. In: *8th International Conference on Computational Linguistics and Intelligent Text Processing, CICLing 2007*. Vol. 4394 of LNCS. pp. 337–347.

- [26] Mihalcea, R., Strapparava, C., 2006. Learning to Laugh (Automatically): Computational Models for Humor Recognition. *Journal of Computational Intelligence* 22 (2), 126–142.
- [27] Oliva, J., Serrano, J., del Castillo, M., Iglesias, A., 2011. SyMSS: A syntax-based measure for short-text semantic similarity. *Data and Knowledge Engineering* 70 (4), 390–405.
- [28] Pedersen, T., Patwardhan, S., Michelizzi, J., 2004. Wordnet::similarity - measuring the relatedness of concepts. In: *Proceeding of the 9th National Conference on Artificial Intelligence (AAAI-04)*. Association for Computational Linguistics, Morristown, NJ, USA, pp. 1024–1025.
- [29] Reyes, A., Buscaldi, D., Rosso, P., 2010. The impact of semantic and morphosyntactic ambiguity on automatic humour recognition. In: *Proceedings of the 14th International Conference on Applications of Natural Language to Information Systems NLDB 2009*. Vol. 5723 of LNCS. pp. 130–141.
- [30] Reyes, A., Potthast, M., Rosso, P., Stein, B., 2010. Evaluating humour features on web comments. In: *Proceedings of the 7th International Conference on Language Resources and Evaluation*. pp. 1138–1141.
- [31] Reyes, A., Rosso, P., Buscaldi, D., 2009. Humor in the blogosphere: First clues for a verbal humor taxonomy. *Journal of Intelligent Systems* 18 (4), 311–331.
- [32] Ruch, W., 2001. The perception of humor. In: Scientific, W. (Ed.), *Emotions, Qualia, and Consciousness*. *Proceedings of the International School of Biocybernetics*. pp. 410–425.
- [33] Saif, M., Cody, D., Bonnie, D., 2009. Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus. In: *Proceedings of the 2009 Conference on EMNLP*. Association for Computational Linguistics, Morristown, NJ, USA, pp. 599–608.
- [34] Shelley, C., 2007. The bicoherence theory of situational irony. In: Gibbs, R., Colston, H. (Eds.), *Irony in Language and Thought*. Taylor and Francis Group, pp. 531–578.

- [35] Sjöbergh, J., Araki, K., 2007. Recognizing humor without recognizing meaning. In: 3rd Workshop on Cross Language Information Processing, CLIP-2007, Int. Conf. WILF-2007. Vol. 4578 of LNAI. pp. 469–476.
- [36] Stock, O., Strapparava, C., 2005. Hahacronym: A computational humor system. In: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. pp. 113–116.
- [37] Stolcke, A., 2002. SRILM - An Extensible Language Modeling Toolkit. In: Proceedings of 7th International Conference on Spoken Language Processing, INTERSPEECH 2002. pp. 901–904.
- [38] Tsur, O., Davidov, D., Rappoport, A., 23-26 May 2010. ICWSM – A great catchy name: Semi-supervised recognition of sarcastic sentences in online product reviews. In: Cohen, W. W., Gosling, S. (Eds.), Proceedings of the Fourth International AAI Conference on Weblogs and Social Media. The AAI Press, Washington, D.C., pp. 162–169.
- [39] Utsumi, A., 1996. A unified theory of irony and its computational formalization. In: Proceedings of the 16th conference on Computational Linguistics. Association for Computational Linguistics, Morristown, NJ, USA, pp. 962–967.
- [40] Veale, T., Hao, Y., 2009. 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. pp. 1376–1381.
- [41] Veale, T., Hao, Y., 2010. Detecting ironic intent in creative comparisons. In: Proceedings of 19th European Conference on Artificial Intelligence - ECAI 2010. IOS Press, Amsterdam, The Netherlands, The Netherlands, pp. 765–770.
- [42] Whissell, C., 2009. Using the revised dictionary of affect in language to quantify the emotional undertones of samples of natural language. *Psychological Reports* 105 (2), 509–521.
- [43] Wilson, D., Sperber, D., 2007. On verbal irony. In: Gibbs, R., Colston, H. (Eds.), *Irony in Language and Thought*. Taylor and Francis Group, pp. 35–56.

## Vitae

Antonio Reyes is a Ph.D. Student at Universidad Politécnica de Valencia, Spain. He is currently a member of the Natural Language Engineering and Pattern Recognition research group at Universidad Politécnica de Valencia, Spain, as well as a founder of the Language Technologies Lab at Superior Institute of Interpreters and Translators, Mexico. His major interests are focused on figurative language processing; especially, on topics related to irony, sarcasm, and humor. He has published several papers in different conferences, workshops and journals.



Paolo Rosso received his Ph.D. degree on Computer Science (1999) from the Trinity College Dublin, University of Ireland. He is currently an Associate Professor at Universidad Politécnica de Valencia, Spain, where he leads the Natural Language Engineering Laboratory of the Natural Language Engineering and Pattern Recognition research group. He has published over 200 papers in different conferences, workshops and journals being involved in many national and international research projects. His main research interests are mainly focused on irony detection, humor recognition, plagiarism detection and geographical information retrieval.



Davide Buscaldi got his Ph.D. “cum laude” in Pattern Recognition and Artificial Intelligence at the Universidad Politécnica de Valencia in 2010 under the

guidance of Dr. Paolo Rosso with a thesis on “Toponym Disambiguation in Information Retrieval”. He is the author of more than 60 papers published in international Journals and conferences. He is currently carrying out a post-doctoral stage at the IRIT of Toulouse (France) on semantic IR. His main research interests are geographic IR, word sense disambiguation, ontology learning and Question Answering.

