

# A Fine-grained Annotation of Irony in Italian Social Media Texts



Alessandra Teresa Cignarella  
748564

Supervisor: Cristina Bosco, Assistant Professor

University of Turin  
Department of Humanistic Studies  
*Master in Linguistics*

April 2017

*Dave Bowman: Hello, HAL. Do you read me, HAL?*  
*HAL: Affirmative, Dave. I read you.*  
*Dave Bowman: Open the pod bay doors, HAL.*  
*HAL: I'm sorry, Dave. I'm afraid I can't do that.*  
*Dave Bowman: What's the problem?*  
*HAL: I think you know what the problem is just as well as I do.*  
*Dave Bowman: What are you talking about, HAL?*  
*HAL: This mission is too important for me to allow you to jeopardize it.*  
*Dave Bowman: I don't know what you're talking about, HAL.*  
*HAL: I know that you and Frank were planning to disconnect me, and I'm afraid that's something I cannot allow to happen.*  
*Dave Bowman: Where the hell did you get that idea, HAL?*  
*HAL: Dave, although you took very thorough precautions in the pod against my hearing you, I could see your lips move.*  
*Dave Bowman: Alright, HAL. I'll go in through the emergency airlock.*  
*HAL: Without your space helmet, Dave? You're going to find that rather difficult.*  
*Dave Bowman: HAL, I won't argue with you anymore! Open the doors!*  
*HAL: Dave, this conversation can serve no purpose anymore. Goodbye.*

2001: A Space Odyssey

# Abstract

This master thesis concerns the development of a fine-grained annotation scheme for pragmatic phenomena to be exploited within the context of automatic sentiment analysis, but also to shed some light on these linguistic phenomena. In particular, it will be used to annotate the rhetorical device of irony in texts from social media. Our prime area of investigation is, in fact, the microblogging platform *Twitter*.

From previous research the recognition of irony and the identification of pragmatic and linguistic devices that activate it are known as very challenging tasks to be performed by both humans or automatic tools. Our goal, is to create an annotated Italian corpus, through which we will hopefully be able to resolve some issues concerning irony formalization and automatic detection.

This thesis collocates in the context of a multilingual project for studying irony and for developing resources to be exploited in training NLP tools for sentiment analysis.

*Keywords:* sentiment analysis, irony, humor, Twitter, fine-grained, annotation

# Contents

<b>1</b>	<b>Introduction</b>	<b>5</b>
1.1	Background information . . . . .	6
1.2	Research question . . . . .	7
1.3	Structure . . . . .	8
<b>2</b>	<b>Sentiment Analysis and Figurative Language</b>	<b>10</b>
2.1	Features exploited in sentiment analysis . . . . .	13
2.2	Lexical resources . . . . .	14
2.3	Datasets . . . . .	17
2.4	Annotation schemes . . . . .	18
2.5	Identifying Irony . . . . .	20
2.6	Related work . . . . .	23
2.7	A broader panorama . . . . .	24
<b>3</b>	<b>Linguistic theories</b>	<b>26</b>
3.1	Early studies on pragmatics . . . . .	26
3.1.1	Austin and Searle . . . . .	26
3.1.2	Grice: towards a new approach . . . . .	27
3.2	Later theories on irony . . . . .	33
3.2.1	Beyond gricean maxims: Sperber and Wilson . . . . .	33
3.2.2	Another point of view: Clark and Gerrig . . . . .	39
<b>4</b>	<b>Twitter</b>	<b>44</b>
4.1	A microblogging platform . . . . .	44
4.2	Twitter as corpus . . . . .	46

<b>5</b>	<b>Methods for data annotation</b>	<b>49</b>
5.1	Development of the corpus . . . . .	49
5.1.1	Data collection . . . . .	50
5.1.2	A new corpus: TWITTIRÒ . . . . .	51
5.2	Following a common scheme . . . . .	52
5.3	Presentation of the scheme . . . . .	53
5.3.1	Level 1: Class . . . . .	55
5.3.2	Level 2: Contradiction types . . . . .	56
5.3.3	Level 3: Categories . . . . .	57
5.3.4	Level 4: Clues . . . . .	62
<b>6</b>	<b>Data analysis</b>	<b>64</b>
6.1	Features and patterns . . . . .	64
6.1.1	Syntactic devices . . . . .	66
6.1.2	Graphic devices . . . . .	69
6.1.3	Semantic devices . . . . .	72
6.1.4	Pragmatic devices . . . . .	76
6.1.5	Mixed devices . . . . .	78
6.2	Discussion: difficulties in applying the common frame . . . . .	79
<b>7</b>	<b>Conclusions</b>	<b>89</b>
7.1	Future work . . . . .	91
	<b>Appendix</b>	<b>94</b>
	<b>Bibliography</b>	<b>110</b>

# List of Figures

2.1	A synset from WordNet . . . . .	16
4.1	Example of Twitter’s interface . . . . .	45
5.1	Annotation scheme for irony . . . . .	55
5.2	Units and their attributes in the annotation scheme . . . . .	63
6.1	Inter-annotator agreement on type, category and polarity in the three sub-corpora. . . . .	65
6.2	Occurrences of negation words in the TWITTIRÒ corpus . . . . .	73
6.3	Positive and negative tweets in the corpus . . . . .	82
6.4	Three different levels of inter-annotator agreement . . . . .	84
6.5	Category-tags exploited by two annotators . . . . .	84
6.6	Distribution of category tags in opposition to OTHER . . . . .	85

# List of Tables

4.1	Examples from TW-BS corpus . . . . .	47
5.1	Distribution of tweets in each corpus [23] . . . . .	50
5.2	Distribution of tweets in TWITTIRÒ corpus . . . . .	51

# Chapter 1

## Introduction

*Are machines going to be able to process, understand and generate speech like humans do?* This one is a question that has been echoing since the beginning of the digital era and the early developing of computers. A vast selection of sci-fi books and movies also engages the controversial theme of artificial intelligence and human-like robots, describing their capabilities and impressive skills in many and various picturesque ways. Sometimes androids move like humans, think like humans and in particular speak like humans.

Of course many of this stories are to be considered as a pure form of art and fiction, but if we pause for a moment and reflect upon nowadays world and technology, we may understand that not all of what is represented is false: in the past ten years combined efforts of different disciplines such as computer science, artificial intelligence, cognitive psychology and computational linguistics have been developing softwares and applications that resemble under some respect what is described in those stories. I am referring to devices such as vocal assistants, text-to-speech applications or chat-bots. Present technology is so advanced that we are actually capable of talking to our mobile phone and schedule an appointment in our agenda without touching the screen, call our friends by simply spelling their name out loud or even decide whether to take the umbrella or not on our way to work because we already know if it is going to be raining or not. That is certainly impressive, but when we talk to a tech-device we have to be extremely precise in our commands, it is, after all, just a machine that simply follows our orders: it doesn't have a conscience



neither a free will. Sometimes it can happen that the machine simply does not function because we might have spoken too fast or too “humanly” without paying attention to output a clear and loud pronunciation that may undermine performance of the system. Adding to these routinely common difficulties, one other recurrent problem in science-fiction, and as well in reality, is that even though machines can successfully complete highly developed tasks they still seem to lack something important.

What I have always been interested in was to find an answer to describe precisely what that flaw is and what we might do to overcome its perception. Among many others, one reply that I always found for the most part appropriate is the fact that what best defines a human being is the capacity of generating and understanding natural speech. Furthermore, looking from a more disciplined and far less naive point of view, I would also add that one other big interrogation on natural speech is “how does it come that humans also understand meaning that lies beyond what it is literally said and a machine can not?”.

Taking into account linguistic theories, pragmatics, previous research, and state-of-the-art technologies, I will try to describe this flaw – in a very humble and introductory perspective – from one precise angle, that is: finding possibly interesting patterns of irony in texts and aim to modeling irony in some way. Hopefully this work will be a good starting point to reply to that big and almost philosophical question.

## 1.1 Background information

The present master thesis is developed in COMPUTATIONAL LINGUISTICS, a relatively recent subject that sees the connection of Linguistics and Computer Science. In particular, the dissertation broadens within the field of SENTIMENT ANALYSIS, that deals with the computational treatment of opinion, sentiment, and subjectivity in texts.

Firstly, using statistics, features derived from linguistic theories, sentiment

analysis tries to extract and characterize the SENTIMENT content of a text unit.

Secondly, the topics of texts of interest for sentiment analysis can be several, and we must take into account also the coverage of domains when we design a resource. The same problem of coverage can be related to languages, since several resources are available for a few languages when they are almost missing at all for several others.

Namely, our principal target is that of creating an ANNOTATED CORPUS FOR ITALIAN language in social media texts. Previous work that inspired this master thesis is to be found in the research activities conducted by the *CCC group* in the *Computer Science Department of University of Turin* in relation to sentiment analysis.

Many other researches contributed inspiring the work I have done in the thesis and they will be quoted in the following chapters and step by step will be added to the bibliography. Some of them are related to linguistic theories, others are connected to sentiment analysis and more specific aspects of this technology, that is IRONY DETECTION.

## 1.2 Research question

Taking into consideration the fact that irony detection is a field that has been growing very fast in the past years [16, 27] and that such technology needs a great amount of data to be improved; and also taking into account the fact that generation of irony (whether it is spoken or written) is rigorously dependent on the language and on the culture in which is expressed, the aim of my thesis is to reply to this simple question: “Is it possible to formally model irony? If so, how?”. Finally, my work concerns the use and adaptation of a fine-grained scheme for irony annotation and the subsequent creation of a corpus for Italian.

Through the present thesis, we aim at describing the phenomenon of irony not only in Italian, but rather in a multilingual perspective. That is: we hope that studying the use of figurative language in Italian social media texts, will

help us to better understand the developing of humor itself, but that also will lead us to the discovery of features and patterns that can be shared and confronted in a following moment with similar projects in other languages.

In the following chapters are going to be described the main characters that play a role in the development of this thesis: the fields of sentiment analysis and opinion mining; previous research which is also our main source of data; the social network platform called Twitter; linguistic theories on humor (focusing on irony) and the influence they have on computational linguistics tasks such ours.

## 1.3 Structure

The present thesis is organized as follows.

In chapter 2 (Sentiment Analysis and Figurative Language), I will provide an introductory overview of sentiment analysis and opinion mining, a sub-discipline in the field of NLP that has been gaining more and more attention in the past decade. Here, I will also introduce the reader to the main areas in which sentiment analysis extends, I will mention some of the most important approaches and resources that are used in this field. I will, lastly focus on the main topic concerning this thesis: irony detection, and I will further broaden the panorama of similar studies that inspired the present work.

In chapter 3 (Linguistic Theories), I will quote and describe many theories of pragmatics, that have been and still are a great source of inspiration for computational studies on natural languages. In particular, I will focus on Austin and Searle, reporting the *Speech Acts Theory* through which we will arrive to Grice and his theories on conversation. The present thesis in computational linguistics is highly in debt with the theories of Clark, Gerrig, Sperber and Wilson. The last two, in particular, provide the sufficient linguistic background that we exploit to study figurative language from a computational perspective.

In chapter 4 (Twitter), a description of the microblogging platform is given. Our main concern is to make the reader conscientious that the present work

is an inquiry of irony in the context of Italian social media texts. Chiefly, the data studied here are Twitter messages, therefore making the social networking site our prime source for the creation of a corpus.

With chapter 5 (Methods for Data Annotation) we enter the empirical part of the present master thesis. In this chapter, I present the process that leads to the creation of a dedicated corpus for studying irony in social media text: TWITTIRÒ. I will thoroughly describe the corpus, its composition and its characteristics in relation to pre-existing corpora exploited for other studies. After that, I will introduce through many explicative examples, a detailed overview of the multi-layered annotation scheme for irony that we used to complete our task, which is also available in the Appendix in its full length.

In chapter 6 (Data Analysis), for the most part, I discuss the tweets from the dataset and point out the existence of some recurrent features and patterns. The idea, in this chapter is to look for connections between linguistic theories and their exploitation in sentiment analysis. Further on, I will discuss the perplexities that have come up during the annotation work and during the analysis of data. Lastly, I will propose some enlightening on new techniques and procedures that might be followed in recent research and further work.

In the Appendix we provide the original text of the exploited *Multi-layered Annotation Scheme for Irony* developed in [22] and largely exploited in [23].

## Chapter 2

# Sentiment Analysis and Figurative Language

In the vast area of research that is Natural Language Processing (NLP) a sub-discipline has been gaining much interest in the past decades due to the quick development of the World Wide Web: **sentiment analysis** (sometimes also mentioned as *opinion mining*), which deals with the computational treatment of opinion, sentiment, and subjectivity in texts.

One important thing that affects our lives and has always made the world spin, is the answer to the question “What do other people think?”. Internet and the Web have nowadays made it possible to find out about the opinions and experiences of a lot of people that are not our friends and that we have even never met. As I previously mentioned, the area of sentiment analysis has recently experienced a peak of interest in research activity, due to many factors such as:

- the rise of machine learning methods in natural language processing and information retrieval;
- the availability of datasets for machine learning algorithms to be trained on, due to the blossoming of the World Wide Web and, specifically, the development of review-aggregation web-sites;
- the realization of the fascinating intellectual challenges and commercial and intelligence applications that the area offers. [31].

The area of interest in the early projects of sentiment analysis was focused

mostly on interpretation of metaphors, narratives, expressions of a points of view, emotions, and related areas. Later on, also bridging with other kindred subjects such as computer science, artificial intelligence, psychology, anthropology and sociology, sentiment analysis began a new phase that Pang and Lee define “**social media monitoring and analysis**” [31]. They describe this definition as a placeholder, to be used until something better (and shorter) takes hold in English.

The main research covered in the field of sentiment analysis is the computational treatment of *opinion*, *sentiment*, and *subjectivity* expressed through *polarity* in text [25]. Therefore, it is of great relevance also the classification of text according to the opinion and sentiment there expressed.

Furthermore, the subject in question is strictly connected to the world of commerce and economics where the trendiest interrogations are, for example: “*Is this product review positive or negative?*”, “*Is the customer writing this email satisfied or dissatisfied?*”, “*Based on a sample of texts, how are people responding to this product release?*” or even “*How did bloggers’ attitudes of presidential candidates change since the election?*”.

Certain matters do not only provide everyday challenges for researchers but also their results may have a huge impact on nowadays society.

In this regard, it is correct to say that sentiment analysis researches the deeper level of human language: the one that concerns the expression of emotions. By any means, being able to describe, or formalize human emotions is complicated. Sentiment analysis, in particular, using statistics, or machine learning methods, tries to extract or identify, and characterize the *sentiment* content of a text unit [31].

Studies explain that humans follow regular patterns and schemes when they utter a sentence, sometimes provided by the grammatical rules of a certain language and sometimes provided from the situational context in which the sentence is written/expressed. But we also know that natural language could be at the same time obscure, ambiguous and unpredictable.

One of the main challenges in sentiment analysis is certainly given by the

automatic recognition of frequent expressions naturally used in human language. And, for example, decipher if an expression is positive or negative, thus expressing its *sentiment*. Among many different applications, it is fairly common to study *polarity orientation* through the repetitive use of certain patterns or the exploitation of features.

General and common applications of sentiment analysis often come from pre-existing disciplines that have been recently remodeled through a new point of view: that of searching a sentiment, opinion, or in general a content that expresses subjectivity of the author.

If we consider text-categorization, for example. Traditionally, it sought to classify documents by topic (which could be also several dozens). Now, in contrast, sentiment analysis classifies texts only within relatively few classes (e.g. *positive* vs *negative* or some degree of them).

As we will see in a later section it has always been hard to formalize rhetoric devices even in a traditional framework, and the harder it is now in a computational perspective. Just to mention the fact that people express opinions in rather complex ways, with a wide and rich vocabulary choice.

Moreover it has been demonstrated that in opinion texts, the study of lexical content alone can be highly misleading due to the common presence of intra-textual reversals, negations or sudden topic changes. Additional devices that can completely misdirect the correct comprehension of a text are rhetorical modes such as general humor, sarcasm, irony, implications and so on. The present thesis will focus on the annotation of irony in perspective of an improvement of results for automatic detection.

With the explosion of Web 2.0 platforms, such as blogs, discussion forums, peer-to-peer networks, and various other types of social media ... consumers have at their disposal a soapbox of unprecedented reach and power by which to share their brand experiences and opinions, positive or negative, regarding any product or service. [31].

The quotation above is intended to disclose the prime reason of why we have a deep connection with social media in this thesis: first of all they are a huge

pond of freely accessible data, secondly they mostly contain sentiment and opinion-texts.

## 2.1 Features exploited in sentiment analysis

Let us consider the most exploited sets of features used for the retrieval of sentiment and polarity applied by systems. One interesting challenge would be to ask ourselves if the features that function well with computers are the same that also humans exploit when they speak, read or and try to decode polarity.

1. **Term presence and frequency.** That is exploiting the presence or of a certain word or set of words to determine polarity. Note that also the position of a token in a text unit can potentially have important effects the overall sentiment textual unit. As we know from linguistic theories a distinction can be drawn between FULL WORDS such as nouns, verbs conveying semantic content, and adjectives and EMPTY WORDS such as conjunctions, articles or prepositions, and that must be studied carefully. For example, on the one hand the conjunction “*and*” is one of the most present and frequent tokens in any type of text of every different topic but it provides a scarce contribution for the determination of sentiment and polarity. On the other hand rare words and even *hapax legomena* could be much more significant in their contribution to sentiment analysis.
2. **Parts of speech.** One simple reason holds for general textual analysis: part-of-speech tagging can be always considered useful and can be exploited in various different ways, it is, for example, a good premise for word-sense disambiguation. Dictionaries are usually organized in lemmata, therefore if we want to exploit PoS-tagging and dictionaries combined we also need lemmatization. We know, that normally the core of semantic information in a given sentence is found in full words such as nouns or adjectives, and even verbs. Part of speech tags could be



useful to discard empty words, that would not prove useful for sentiment analysis tasks. Furthermore certain adjectives are good indicators of sentiment, and sometimes have been used to guide feature selection for sentiment classification.

3. **Syntax.** Relating syntax to sentiment is a field of research that is still undergoing through a lot of study. It can be useful when compared to study, for example, collocations and even more complex syntactic patterns. State-of-the-art NLP technologies are capable of complex analysis exploiting syntactic structures, but they work well only in certain domains (e.g. newspapers, law and judicial texts). Nowadays they still do not work properly on social media texts, a field in which syntax is completely distorted from common writing rules. The same problem applies to PoS-tagging.
4. **Negation.** Handling negation can be an important concern in opinion- and sentiment-related analysis, being negation a huge polarity reversion element (just think to the sentences “I LIKE APPLES” vs. “I **DON’T** LIKE APPLES”).

I believe these four previous features to be the most representatives among all those that could be exploited in sentiment analysis tasks. Additionally to them, especially related to semantic content, one other useful approach is to use defined lexical resources as I will explain in the next section.

## 2.2 Lexical resources

In a work written by Pang [32] in which general aspects of polarity are described, such as the fact that it is commonly taken for granted that polarity is expressed through a binary opposition (even though some studies also admit a mixed polarity).

On this foreground, it is usually engaged a semantic approach, that is, human subjects (*annotators*) are asked to pick keywords that would be good

indicators of sentiment polarity (e.g. for/against, like/dislike, good/bad...). Words, in fact, are said to have a so called “*semantic orientation*”, meaning that they are either intrinsically positive or intrinsically negative. Once again, just this simple task might possibly lead to a series of difficulties because it is based on human perception of words’ meanings. It is certainly true that humans have a higher sensitivity than machines; but it is also true that at the same time they are prone to error. A technique that partially covers this problem is the development of *sentiment lexica* which are sort of dictionaries (basically: list of words) in which lemmas are annotated with a precise polarity.

Among the vast pond of lexical resources used in sentiment analysis we find **Wordnet**<sup>1</sup> a lexical database for English with emphasis on synonymy. In this “dictionary” nouns, verbs, adverbs and adjectives are grouped into synonym sets called *synsets*. The main idea is that different words are linked according to lexical and conceptual relations thus creating a “net”.

Figure 2.1 shows how a synset is to be intended: an ontological representation. In particular, WordNet is not specifically sentiment-oriented, but it has been used to help derive sentiment and related information. Hu and Liu [20], for instance, identify polarity for adjectives using Wordnet; they start with a set of *seed* adjectives of known (and obvious) orientation: such as “good”, “fantastic”, “wonderful”, “awful”, “terrible”, “bad”. For uncertain adjectives, they measured proximity via synonymy/antonymy relations with core adjectives<sup>2</sup>. What one can do is simply to evaluate the sentiment of sentences based on counts of positive vs negative polarity words (as it is mechanically determined by the Wordnet algorithm).

Another useful tool is **SentiWordNet**<sup>3</sup> a lexical resource for sentiment analysis, which is based on Wordnet synsets. This time, at each synset is manually assigned one out to three sentiment scores: positivity, negativity, and objectivity [3].

---

<sup>1</sup><https://wordnet.princeton.edu/>

<sup>2</sup>If an adjective is close in synonymy to positive words, or close in antonymy to negative words, it’s positive.

<sup>3</sup><http://sentiwordnet.isti.cnr.it/>



and multiword expressions still seem to create enough failures in automatic systems concerning polarity.

## 2.3 Datasets

Datasets in the field of Linguistics are commonly called corpora (plural for corpus). A corpus is, in fact, a large collection of texts. It is a body of written or spoken material upon which a linguistic analysis is based. A corpus can be either monolingual or multilingual, meaning that it can represent only one language or more languages.

A corpus may be composed of written language, spoken language or both and it provides grammarians, lexicographers, and other interested parties a better descriptions of a language under some respect. Computer-processable corpora allow linguists to adopt the principle of total accountability, retrieving all the occurrences of a particular word or structure for inspection or randomly selected samples. Corpus analysis provide lexical information, morpho-syntactic information, semantic information and pragmatic information.

Corpora are used in the development of NLP tools. Applications include spell-checking, grammar-checking, speech recognition, text-to-speech and speech-to-text synthesis, automatic abstraction and indexing, information retrieval and machine translation. Corpora are also used for creation of new dictionaries and grammars for learners.

Some corpora do not just provide a certain amount of text but have further structured levels of analysis applied. In particular, the more representative resources in this sense are those called *Treebanks*<sup>4</sup>, where texts are associated with annotations that describe them from a morphological and syntactic point of view.

Other levels of linguistic structured analysis are possible, including annotations for semantics and pragmatics. In each case, the annotation follows a set of principles encoded in a schema, the so-called annotation scheme, that

---

<sup>4</sup>Tree are parsed texts corpus where morphological and syntactic sentence structures are annotated.

is introduced in the following section. In our case a corpus will be annotated on the pragmatic level (concerning irony) and on sentiment polarity and an annotation scheme for this purpose has been developed.

## 2.4 Annotation schemes

Each classification task, like e.g. sentiment analysis, always has to start with a beginning phase of SUPERVISED LEARNING through examples, therefore knowledge-bases are often needed. This type of dataset is usually an annotated corpus which can have annotations on several different levels and thus be useful for different tasks.

Some researchers consider the possibility of retrieving sentiment (or *attitude*) in social media as a «fueling revolution in computational linguistics» and as a potential tool to get «real-time opinion from people around the world» [15]. Nowadays the attention is particularly focused on social media texts because in the past ten years an enormous quantity of publicly available texts has grown.

A 2011 paper [15], for example, deals with the pros and cons of a coarse- and fine-grained sentiment analysis of social media texts focusing on how important and essential human contributions on annotation are. Since sentences might contain multiple readings and unclear nuances, human annotations prove once again to be an essential and irreplaceable part of NLP. In their research they focus on three main aspects of texts, which they call: *polarity/valence*, *source* and *target*.

The concept of polarity should be clear by now; thus I will spend a few lines explaining the other two categories. Source is banally the person or group of people who express the opinion; while target is the thing, product or person towards the attitude is expressed. In their research the authors instruct three pairs of skilled annotators in this is a central and delicate part in every procedural work, because annotators have to work and annotate linguistic features according to their experience as native speakers but at the same time

they can not be left completely free; they have to follow a common frame, otherwise the annotations would be too sparse. That is precisely why it is important to have a unified annotation scheme and instruct all the annotators beforehand with the same weighted directives (I will further comment on this subject in the following chapters).

A common practice that usually follows annotation of data, is to calculate the *inter-annotator agreement*<sup>5</sup>. For instance, Fink et al. [15] define a variable called Krippendorff’s  $\alpha$ <sup>6</sup> which «generally if it is greater than 0.8 indicates superb agreement, whereas between 0.667 and 0.8 indicates good agreement».

The inter-annotator agreement is certainly a quantitative measurement, but not only. It has to be regarded also as qualitative measurement, considering that it provides a representation of a statistical sample of humans and how they relate to the use of language. In fact, annotators not only are essential for the successful completion of the annotation task, but also, being native-speakers of the language they deal with, they are representatives for the whole population speaking that language.

As demonstration of the granularity and specificity of annotations schemes, there are several studies of sentiment analysis more specifically focused on humor recognition [28, 29, 33, 34], where specific-designed schemes were designed. A statement that seems shared among many, is that there have been just few attempts to create computational models for humor recognition or generation. Of course we find ourselves in the field of creative use of language which goes «beyond banal useless humorless prose» as Mihalcea and Pulman define it [28]. In their paper they exploit humor-datasets with different characteristics to give automatic recognition machines a broad selection of features. For example they define a specific type of humorous text called “*one-liner*” which, as the name suggests, presents as a joke in one line and must produce the funny effect “in one shot” (very specific kind of humor). Then, they also retrieve

---

<sup>5</sup>It is the degree to which multiple human annotators arrive at the same annotations, when confronted with the same source text

<sup>6</sup>It is a statistical measure of the agreement achieved when coding a set of units of analysis in terms of the values of a variable.

data from the U.S. online newspaper *The Onion*<sup>7</sup> which is defined as «the best satire magazine in the U.S. and the best source of humor out there» [28] (for the purpose of the present thesis – that deals with Italian – the immediate association to the satirical blog *Spinoza*<sup>8</sup> is evoked).

## 2.5 Identifying Irony

We will now enter the area of research directly linked with the present thesis. In this paragraph we will provide some information on the detection of irony in sentiment analysis, discuss some of the most common methods and procedures that are widely used in this area and provide some examples of useful NLP tools, lexical devices and humor-related corpora.

First of all, as it has been shown both in psychology and linguistics, remember that humor and irony are strictly linked to one’s personal perception, one’s taste in jokes and one’s personal background. For these reasons creating a general and always-true annotation scheme which could be readable from a machine proves to be a rather challenging task to satisfy.

This means, in particular, that only a portion of the annotated material will be considered as useful for training sentiment analysis systems, that more representative of the shared knowledge about irony.

Already a lot of researchers in the past decade have put their efforts in small sub-tasks that in a hopefully bright-future will be connected one with the other and resolve our resolution. Also our research and particularly this thesis collocate in the same horizon. Specifically the creation of a corpus of humorous content, humanly annotated with the help of a fine-grained scheme.

Generally, as we have just mentioned, a corpus is certainly one of the most exploited resources in NLP, but their creation is also one of the more time-consuming tasks. As we have briefly mentioned before, corpora can be completely different from one another depending on which their purpose will be. In our area, that is sentiment analysis, a corpus usually must contain at

---

<sup>7</sup><http://www.theonion.com/>

<sup>8</sup><http://www.spinoza.it/>

least information about the *polarity* of the sentiment expressed in texts.

In a paper by Mihalcea and Pulman from 2005, the two authors inquire the identification of features characteristic of verbal humor and they find some interesting facts among their dataset, such as: a large use of negation (grammatical), a rather negative orientation (sentiment polarity) of texts, the fact that certain professional communities are repeatedly used as targets, a broad use of a “human-centric” vocabulary and many references to human “weakness” [29].

All these categories are to be intended in a semantical approach and of course they do not just emerge from nothing, but they clearly make reference to previous linguistic theories such as Attardo’s [1]. As more relevant than others, they propose:

- ALLITERATION – Jokes show structural and phonetic properties that are at least as important as their content.
- ANTONYMY – Humor often relies on some type of incongruity, opposition or contradiction.
- ADULT SLANG – Humor might include sexuality references or vulgarities (often realized through polysemy).

Reyes et al. [33, 34] propose a study with likewise features but with the addition of categories such as: SIMILARITY, STYLE or IDIOMATIC EXPRESSION.

The definition of humor-specific stylistic features proves to be fundamental in every following work with a similar intent (including ours).

To summarize, sentiment analysis is unquestionably a difficult task, and the more the nuance and complexity of opinions expressed increases, the more difficulty in retrieving good information rises. If we imagine all the quantity of things that is currently written on the web and domains or targets addressed in texts, we can imagine different levels of difficulty: as product reviews are relatively easy to retrieve correctly, while books, movies, art, music are more difficult. On another level but still hard to formalize are political discussions,



indirect expressions of opinion, because – for instance – political leanings are usually non-binary sentiments. And also, to conclude: is it true that words have a universal polarity? Or is it more plausible that a word’s polarity might slightly change when it occurs in different positions or contexts? Additionally, communication is much more than sharing a common code and often people are meant to infer meaning that goes beyond syntax or semantics. As we have seen, this proves to be a hard challenge for computers but it is the same for humans as well.

The present dissertation focuses on data entirely in Italian; currently all the resources I have listed so far, give satisfying results just for English. All the lexica and tools that exist for other languages, Italian included, have been built on automatic translation from the preexistent English resources.

A further challenge to be dealt are languages. For a language with a rich morphology such as Italian, sentiment analysis (and in general NLP) is prone to larger failures in data results. For example, this can be evident if we consider lexica currently exploited in sentiment analysis which result from translation from English. Let’s assume that the word-form *friend* in SentiWordNet has a positivity score of +3; an automatic translation in Italian would just output the masculine singular form *amic-o* with a score of +3, but it would not output the feminine form *amic-a* thus creating a defect in automatic retrieval of sentiment in two identical sentences one of which contains the feminine word-form and the other the masculine word-form: “Domani mi vedo con un’amic-**a**” vs. “Domani mi vedo con un amic-**o**”. Same problem applies also with verbs: Italian verbs have a different ending for each person in almost every mood and tense, while English appears to be described as almost the opposite situation. Thus, creating a new lexicon based on specific word-forms would help in one case but would be enormously difficult and time-consuming to build, not to mention the time of computation every time it has to be exploited, for instance with a common laptop of everyday use.

## 2.6 Related work

In this section we focus on sentiment analysis tasks on Italian and, in particular, on those research projects that are somehow predecessors to this work, because they concern the development of similar resources to the one that has been developed for this thesis. Only a few resources has been released until now for Italian, while more corpora has been released for English. Among them we will cite those that are under some respect more related to the present study.

In particular, a recent work by Stranisci et al. [39] aimed at applying sentiment analysis and opinion mining techniques to user-generated texts on social media with the intent of studying political sentiment, political alignment or generally speaking socio-political issues. This research involved the creation of a corpus for Italian, based on the topic of the controversial educational reform “Buona Scuola” (Good School). The dataset was collected retrieving the most popular hashtags referring to the school reform and later on was provided a manual annotation both for polarity and irony. The choice of creating an Italian corpus based on political topic is not random: the idea – strongly supported by linguistic theories – is that when it comes to express political believes or opinions, it is more likely for humans to use figurative language and rhetorical devices instead of speaking clearly and literally.

A similar research [23] studied how Twitter’s users exploit irony devices within their communication strategies in texts. In doing so a multi-layered annotation schema for irony was proposed for the first time, exploited in this work it was exploited with the broader aim of building a multilingual corpus for English, French and Italian. The annotation schema had three layers for the cataloging of 1) irony activation type, 2) irony categories and 3) irony markers; while the study itself focused mainly in finding correlations between the occurrence of certain values in the three different levels. An in-depth and detailed description of these two studies and some other previous related works will be handled in later chapters.

## 2.7 A broader panorama

As we have seen in the previous chapter, the present thesis collocates itself in a broader panorama of linguistic efforts into the annotation of humorous social media texts in Italian. This so called “broader panorama” could be summarized just by making a reference to EVALITA<sup>9</sup>, which is the *evaluation campaign of Natural Language Processing and Speech Tools for the Italian language*. The aim of the above-mentioned campaign is, in fact, to improve and support the development of resources and technologies for Italian, also covering the analysis of both written and spoken language at various levels of processing. Its first edition dates back in year 2007 and after four other editions, the last one was EVALITA2016 which includes the development of six different tasks – our major scope of interest. Among the various tasks, of notable mention in regard of the theme of this thesis, we find *SENTIPOLC – SENTiment POLarity Classification*<sup>10</sup>, whose aim consists in automatically annotating tweets indicating the message’s subjectivity, its polarity (positive or negative), and whether a message is ironic or not [4].

Another task worth mentioning is *PoSTWITA – POS tagging for Italian Social Media Texts*<sup>11</sup>, whose task consisted in Part-Of-Speech tagging tweets [9] and as dataset exploited the tweets that were part of the former EVALITA2014 SENTIPOLC corpus.

Even if it is not *strictly* part of the EVALITA campaign, a mention also goes to another “older” work that shares similar dataset of the two cited above. *Senti-TUT*<sup>12</sup>, «a project for the development of a novel Italian corpus for sentiment analysis, which includes sentiment annotations concerning irony and consists in a collection of texts from Twitter» [8]. This resource not only provides important annotations on polarity but also a second set of annotations concerning morpho-syntactic structures, in order to make available further exploitation related to sentiment analysis. Senti-TUT has a focus on politics,

---

<sup>9</sup><http://www.evalita.it/>

<sup>10</sup><http://www.di.unito.it/tutreeb/sentipolc-evalita16/>

<sup>11</sup><http://corpora.fclit.unibo.it/PoSTWITA/>

<sup>12</sup><http://www.di.unito.it/tutreeb/sentiTUT.html>

a domain where irony is frequently exploited by humans. The polarity and humorous tags exploited in this corpus are:

**POS** (positive),

**NEG** (negative),

**HUM** (ironic),

**MIXED** (POS and NEG both),

**NONE** (objective, none of the above).

As we can see from the five tags above, there is a coarse-grained choice to be done in order to annotate polarity. According to this scheme a tweet is to be tagged with just one of these categories; so a tweet, for example, is either POS or NEG: two tags can not be used at the same time. To represent that specific case a human annotator would have to use the MIXED tag. But what if a tweet is considered to be humorous and negative at the same time? Why should the presence of one tag exclude the use of another? To avoid such an impasse and to mark irony, in Stranisci et al «introduced two polarized ironic labels: **HUM NEG**, for negative ironic tweets, and **HUM POS** for positive ironic tweets» [39]. Later on in this chapter I will comment on the pros and cons of coarse- and fine-grained annotation schemes in comparison with our tagging experience.

Additionally, the Senti-TUT corpus is organized in two subcorpora: TW-NEWS (containing around 3,000 tweets, published in the weeks after the new Italian prime minister Mario Monti announced his Cabinet. October 2011 – February 2012) and TWSPINO (containing more than 1,000 tweets extracted from the Twitter section of Spinoza. July 2009 – February 2012).

As I previously mentioned, my work is not intended to be read alone but it is part of a bigger frame that focuses on the fields of language use in Italian social media texts, humor detection, especially irony.

# Chapter 3

## Linguistic theories

Among the many existent fields of linguistics and semiotics we will mainly focus on the area that holds into its core the importance of meaning and that primarily studies the ways in which context influences meaning itself when a sentence is uttered. We are going to deal with pragmatics and the most influential theories that have been expressed by linguists and philosophers of language.

### 3.1 Early studies on pragmatics

#### 3.1.1 Austin and Searle

As a first warm-up discourse on linguistic theories we definitely must quote the *Speech Acts Theory* which was mentioned for the very first time in 1955 by professor John Langshaw Austin during a lecture entitled *How Do Things With Words?*<sup>1</sup> at Harvard University. Later on, the same theory was recalled by the linguist John Searle and even better discussed by in his major work [35].

Austin, deeply inspired by the philosophy of Frege and Wittgenstein presented a new picture to analyze meaning. In his thoughts, meaning is described as a relation among three different entities: i.e. the linguistic conventions correlated with words/sentences, the situation where the speaker actually says

---

<sup>1</sup>The *Speech Acts Theory* was officially recorded on paper only posthumously, in 1962, in a book with the same title [2].

something to the hearer (context), and the associated intentions of the speaker. In Austin, the new description of meaning is truly understandable only if one takes into consideration the connections that exists among these three objects, which are also successfully depicted by the concept of *acts*. In uttering a sentence, that is, utilizing a linguistic convention, the speaker performs a linguistic act to the hearer with an associated intention.

Few years after Austin's new ideas had spread, the key concepts concerning a new definition of meaning were again studied and deepened by Searle. In his main work [35] he provides an exhaustive description of three facets of a speech act, stating that we are always able to find three coexistent "forces" or acts, entwined in one another, that take place every time a linguistic production is uttered. Speech acts are basically three distinct levels of action beyond the act of utterance itself. Searle distinguishes 1) the act of saying something, 2) what one does in saying it, and 3) the action that one performs after saying it. Respectively:

1. *locutionary act*: the production of a meaningful linguistic expression,
2. *illocutionary act*: the action intended to be performed by a speaker in uttering a linguistic expression, by virtue of a conventional force associated with it (illocutionary force).
3. *perlocutionary act*: the actual realization of consequences or effects on the audience through the uttering of a linguistic expression.

The Speech Acts Theory has proven extremely important and has been a relevant turning point to other theoreticians who dealt with pragmatics in the immediate following years as well as researchers that are dealing with computational approaches of pragmatics today.

### 3.1.2 Grice: towards a new approach

Almost in the same years, Paul Herbert Grice published *Logic and Conversation* [17] where he compares philosophical logic to natural language and thus introduces another key concept in the field of pragmatics: that one of *implicature*. He declares that there is a clear parallelism between some of what he

calls FORMAL devices:  $\sim$ ,  $\wedge$ ,  $\vee$ ,  $\supset$ ,  $(x)$ ,  $\exists(x)$ ,  $\int x$  and their counterparts in natural language i.e. *not*, *and*, *or*, *if*, *all*, *some* (or *at least one*), *the*. Because of this reason, Grice argues that it should not be difficult – theoretically – to formalize all expressions of natural speech, but at the same time he also provides evidence against this same opinion stating that it might be hard to precisely define, in practice, some expressions, which may not have a single truth value (i.e. one statement equal one meaning).

This controversial thought arises when the author explains that expressions, as they are commonly used in natural speech, cannot be regarded as finally acceptable and may turn out to be not fully intelligible. Sometimes there is a clash between what is uttered, what is literally meant and thirdly what is REALLY meant.

Trying to use Searle's terminology, one might say that in some situations there is a clash on the levels of locutionary and illocutionary acts. Here below an example:

[...] Suppose someone to have uttered the sentence *He is in the grip of a vice*. Given a knowledge of the English language, but no knowledge of the circumstances of the utterance, one would know something about what the speaker had said, on the assumption that he was speaking standard English, and speaking literally. One would know that he had said, about some particular male person or animal  $x$ , that at the time of the utterance (whatever that was), either 1)  $x$  was unable to rid himself of a certain kind of bad character or trait or 2) some part of  $x$ 's person was caught in a certain kind of tool or instrument. [17].

What Grice is stating through this example, is that the sentence here exploited is not completely intelligible at first sight, but it is rather unclear, and to have a full identification of what the speaker had said, one should need to know a few more things, such as the identity of  $x$ , the time of utterance, the meaning, on the particular occasion of utterance of the phrase *in the grip of a vice*.

## Implicatures

To better clarify the ambiguities like the one we have just described, which are more common in natural speech than people might think, Grice introduced a new terminology in linguistics: the verb *implicate* and the related nouns *implicature* (cf. *implying*) and *implicatum* (cf. *what is implied*). The creation of these new terms is due to the fact that Grice considered traditional description of utterance meaning and literal meaning not exhaustive enough and a finer distinction in terminology was needed. Grice was the first to systematically study cases in which what a speaker means differs from what the sentence used by the speaker means. He also wished to represent a differentiation between what one SAYS strictly speaking and what one INTENDS TO SAY with the same words. On this regard, the newly introduced term “implicature” denotes the act of meaning or implying one thing by saying something else, therefore, “implicating” is also what Searle [35] called an *indirect speech act*.

Conversational implicatures have become one of the most discussed subjects in pragmatics; but beware: they are not to be confused with entailments<sup>2</sup>. The most widely recognized forms of implicature are the figures of speech. Irony, overstatement (hyperbole), understatement (meiosis and litotes), metonymy, synecdoche, and metaphor have been known at least since Aristotle. Later on in this chapter we will focus on such figures of speech to provide familiar examples of clashes between literal meaning and intended meaning just as Grice intended to explain in his major work and we will clearly see that most of the problems for such theories are linked to the fact that speakers often have other goals in mind but to communicate.

---

<sup>2</sup>In semantics and pragmatics, an entailment is the principle by means of which – under certain conditions – the truth of one statement (A) ensures the truth of a second statement (B). Entailments can also be found in literature as *strict implications*, *logical consequences* or *semantic consequences*. One famous example is *The president was assassinated.* entails (B) *The president is dead.* Notice also that if (B) is false, then (A) must necessarily be false.



## The cooperative principle and the maxims

One other great contribution in the field of pragmatic studies, that was formulated by Grice, is the one of the COOPERATIVE PRINCIPLE. It is basically a sort of script that – according to the author – should be followed in every production of speech whatsoever. It is the assumption that:

Our talk exchanges do not normally consist of a succession of disconnected remarks, and would not be rational if they did. They are characteristically, to some degree at least, cooperative efforts; and each participant recognizes in them to some extent, a common purpose or set of purposes, or at least a mutually accepted direction. [17].

Grice describes the Cooperative Principle formulating four different categories: Quantity, Quality, Relation and Manner. **QUANTITY** relates to the amount of information to be provided, and it is better explained by the following maxims:

1. Make your contribution as informative as is required (for the current purposes of the exchange).
2. Do not make your contribution more informative than is required. [17].

**QUALITY** is represented by a general statement “Try to make your contribution one that is true” and two more specific maxims:

1. Do not say what you believe to be false.
2. Do not say that for which you lack adequate evidence. [17].

**RELATION** is defined only through one single maxim: “Be relevant” which – Grice himself admits – is a bit blurry and vague.

**MANNER** is the only one on four categories that do not focuses on *what* it is said but rather on *how* it is said. It includes the maxim “be perspicuous” and others such as:

1. Avoid obscurity of expression.
2. Avoid ambiguity.
3. Be brief (avoid unnecessary prolixity).

4. Be orderly. [17].

Of course each one of these categories and the associated maxims have a different weight and a different importance. There are, for example, all sorts of other maxims such as “be polite” that are normally observed even outside the boundaries dictated by the cooperative principle.

### **Flouting of the maxims**

It is time for us to underline more clearly why these preambles are fundamental to be kept in mind while reading the following chapters and why they are essential to the development of this thesis. While the definition of the cooperative principle is now rather clear, one can not say the same when it occurs to describe its application to real world conversations. For example, a participant may fail to fulfill a maxim when they speak. Here below some examples:

1. They may quietly and unostentatiously VIOLATE a maxim; if so, in some cases they will be liable to mislead.
2. They may OPT OUT from the operation of both maxims and cooperative principle; they might be unwilling to cooperate in a way maxims require.
3. They may be faced by a CLASH: they might be unable, for example, to fulfill the first maxim of Quantity without violating the second maxim of Quality.
4. They may FLOUT; that is, they may BLATANTLY fail to fulfill it.

After having introduced many key concepts of pragmatics it is now time for us to draw a precise line of connection between those theories and the core topic of our research i.e. irony.

As we already mentioned plenty of times, the generation of natural speech something that is really hard to define in a fixed model, because it entails procedures that follow rules, and other procedures that are completely unpredictable and directly connected to the context. Because of that, it might occur that participants in a conversation do not strictly follow the maxims and the cooperative principle, thus creating a series of particular situations which are not easy to be depicted into a traditional frame. For instance, there are situations, in which no maxim is violated, or at least it is not clear that

any maxim is violated. There are cases then, in which a maxim is violated, but its violation can be explained by the supposition of a clash with another maxim (e.g. being vague in an reply to a question – conflicting with Quantity – because they don’t know the precise answer – maintaining accordance with Relevance and Quality). Another frequent situation, and here it is where we finally connect with our study, is the case in which a maxim is flouted for the purpose of getting in a conversational implicature by means of something of the nature of a figure of speech [17]. The exploitation of rhetorical devices such as irony, metaphor, hyperbole, euphemism can sometimes lead to misunderstanding among participants of a same conversation. In those situations, it is often said something which is false or even untrue, but participants can understand the violation of maxims in regard of the cooperative principle and they might even share a common knowledge of the world. For instance, according to Grice, irony is a conversational situation in which the first maxim of Quality is flouted:

IRONY — X, with whom A has been on close terms until now, has betrayed a secret of A’s to a business rival. A and his audience both know this. A says “*X is a fine friend*” [...] It is perfectly obvious to A and his audience that what A has said or has made as if to say is something he does not believe, and the audience knows that A knows that this is obvious to the audience. So, unless A’s utterance is entirely pointless, A must be trying to get across some other proposition than the one he purports to be putting forward. This must be some obviously related proposition; the most obviously related proposition is the contradictory of the one he purports to be putting forward. [17].

What we have been asking ourselves while conducting this research is precisely inquiring the nature of that thing or common ground that makes “*what A has said [...] and he does not believe [...] perfectly obvious to A and his audience*”. Furthermore, is it something that we can describe in a formal way? And again: is it always true that irony has always to be decoded as the opposite of what one has just uttered? If so, how can humans understand each other while they

speak? To answer these questions and many others related to the same topic, we must consult further linguistic research on the subject of irony.

## 3.2 Later theories on irony

### 3.2.1 Beyond gricean maxims: Sperber and Wilson

A new epoch of linguistic theories concerning irony, embraces the former fundamental work of Grice and thus some later scholars are usually depicted as “Neo-Griceans”. Many of them took Grice’s theories and added some modifications or instead criticized Grice for his lack of precision and under-specification. Among all, the most influential alternative to Gricean and Neo-Gricean theories worth mentioning is the *Relevance Theory* developed by Dan Sperber and Deidre Wilson. The authors argue that «all Grice’s maxims can be replaced by a single principle of relevance». [38] In fact, if a speaker is as relevant as possible in every circumstance (and that is granted from the cooperative principle), *Relevance* alone provides a sufficient proof for all situations that Grice’s maxims were designed to explain.

According to the authors, in fact, the word *relevance* is given in a highly technical sense, roughly meaning *communicative efficiency*. The **PRINCIPLE OF MAXIMAL RELEVANCE** (COMMUNICATIVE EFFICIENCY) yields: «contribute that which has the maximum ratio of contextual effects to processing cost» [38]. This means that while Grice’s maxims force the speaker to communicate efficiently, they do not require maximization (as intended by Sperber and Wilson). Conversely, the principle of maximal relevance does not force to follow Grice’s rules. Let us compare three sentences to provide an explanation. Imagine a situation in which the speaker is a waiter and his reply choices are confined to these following alternatives:

1. We are serving chicken.
2. We are serving meat.
3. We are serving chicken or  $(7^2 - 3)$  is not 46.

Sperber and Wilson explain that (1) would be maximally relevant because it contains everything that also (2) does and it is even more specific, while being as easy to process. Instead, (3) has the same *contextual effects*<sup>3</sup> as (1), but is harder to process. Numerically speaking: in (3) there are more words than in (1) and half of them is irrelevant or useless to the purpose of mutual understanding. Sperber and Wilson, thus suggest that Grice's theory has some flaws, namely that it does not support situations in which an utterance is "*maximized*" at its best. It is in a sense contradictory. However, also their theory has similar deficiencies. For example, the principle of maximal relevance clashes with the simple PRINCIPLE OF POLITENESS, that we mentioned before, as badly as the cooperative principle does. Imagine parents deciding what to say after listening to their daughter play the clarinet for her school recital. "Your performance was horrendous" (maximized sentence) seems as easy to process as "Your performance wasn't perfect" (figure of speech – litotes). As human speakers, we understand immediately, then, that *maximization* as proposed by Sperber and Wilson also would not be the perfect solution in every situation.

### **A first glance to figurative language**

Later on in their research path, also Sperber and Wilson dealt with the ambiguous topic of figurative speech, in particular – interestingly for our study – they analyzed irony and provided a new theory and an innovative formalization for it [37, 40]. They point out that traditional studies on irony usually subsume that any ironical utterance is LITERALLY saying one thing and FIGURATIVELY meaning the opposite. They argue that a traditional explicit semantic theory is not good enough to express the deeper level of analysis of certain type of utterances; for example – they state – «why a speaker should prefer the ironical utterance *What lovely weather* to its literal counter part *What awful weather*» if the second sentence is exactly what the speaker thinks? According to them

---

<sup>3</sup>Roughly, what a proposition adds to the representation of the world that is already given in the context.

a traditional overview is not clear enough to explain those kind of utterances and that is precisely why we need to look things from another point of view. Grice [17, 18] for example, moves a step forward using a pragmatic approach to irony and it looks more promising than the traditional semantic one.

Grice's overview proposes that the utterance *What lovely weather* has no figurative meaning at all, but creates instead a conversational implicature, that is: the weather is awful. Grice, in fact, exploits the new terms he created (i.e. implicature) to express his idea on figurative language. However, as we can see, and also as Sperber and Wilson argue, there is still enough room to debate on this topic and certainly all problems are not solved simply by shifting from semantic to pragmatic domain. Grice's theory seems to hold when it has to justify WHAT happens when we utter one ironic sentence (which entails the creation of an implicature), but it can not provide a good reason of WHY someone should choose to say something when they actually intend to express its opposite. Also, uttering an ironic sentence, in a sense, could be considered as a violation of the cooperative principle or at least of one of the maxims.

Sperber and Wilson [37], after having expressed a comparison between rhetoric and linguistics, explain that, in general, approaches on figurative language should be taken really carefully being that rhetorical judgments or linguistic jokes are solely inferable from one's own intuition. The full comprehension of a joke or an ironic utterance is not universal, although they have been able to draw a general theory of rhetorical devices. Their idea is entirely based on psychological and interpretative mechanisms, which – they add – remain invariant from culture to culture.

After having explained that «the notion of irony is an abstract one, based on a rather arbitrary range of examples which have themselves been rather inadequately described [...] there is a whole range of utterance-types that can be more or less loosely called ironical» [37]. Therefore, they propose a new conceptual scheme of irony, deriving their inspiration from the framework of classical rhetoric.

After their doubts on the functionality of Grice's maxims, another con-

fictual point they argue against their peer is the one of figurative language. They say that in a situation in which figurative language is used, the speaker normally intends to convey something **INSTEAD OF** one of the literal senses of his utterance.

From Grice's perspective, instead, using figurative meaning (such as irony) contributes to the creation of an implicature that has to be seen as **SUBSTITUTING FOR** the literal sense. In his theory, one must look for a sort of "duplication" of sentences: on the one hand there is the figurative expression that is uttered (pointless) on the other hand there is a new implicature (meaning the opposite of what is uttered) and, according to Grice, here it is where the actual meaning is conveyed.

For Sperber and Wilson, the idea that an implicature, generated from a sentence, could actually contradict the literal sense of that precise sentence itself, simply does not hold. As we have previously commented, *an implicature denotes the act of meaning or implying one thing by saying something else*, therefore – in ironic situations – it cannot simply negate the meaning of what has just been uttered. Grice's and Sperber and Wilson's point of views highly separate in the description of where, how and why irony is actually generated.

## **A description of irony**

In order to better clarify the notions of figurative meaning and in particular the device of irony, Sperber and Wilson propose some data and a new theory for irony description. Here below a clear example that they provide to express their innovative idea.

Consider the sentences (1)-(8) below exchanged between two people caught in a thunderstorm in circumstances that are otherwise normal:

- (1) *What lovely weather.*
- (2) *It seems to be raining.*
- (3) *I'm glad we didn't bother to take an umbrella.*
- (4) *Did you remember to water the flowers?*
- (5) *What awful weather.*
- (6) *It seems to be thundering.*

(7) *I'm sorry we didn't bother to bring an umbrella.*

(8) *Did you remember to bring in the washing?*

There are two obvious ways of grouping these examples. First, there are close syntactic and lexical parallels between (1) and (5), (2) and (6), (3) and (7), and (4) and (8). Second, in a less straightforwardly definable way, (1)-(4) have something in common which distinguishes them from (5)-(8). [37].

What we are mostly interested in, of course, is the “less straightforwardly definable way”, i.e. ironic *vs.* non-ironic distinction. And precisely we will focus more on the first four sentences.

The interesting question that we want to ask ourselves and the reader is: How do can irony be recognized? (Especially when no particular distinctive intonation is used). If we perceive irony, astonishingly unambiguously as just humans can do, and we are certain about the correctness of our detection, but also we know that ironic sentences usually mean the opposite of what they say, we should definitely ask ourselves why – for instance – the speaker in (1) thought the weather was awful but did not say so directly? What is the difference in saying *What lovely weather* ironically and saying *What awful weather* (5) literally?

In (2) rather than expressing the contrary, irony is expressed by saying just less (remember, the participants find themselves in a thunderstorm) and being thus, **IRONICALLY UNINFORMATIVE**. This example is not different from what we said some paragraphs before about the blatant flouting of maxims used to produce clash in meaning.

Example (3) is an understatement, similar to what happens in (1); because the speaker intends to say the opposite of what it is uttered but it might be an echoic mention or a reference to something that happened before. For instance, as to mention a previous “linguistic context”, we might imagine a situation in which, two friends were arguing whether to take an umbrella or not before going out. Imagine one of the two said something like “Don’t bother to take the umbrella” and at the end they decided to leave it at home. As we already know, they would soon find themselves completely soaked in water due to an



unexpected thunderstorm. On this terms, (3) is immediately recognized both as ironical and as echoic mention.

Sperber and Wilson do not provide names nor a deeper description for such different uses of irony, but only going through some brief example sentences, we can brightly see in how many different ways ironic devices can be exploited.

### **The Use-Mention Distinction**

One of the most discussed theories by Sperber and Wilson is the rather philosophical *Use-Mention Distinction* in which they express the importance of drawing a line to differentiate *i*) sentences in which the **USE** of an expression involves reference to what the expression refers to (external reference); and *ii*) those cases in which the **MENTION** of a sentence involves reference to the expression itself (meta-linguistic).

The latter case is clearly evident in (4), where someone who asks *Did you remember to water the flowers?* during a thunderstorm, can not possibly be serious and intend it literally. Therefore, it is clear that the sentence must be read with a figurative meaning. The further implication is that these types of question are usually pointless and the focus is not on WHAT IS asked but just on the fact that IT IS being asked. It is not simply *used*, but it is *mentioned*. They indeed have something in common with rhetorical questions.

As a result, what the theory proposed by Sperber and Wilson implies, is that all standard and non-standard cases of irony seem to involve (generally implicitly) a mention of a previous proposition. Additionally, a certain degree of pointlessness is also present; i.e. irony entails the presence of something useless, redundant or not fully related.

Another relevant point discussed by Sperber and Wilson and for which they provide good evidence for is the one that says ironical utterances are aimed at a particular target (later named “victim” also in Clark and Gerrig [13] as we will see in the next section) and if the audience who is intended to be the addressee of the irony fails to understand it, the cause is to be looked for in a difference in common ground or knowledge of the context.

### 3.2.2 Another point of view: Clark and Gerrig

A brief mention also goes to Clark and Gerrig [13] and their *Pretense Theory of Irony* which has proved, at the same time, to be both in agreement and disagreement with Grice, Sperber and Wilson theories. Renewing some key concepts studied from their colleagues, Clark and Gerrig propose a notion of irony in which the participants of a conversation are not simply USING a proposition in order to be ironic but they are, in fact, PRETENDING to use that proposition (hence the name *Pretense Theory*).

What they mean, is that when a participant uses the rhetorical device of irony, he is, in fact, behaving like an actor who portrays his fictional character in a play. They refer directly to Grice's theories according to which «to be ironical is, among other things, to **pretend**, and while one wants the pretense to be recognized as such, to announce it as a pretense would spoil the effect» [18]. Clark and Gerrig explain pretense theory as it follows.

Suppose S is speaking to A, the primary addressee, and also to B, who may be present or absent, real or even imaginary. When he speaks ironically, S is pretending to be someone else. Furthermore S and A share more common ground than the one that S and B share. Therefore A immediately understands when S is being ironic, while B in naive ignorance, is intended to miss the *pretense*, because (in virtue of the cooperative principle) he takes S words as he was speaking sincerely (and literally).

Because of this reason, A is part of the so called “inner circle”, that is A has the same common ground in knowledge as S. Hence, A sees everything: he sees what S utters, he understands the pretense, he immediately comprehends S's figurative talking, he sees B's ignorance and difficulty to understand what “lies beyond” the actual literal meaning.

I have found their proposal on the description of irony to be interesting and original and I would like to underline some relevant highlights which globally describe their idea on the subject: 1) asymmetry of affect, 2) victims of irony

and 3) ironic tone of voice. Aside from being particularly iconic in their description, they also point out some key nodes that are not dealt with by other theoreticians. Looking back at Jorgensen [21], they explain that an ironic participant in a conversation is more likely to say “*what a clever idea!*” of a bad idea than “*what a stupid idea!*” of a good one. They explain that the basis for that to happen is to be found in cognitive psychology and theories that explain how it is more common to see the world through norms of success and excellence, therefore using a positive expression to refer to something negative – they call it *asymmetry of affect*. A second category they introduce is called *victims of irony* and takes for granted the fact that irony is uttered through a process that produces victims. According to the general rule of the pretense theory, victims could be of two kinds: S – the unseeing or injudicious person the ironist is pretending to be and b the uncomprehending audience not in the inner circle. The third and last class the describe is the one of the *ironic tone of voice*, stating that «if speaking ironically has to be [...] the expression of a certain sort of feeling, then a tone suitable to such an attitude is mandatory at any rate» [13].

To better clarify, let us assume three friends Paul, John and Sarah are going to the cinema together to see the latest Tarantino’s movie. As foreground, we know that Paul and John absolutely love the director and especially, as they are talking before entering the cinema, they hope that the new movie, would be as full of blood scenes as Tarantino’s movies usually are.

Although Sarah, is somewhat fragile and she does not like particularly violent scenes (the two boys both know it), she decides to go to the cinema with them anyway. Additionally the girl has never seen any movie directed by Tarantino, therefore she does not know what to expect from the night out. She does not even know that Tarantino is John’s and Paul’s favorite director.

After some time, when the movie has ended and it proves to be, in fact, full of violent bloody scenes – as expected by the two boys – Paul says to both his friends: “*Great plot! It’s just a pity there weren’t that many bloody scenes*”.

Imagine now, and referring to Clark and Gerrig theory, that Paul is S, John

(who shares a common background with Paul) is A and Sarah, who is not part of the *inner circle* is B.

What happens next, is that John immediately understands that Paul must be ironic. The movie, in fact, has proven to be as great as they both expected it to be (and they were speaking of it before going in the cinema). Meanwhile Sarah, who is still shocked by the vision, does not understand Paul's intention to be ironic. In this example the **PRETENSE** is shown by the fact that Paul acts as if he is disappointed by the vision.

Sarah is the **VICTIM OF IRONY** because, not being present when the two boys were commenting about before the movie even started, she does not share the same knowledge background with them, hence fails to understand what Paul *really* wants to convey.

The fact that Paul states that the bloody scenes were not enough, when he namely means the opposite and he is fully satisfied, is at every level an **ASYMMETRY OF AFFECT**.

We can imagine a situation in which Paul could have used a blatant **IRONIC TONE OF VOICE** and also Sarah could have understood the irony and the pretense immediately.

### **Pretense or echoic mention?**

Furthermore, Clark and Gerrig also move a critic against the theory of irony proposed by Sperber and Wilson arguing that *Pretense theory* appears to be superior to the *Use-Mention theory* particularly when considering the psychological account of all the functions and processes of irony.

As we explained before, for Sperber and Wilson irony in an utterance is either use or mention. However, being that simplistic, their theory is forced to say that many ironic expressions are merely *implicit echoes* – echoic mentions of popular wisdom or received opinion – but it does not describe any criteria for deciding what is a possible implicit echo and what is not. According to Clark and Gerrig, Sperber and Wilson's theory is lacking of something in the sense that there should be something as a list or a precise group of features in

order to establish what can be an echo and what can not.

Let us consider the following example provided by Clark and Gerrig [13] about Johnathan Swift's essay *A Modest Proposal*<sup>4</sup>. In this work, the author suggests that the impoverished Irish population might ease their economic troubles by selling their children as food for rich gentlemen and ladies. This satirical hyperbole mocked heartless attitudes towards the poor, as well as British policy toward the Irish in general.

This famous piece of work of English literature is often pointed out as a true example of ironic text; Clark and Gerrig point out that it is mainly because Swift exploited a very peculiar technique to express his satire. Namely he used the rhetorical device of creating the *Proposer* (suggesting the purchase of children) with a serious and pompous tone of voice to highlight the absurdity of his proposal.

Clark and Gerrig argue that if one had to refer to *Use-Mention theory* to explain the irony present in *A Modest Proposal*, they would have to pinpoint that the ENTIRE ESSAY is an echoic mention. Which is absurd, because echoic mentions – as intended by the two theoreticians – are precisely recalled word for word. For this and other examples provided in [13] it appears that, on this point, the *Pretense theory* proves to have a more powerful argumentation than the rival one.

Additionally, they underline the fact that they give more attention also to the psychological side of irony production, considering into their research the background between speaker and their audience. They declare, through quite clever examples, that the perception of irony is bound to the notion of common ground between the participants of a conversation. It is often subdued to subtle judgments of what is common ground to whom; so a listener not supplied with the right information may not make these judgments accurately (and consequently fail to understand irony itself).

On this account the pretense theory – even though at first sight might ap-

---

<sup>4</sup>Original title: *A Modest Proposal For preventing the Children of Poor People From being a Burthen to Their Parents or Country, and For making them Beneficial to the Publick*, 1729.

pear as a harsh critic – provides transparent explanations for several important features of irony mentioned by Sperber and Wilson [37].

To summarize, as it was for Grice, also for Sperber and Wilson irony involves a change of meaning. This particular type of figurative speech is called **VERBAL IRONY**. Differently from Grice, noticing that an ironic sentence not always represents the opposite meaning of what it is uttered, Sperber and Wilson feel the urge to distinguish between the concepts of *use* and *mention*.

The two authors also discuss the existence of an “ironical tone of voice” in traditional framework stating that the introduction of a suprasegmental description is quite bizarre and they argue that if that is the case there should be also a “metaphorical tone of voice” or a “synecdochical tone of voice”. According to them it is ordinary for ironical utterances to involve a switch in style and register, hence «it is quite common [...] a change to a more formal or pompous style» [37] but there is no mention of it in the traditional framework.

One other point they deal with, is that one they refer to as *asymmetry* (also renewed by Clark and Gerrig [13]); that is it seems more likely to express something bad with a positive sentence than the other way round.

After having introduced and described all these theories and facts, it is finally time for us to explain their connection with the precise goal of our research and disclose their degree of importance in our study, namely irony annotation in social media texts. Specifically we have found ourselves deeply in agreement with the theory provided by Sperber and Wilson, with whom we share common beliefs and methodology.

# Chapter 4

## Twitter

In this chapter I will introduce to the reader one of the resources that has been fundamental for the developing of this thesis: the microblogging platform Twitter. In particular, this networking service has proven to be useful when we collected the texts to build our corpus, given the free accessibility of data that is provided. As I will explain in the following chapters, our analysis is focused on Italian in the context of social media texts.

### 4.1 A microblogging platform

Twitter<sup>1</sup> is an online news and social networking service that was created in March 2006 and launched in July of the same year. Sometimes defined as “the SMS of the Internet” it has created an online environment in which users can type short messages (*tweets*) that can reach a maximum length of 140 characters. Nowadays Twitter is the second largest microblogging platform and it counts more than 500 million subscriptions and more than 340 million tweets per day worldwide [12].

A tweet contains a lot of interesting data: apart from its text, many other fields (metadata) with additional information about the tweet are provided. For example, tweets contain a field with their creation date, whether or not it is a retweet (a tweet that has been copied from another user as it is considered interesting), or even coordinates (about 1% of Twitter users provide their geo-

---

<sup>1</sup>[www.twitter.com](http://www.twitter.com)

graphical position more or less consciously through their smartphone's GPS). Additionally, the tweet's user is known.



Figure 4.1: Example of Twitter's interface

As we can see from figure 4.1 each user's profile shows a number of followers and friends, a language, a screen name as author of the tweet (users can exploit either their real name or nicknames), and so on [19].

While writing a tweet, users can exploit some additional resources other than simple text that have become really typical in social networks and microblogging platforms, such as *hashtags*, user's tags, link to URLs or pictures. In particular hashtags were so exploited in the past five years that the term has become so popular to be added to the *Oxford English Dictionary*<sup>2</sup> in June 2014.

Hashtags, depicted from Chiusaroli [11] as «the most peculiar textual elements in Twitter» are nowadays largely exploited in Twitter, Facebook or Instagram; they are basically “labels” used to find messages with a specific theme or content. Users create and use hashtags by placing the hash character # (also known as the number sign or octothorpe) in front of a single word or unspaced phrase (*scriptio continua*).

The most popular hashtags in the last years, aggregated in macro-topic categories<sup>3</sup>, for example, have been: #OneDirection, #NowPlaying for mu-

<sup>2</sup><http://www.oed.com/>

<sup>3</sup>Retrieved from: <https://2015.twitter.com/top-trends>



sic; #TheWalkingDead and #Empire for TV; #iPad, #Android, #iPhone concerning tech world; #ISIS, #CharlieHebdo and #JeSuisCharlie for news. Concerning politics the most used shared labels have been #WakeUpAmerica, #Obama; #NFL, #NBA and #Football for sports and finally #Love, #Travel, #Selfie, #Food for free time.

Looking for a certain hashtag with a search tool, present in almost every web site and platform, will yield each message that has been tagged with it (this technique is largely exploited to retrieve data in many researches). Furthermore, hashtags can be inserted anywhere within a sentence, either preceding it, following it as a postscript, or being included as a word within the sentence (e.g. "It is #sunny today"), thus their importance is both figurative, semantic and syntactic.

## 4.2 Twitter as corpus

Microblogging today has become a very popular communication tool among Internet users. Twitter is used to post emotional states, updates on personal life events or to share beliefs on various different topics. According to a 2009 survey<sup>4</sup> by *Pear Analytics*<sup>5</sup>, 40% of tweets are personal updates, 37% are conversations, 9% are retweets, 6% are advertisement, 4% are spam, and a last 4% are news. As we can see from this digits and as more and more users post about products and services they use, or express their political and religious views, Twitter has increasingly become a valuable source of people's sentiments and opinions [30].

If on the one hand using Twitter as a corpus for sentiment analysis provides many useful information, on the other hand we must take further precaution when using it for social analysis. Moreover it can be especially challenging from a linguistic point of view. Many linguistic studies show, in fact, how written language changes enormously in e-mails, chats, blogs and newsgroups that is

---

<sup>4</sup><http://pearanalytics.com/wp-content/uploads/2012/12/Twitter-Study-August-2009.pdf>

<sup>5</sup>A digital marketing agency with a strong background in analytics and solving complex tracking and conversion issues with online marketing.



when they differ from common grammar rules and instead they produce such iterations?”.

Bazzanella [14] describes this effect as a manner to “obtain brevity by lengthening” which sounds oxymoronic but it clearly explains how briefness in comprehension is achieved employing a visual mean (the exaggerated repetition) in contrast with the grammatical norm, usually exemplary of written texts. Other similar devices that are broadly used in Twitter to better catch other users’ attention are, for instance, sentences written in all-caps (often decoded as yelling), emoticons, abbreviations, hashtags, mentions and URLs [25].

In the first chapters of the present thesis we have dealt with an introduction to sentiment analysis and we described its area of interest. We have mentioned the main linguistic theories linked to pragmatics and especially the description of irony, and introduced the reader to the main characteristics of the famous social media platform Twitter. In the following section I will finally enter the main core of this thesis.

# Chapter 5

## Methods for data annotation

Consider that being the present dissertation a master thesis, not all the resources as in a “real” research project were exploitable. To start, an ideal job would have seen the exact same amount of data annotated from (at least) three different annotators. Here, instead, a prototypical study is presented, which can be usefully exploited for validating the procedure and scheme used for the annotation of the novel resource by studying a smaller sample data set.

The tweets of the corpus were annotated by two annotators and only a smaller portion was annotated from three people, namely, those where the first two did not achieve the agreement in annotation. This provides also some statistical data to further support the argumentations that are going to be treated in chapter 6. Idealistically, the annotation work will not stop after the defense of the present thesis but it will continue, in order to have a complete annotated corpus that could also be usable within the context of sentiment analysis applications and studies.

### 5.1 Development of the corpus

I will now focus on the two main steps that took place during the creation of the corpus, firstly the retrieval and collection of data and secondly the annotation procedure.

### 5.1.1 Data collection

The collection of data for the corpus that I will exploit for this thesis has been relatively simple. As we have already stated numerous times, the preset research collocates in a multilingual project on irony detection in social media texts. Therefore to build a suitable corpus I had to follow pre-existent guidelines that have been already exploited for other languages such as French and English and are explained in Karoui et al. [23].

I did not have to retrieve tweets by myself directly from Twitter but for my thesis I have exploited tweets collected for previous researches [7, 8, 23, 39].

As it is well explained in the most recent article on this same subject, by Karoui et al. [23], the tweets for both the French and English corpora were retrieved by using the Twitter API<sup>1</sup> capturing all texts containing specific *hashtags* that are usually exploited by users to self-mark their humor (*#irony*, *#sarcasm*, *#sarcastic*). The same procedure could not be used to retrieve Italian tweets because, even though Italian users actually exploit a series of humorous hashtags, no long-term established single hashtag is shared among users (an “*#ironia*” hashtag is not really present and exploited). For this reason, Italian data were collected with other methods. They were, in fact, extracted from existing annotated data: the SENTIPOLC corpus [5] and TW-SPINO [8], which were both already annotated for the PRESENCE or ABSENCE of irony.

Corpus	Irony	Not Irony
F	2,073	16,179
E	5,173	6,116
I	806 (Sentipolc) +2,273 (TW-SPINO)	5,642 (Sentipolc)

Table 5.1: Distribution of tweets in each corpus [23]

As we can see from the digits reported in table 5.1 both the French and English corpus have an amount of ironic tweets and an even bigger amount of not ironic

---

<sup>1</sup>In computer programming, an application programming interface (API) is a set of sub-routine definitions, protocols, and tools for building application software.

tweets. This scenario does not apply in the case of the corpus developed for the present thesis, because we have selected only tweets that were at least annotated as humorous if not precisely ironic.

Our goal is not to focus on deciding whether a tweet is ironic or not, but, given only humorous tweets, annotate them with a fine-grained description of the features and patterns that can cause irony in written texts.

### 5.1.2 A new corpus: TWITTIRÒ

In a first phase of our work, about 1,100 ironic tweets were extracted from pre-existing corpora. Some of them had to be discarded because they were advertisement, *retweets*, repetitions, or incomplete. Although, for the present thesis, only a smaller portion of tweets will be annotated.

In table 5.2 are shown the digits of the three sub-corpora that compose the annotated corpus we have created during the development of the present thesis: TWITTIRÒ<sup>2</sup>.

Corpus	Number of tweets
TW-SPINO	198
SENTIPOLC	100
TW-BS	95
<b>TWITTIRÒ</b>	<b>393</b>

Table 5.2: Distribution of tweets in TWITTIRÒ corpus

Even though all the tweets are of humorous content (mostly ironic) the three sub-corpora are really different from each other and present heterogeneous characteristics.

- **TW-SPINO** contains tweets that have been collected from the satirical blog SPINOZA.IT. Before being published on Twitter, they usually undergo a strict procedure of selection already operated from the working team of

---

<sup>2</sup>The complete corpus will soon be available for consultation at the web page <https://github.com/IronyAndTweets/Scheme>. At the same link, also the French corpus will be available. An evaluation campaign on the corpora is still ongoing, therefore data could not be published until the process is over.

Spinoza itself. For this reason, the language used in these tweets usually is more grammatically correct and it is highly refined with a high register and style. The topics concerned are variegated but with a clear preference for jokes concerning the world of politics and general news.

- **SENTIPOLC** is less homogeneous than the previous sub-corpus. First of all, it contains tweets generated by “common users” and not a team of skilled satirical writers. Hence, the use of language is mixed as one would actually expect, given the breath of the Internet. Grammar is sometimes poor and the use of colloquial expressions, vulgarities or other devices typical of CMC is very frequent. There is a frequent use of creative hashtags, mentions to other users and features such as the repetitions of laughters. The topics covered here are mainly politics with references to the technical government of Monti between 2011 and 2012, because of a filtering applied in part of the original data collection.

- **TW-BS** is a corpus containing just tweets on the debate of the controversial reform of Italian school and educational sector, i.e. “Buona Scuola”. Here, as in SENTIPOLC, the tweets retrieved are generated by “common users” therefore the use of Italian language is far from the norms of standard writing. Once again, a lot of devices typically exploited in CMC are shown. Furthermore, being the reform of the education system a highly criticized one, the use of sentences written in ALL CAPS (to decode shouting) is wide.

A further and extensive description of the tweets, will be provided in chapter 6 with an enumeration of the main patterns that are exploited by users to encode ironical content, and a discussion on their effects on readers.

## 5.2 Following a common scheme

The presence of multiple annotators is fundamental in each annotation work. In particular, studies like ours, which deal with figurative use of language need

the presence or more than just one person. Firstly because, as we have seen, the exploitation of irony and its comprehension rely on completely different ground among speakers.

Secondly, because, if understanding irony is somehow “intuitive”, modeling it into a common frame is not. Other annotators provide a good team to talk with and to discuss with any doubt regarding the annotation task. Inevitably, working alone on such a repetitive task as annotation, can cause the generation of certain biases which could spoil the meaningfulness of data.

The final aim is, anyway, to have all the 1,100 tweets annotated by three different annotators and adding even more ironic tweets in order to reach a similar amount of data as the French or English corpora, so that the statistics would be better comparable.

As I just mentioned, three different people were involved in the annotation task for the present master thesis. Two of them are skilled annotators, who have previously dealt with similar annotation tasks and even contributed to the creation of the annotation scheme that was used. In fact, a brand new multi-layered annotation scheme for irony was proposed in Karoui et al. [23] for the first time. It is also available in the Appendix.

### 5.3 Presentation of the scheme

The first steps towards a common scheme were made in [22], where the first attempts to model irony were also made. Despite linguistic theories differ on how to define irony, they all commonly agree that it involves a clash between the literal meaning of an utterance and what is expected about the speaker. Let us consider the tweet in example 5.1:

(5.1)        @BlitzQuotidiano Continua così #Super #Mario #Monti! A breve lavoreremo solo per risanare il PIL! A pranzo e cena veniamo a casa vostra!

Here the sentence “Continua così #Super #Mario #Monti!” can be inferred as ironical only if some additional common knowledge is shared among speakers. That is, the user who tweeted it does not really want Mario Monti to “keep on



doing like that”, in fact, he hopes precisely the opposite. Some more situational or conversational context is needed. In [22], it is shown how there are two ways to infer such context:

- 1) rely exclusively on the **lexical clues** internal to the utterance,
- 2) combine these clues with an additional **pragmatic context** external to the utterance.

Later on, in [23], this same distinction will be marked as the difference between EXPLICIT irony and IMPLICIT irony. Implicit irony is due to an opposition that is created between a lexicalized proposition  $S$ , describing an event or state, and a pragmatic context external to the utterance, where  $S$  is false or is not likely to happen.

In other words, the writer asserts or affirms one thing ( $S$ ) while he really intends to convey another one ( $Q$ ):

- $Q = \text{Not}(S)$

or

- $Q \neq S$ .

The irony occurs because the writer believes that his audience can detect the disparity between  $P$  and  $P1$  on the basis of contextual knowledge or common background shared with the writer.

Beside this first type of contradiction, that generates ironic sentences, in [23], are also studied different types of irony and the features they are expressed with. In figure 5.1, retrieved from the document also available in the Appendix, is presented the scheme for irony annotation in tweets that we followed to create our corpus. The creation of the scheme, and the categories that are described rely completely on the linguistic theories that we mentioned in chapter 3.

The goal of this thesis is also to try the scheme on a real data sample, and to discuss it showing that it can be further extended for representing specific cases which often occur in the data set annotated until now.

The scheme presents four different levels of annotation: level 1: CLASS, level 2: CONTRADICTION TYPES, level 3: CATEGORY and level 4: CLUE WORDS.

Those categories were structured to reply to simple questions such as: “Can the types of irony studied in Linguistics be found in social media such as Twitter?”, “If yes, what are the most frequent types?”, “Are these types explicitly marked? What are the correlations between irony types and these markers?” and “How can we exploit these correlations in a purpose of automatic detection?”.

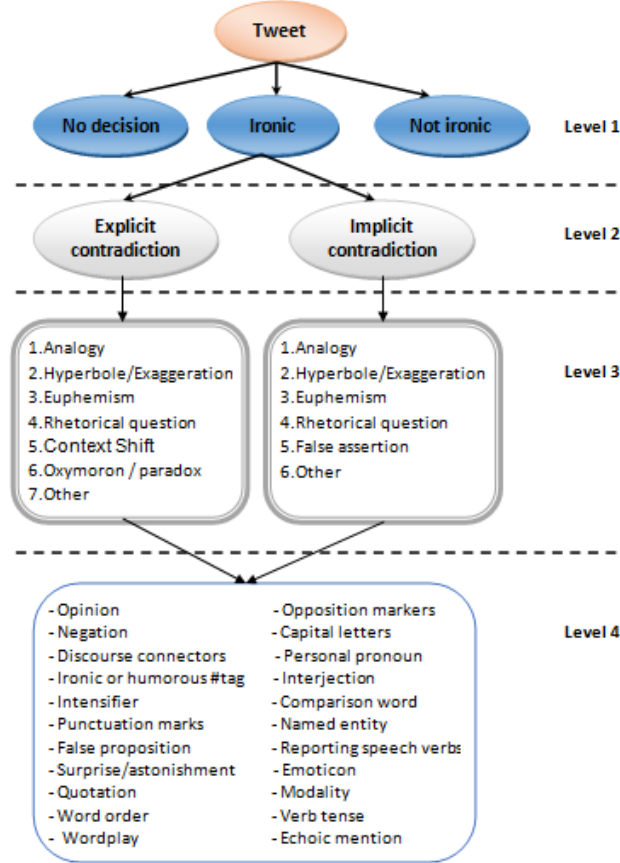


Figure 5.1: Annotation scheme for irony

In the following paragraphs I will focus on giving an explanation of the four layers. I will exploit significant tweets extracted from TWITTIRÒ as examples.

### 5.3.1 Level 1: Class

The first layer would involve the classification of tweets into **ironic** or **not ironic**, but it does not apply to the work of this thesis, because, in our case,

a pre-selection was made, and we are hence treating only ironic tweets.

### 5.3.2 Level 2: Contradiction types

As stated from various linguistic theories, irony is often exhibited through the presence of a clash or a contradiction between two elements. In tweets, it could be a clash between two entire propositions, for example, P1 and P2.

The propositions P1 and P2 can be both part of the internal context of an utterance (explicitly lexicalized), or one is present and the other one implied, as part of some external context. Because of that, we operate the following distinction:

EXPLICIT	IMPLICIT
Irony is activated only by two lexical clues present in the utterance.	Irony is activated by clues present in the utterance plus additional pragmatic context.

#### EXPLICIT CONTRADICTION

It can involve a contradiction between proposition P1 and proposition P2 that have opposite polarities. In example (5.2) the opposition is shown from by the pair of semantically opposite words *liberate* and *processate*.

- (5.2) [Liberate]<sub>P1</sub> Greta e Vanessa. Saranno [processate]<sub>P2</sub> in Italia. [@maurizio-neri79]

Explicit contradictions can also present a clash between lexicalized elements that would not be expected to appear together. Not precisely opposite as in example 5.2. But, for instance, in (5.3) the mother of a grown up politician subscribes him to a new political party. In (5.4), instead, *antidepressants* and *European elections* are clashing because they not pertain to the same semantical sphere.

- (5.3) [La mamma]<sub>P1</sub> di Civati: “Nel Pd non si può più stare”. Se continuano così [lo iscrive a un altro partito.]<sub>P2</sub> [@pirata\_21]

- (5.4) #NicolaMorra su #labuonascuola in preda ad allucinazioni; sono la conseguenza di abuso di [antidepressivi]<sub>P1</sub> assunti dopo [elezioni europee]<sub>P2</sub>.

## IMPLICIT CONTRADICTION

On the other hand, an implicit contradiction rises from a contradiction between a lexicalized proposition  $P_1$  describing an event or state and a pragmatic context external to the utterance in which  $P_1$  is false or is not likely to happen.

The irony occurs because the writer believes that his audience can detect the disparity between  $P_1$  and  $P_2$  on the basis of contextual knowledge or common background shared with the writer.

- (5.5) ["Se davvero abbiamo pagato è uno schifo" ha detto Salvini guardando la laurea di Renzo Bossi.] <sub>$P_1$</sub>  [faro]  
→  $P_2$ : Renzo Bossi got his Master's degree by paying with his father's party money.
- (5.6) [La buona scuola e le sillabe - <http://t.co/nS42fRjAKp>] <sub>$P_1$</sub>   
→  $P_2$ : The official document that presented the reform had hyphenation mistakes.
- (5.7) [Avvincente la nomina dei sottosegretari del governo Monti. Che tanto mica abbiamo problemi più seri...nooooo, noooooo...] <sub>$P_1$</sub>   
→  $P_2$ : The nomination is not compelling at all. And there are certainly more serious problems.

There are cases in which irony is activated in multiple ways inside a tweet. It might occur, that on one superficial layer irony is explicitly activated from lexicalized cue words, and on a second "hidden" layer there is a deeper level of irony, inferable only through additional pragmatic knowledge. In this case, the tweet has to be annotated as **IRONIC WITH IMPLICIT CONTRADICTION**.

### 5.3.3 Level 3: Categories

Explicit and implicit contradictions, according the common scheme, can be expressed through different rhetorical devices, patterns or features that are grouped under the label of *categories*. In [23], eight different categories are defined at a further level of analysis. Two of them can occur only in presence of an explicit type of irony, while one is specific for the implicit type. The remaining five can be associated with both types.

Categories are not mutually exclusive, in fact they can co-occur in the same tweet. I will mostly explain them through the advice of examples.

For a broader description of the eight categories, refer to the Appendix.

## ANALOGY

In the category of analogy are summoned also other figures of speech that comprehend mechanisms of comparison, such as SIMILE and METAPHOR. In (5.8) an analogy is drawn between the footballer Lionel Messi and the Italian minister Maria Elena Boschi because of their authoritarian fathers. **Pure** (furthermore) serves as clue word. In (5.9) the analogy is activated by the construction “*x reminds me of y*”. Instead, in (5.10) the clue words are “*there is x and also y*”.

- (5.8) Leo Messi: "Firmo quello che mi dice papà". **Pure** la Boschi. [notturmoconcertante]
- (5.9) Il governo #Monti **mi ricorda** la corazzata kotikmin.
- (5.10) @fattoquotidiano Quest'anno è peggio del solito: **oltre** all'amianto c'è **anche** #labuonascuola.

## HYPERBOLE/EXAGGERATION

It is a figure of speech which consists in expressing an idea or a feeling with an exaggerated way. It can be expressed through the use of superlative adjectives such as in (5.11) or (5.12). Either with the use of hyperbolic expression as **aberrazione** (aberration) in (5.13), or indefinite collective adjectives (or pronouns) such as **tutti** (everyone) in (5.14).

- (5.11) Il trionfo di Tsipras in Grecia è **il più grande successo** della sinistra italiana dai tempi di Zapatero in Spagna. [@gmbugs]
- (5.12) #labuonascuola “Sarà **bellissimo** ascoltare la voce di tutti,...” @matteorenzi la senti questa voce?....(cit. coro da stadio )
- (5.13) @masechi Si è già assistito a Porta a Porta alla simulazione di un governo Monti con ministri La Russa e Bindi. **Aberrazione** audiovisiva.
- (5.14) #M5S #Renzi, se tra un anno non ci saranno 170 mila insegnanti di ruolo in più, te li porto **tutti** a @Palazzo\_Chigi #labuonascuola.

## EUPHEMISM

It is a figure of speech which is used to reduce the facts of an expression or an idea considered unpleasant in order to soften the reality. In (5.15) is exploited the use of the partitive expression “*a few/a little*”. Instead in (5.16) irony through euphemism is activated by the expression **senza fretta** (no haste). In (5.17) is used the common device of punctuation such as quotations to soften one’s way to express their own opinion.

- (5.15) Nel 2006 Charlie Hebdo aveva pubblicato delle vignette satiriche su Maometto. Ci hanno messo un po’ a capirle. [nicodio]
- (5.16) @palazzochigi Professor Monti, c’è una certa attesa per la lista #ministri. Senza fretta, però tenga conto della nostra ansia...
- (5.17) “”” buona scuola ”””.

## RHETORICAL QUESTION

It is a figure of speech in the form of a question asked in order to make a point rather than to elicit an answer. It can be direct and explicit as in (5.18) and in (5.19). Or it can be an indirect rhetorical question as in (5.20)

- (5.18) Giovanardi applaude la sentenza Cucchi. Cosa vi aspettavate da una frase che inizia con “Giovanardi”? [CONTINUA su <http://t.co/oDPUtx2DvV>].
- (5.19) Mario Monti? non era il nome di un antipasto? #FullMonti #laresadeiconti #elezioni #308.
- (5.20) #renzi blocca gli stipendi per 10 anni, noi dobbiamo fare #labuonascuola non so se sanno che anche gli insegnanti hanno una famiglia.

## CONTEXT SHIFT (EXPLICIT ONLY)

It occurs by the sudden change of the topic/frame in the tweet, as in (5.21), where the first sentence is about pupils’ pocket money, while the second is about the price of ice cream. The same happens in (5.22), where the first clue word is about a Romani encampment, while the second about a safari journey.

- (5.21) @matteorenzi Più che la #labuonascuola direi #carascuola visto che ci vogliono più di **800 euro a piscello**....quasi quanto **5 kg di gelato**.
- (5.22) L’auto di Salvini assalita al **campo rom**. Rovinato il **safari**. [@paniruro]

Under this broader category are also included the cases of **REGISTER CHANGING** in which the “context shift” is due to a sudden change of linguistic style, exploitation of vulgarities as in (5.23) or, on the contrary, a rather pompous style. In Italian tweets, users often recur to the exploitation of dialectal expression as in (5.24).

(5.23) Salvini: “Al Quirinale vorrei Vittorio Feltri”. È già abituato a mettere la firma su delle **stronzate**. [pirata\_21]

(5.24) Mario, Monti sulla **#cadrega**.

#### **FALSE ASSERTION (IMPLICIT ONLY)**

It indicates that a proposition, fact or an assertion fails to make sense against the reality. The speaker expresses the opposite of what he thinks or something wrong with respect to a context. External knowledge is fundamental to understand the irony (it is, in fact, implicit only). In the following examples, the sentences written in correspondence of the right arrow are propositions, which are not lexicalized, but must be inferred from the reader to understand irony.

(5.25) “Potrei non opporre veti a un presidente del Pd”, **ha detto Berlusconi iscrivendosi al Pd**. [CONTINUA su <http://t.co/oDPUtxTq-U7>]  
→ Berlusconi never subscribed to the PD party.

(5.26) Vedo che c'è molta disinformazione sul referendum del **17 maggio**. [Mister-Donnie13]  
→ The referendum is on April, 17th – not May.

(5.27) Totoministri per il governo Monti: **Gelmini ai lavori pubblici, farà il tunnel dei neutrini!**  
→ Minister Gelmini was never in charge of public work administration. It is also a reference to an erroneous statement about neutrinos that the Minister had previously uttered.

#### **OXYMORON/PARADOX (EXPLICIT ONLY)**

This category is equivalent to the category **FALSE ASSERTION** except that the contradiction, this time, is explicit. Also in this subsection, the sentences written in correspondence of the right arrow are propositions, which are not lexicalized, but must be inferred from the reader to understand irony.

- (5.28) Legge elettorale, il Pd si divide. Non vedono l'ora di provarla. [@maurofodaroni]  
 → It is absurd to think that the Italian political party PD has undergone an internal division in order to try a new electoral law they promoted.
- (5.29) Individuata una mafia tipicamente romana. Prima di mezzogiorno non prendeva appuntamenti.  
 → It is common knowledge that people from Rome are often late, thus the paradox of creating a criminal organization that is also often late.
- (5.30) Brunetta sostiene di tornare a fare l'economista, Mario Monti terrorizzato progetta di mollare tutto ed aprire un negozio di pescheria.  
 → It is paradoxical that the Italian Prime Minister would leave the government to open a fishmonger's.

## OTHER

This last category represents ironic tweets both with explicit and implicit contradiction, which can not be classified under one of the other seven previous categories. It can occur for example in case of humor or situational irony.

The tag “other” is also applied when there is a number of overlapping categories, and thus, it is hard to define which one should be tagged first.

In (5.31), for example, we notice a pun in bold. *Larga intesa* is a common collocation in Italian to describe a good shared agreement, while *BANCA INTESA* is a well known Italian credit institution. Furthermore, the graphic device of caps lock is used.

Instead, in (5.32) irony is expressed through the use of another type of pun, exploiting assonance “buono-scuola” (monetary ticket used to buy books or office products) vs. “buona scuola” (the reform).

- (5.31) Il Governo Monti parte....c'è **larga BANCA INTESA**. <http://t.co/x0-u6nt7b>.
- (5.32) Dal **buono-scuola** alla **buona scuola** renziana: buone ragioni per lo sciopero <http://t.co/kduFO6MLdd> #10\_ottobre #quota\_96 #sciopero.
- (5.33) L'Euro entrò in vigore l'**1 gennaio 2002**. Trasformandolo immediatamente nel **2 gennaio 4004**. [CONTINUA su <http://t.co/pJhE2CMi70>]
- (5.34) Sicilia, **arriva barcone di migranti** e a bordo c'è anche un **gatto**. **Vengono a rubarci i nostri like**. [@LughinoViscorto]

In the tweet (5.33) we have the overlapping of a *false assertion* (what it is stated is not true), an *hyperbole* (the mathematical doubling of a date) and the *need*



*of external knowledge* (prices doubled in all Europe with the introduction on the new communitarian currency). So maybe also an *analogy*?

In the example (5.34) is shown the application of racism, through the exploitation of a shared knowledge, almost an idiomatic expression: “*foreigners come here and they steal X from us*”. Furthermore, to understand a second layer of irony, the additional knowledge of social media is needed: cats are a subject that makes people on the net turn crazy, and hence, they receive a big amount of LIKES (Facebook jargon).

#### 5.3.4 Level 4: Clues

Clues represent words that can help annotators to decide in which category belongs a given ironic tweet, such as **like** for analogy, **very** for hyperbole/exaggeration. Clues include also negation words, emoticons, punctuation marks, interjections, named entity (and mentions), as depicted in figure 5.2.

Since the extraction of the information about this level can be done, to a great extent by automatic tools, we did not address this specific annotation task within the present research.

For a more detailed explanation of the four layers, please refer to the Appendix.

During the development of this thesis, it was not given a particular priority to the tagging of level 4 and clue words, but as we will see in chapter 6 precedence was given more to level 2 and level 3.

Annotation of clue words, is something that has to be done through the exploitation of a dedicated annotation tool: **Glozz**<sup>3</sup>. It is a task that, in any case, will be further continued also after the printing of this master thesis.

---

<sup>3</sup><http://www.glozz.org/>

Units	Attributes
<b>Irony with explicit contradiction</b>	Analogy = {No, Yes}
	Hyperbole/Exaggeration = {No, Yes}
	Euphemism = {No, Yes}
	Rhetorical question = {No, Yes}
	Register changing = {No, Yes}
	Oxymoron/paradox = {No, Yes}
	Other = {No, Yes}
<b>Irony with implicit contradiction</b>	Analogy = {No, Yes}
	Hyperbole/Exaggeration = {No, Yes}
	Euphemism = {No, Yes}
	Rhetorical question = {No, Yes}
	False assertion = {No, Yes}
	Other = {No, Yes}
<b>Not ironic</b>	No attributes
<b>No decision</b>	In case of indecision (annotators do not understand the tweet because of cultural references or lack of background knowledge)
<b>Clues</b>	Discourse connectors = {No, Yes}
	Punctuation = {No, Yes}
	Opinion = {+, -, neutral, No}
	Emoticon = {+, -, neutral, No}
	Opposition markers = {No, Yes}
	Capital letters = {No, Yes}
	Intensifier = {No, Yes}
	Comparison word = {No, Yes}
	Modality = {No, Yes}
	Negation = {No, Yes}
	Quotation = {No, Yes}
	Interjection = {No, Yes}
	Personal pronoun = {No, Yes}
	Reporting speech verb = {No, Yes}
	Surprise/astonishment = {No, Yes}
	Named entity = {Person, Object, Place, Animal, Event/Fact, Function, Organization, Age, Duration, Date, Period, Day, Weight, Temperature, Percentage, Length, No}
	False proposition = {No, Yes}
	Irony or humorous #tag = {No, Yes}
	Word order = {No, Yes}
	Wordplay = {No, Yes}
	Verb tense = {No, Yes}
	Echoic mention (proverbs, songs, poem)={No, Yes}
<b>Tweet</b>	Tweet Id → already defined
	Internal Id of the tweet → already defined
	Global opinion of the tweet = {+, -, neutral, No}

Figure 5.2: Units and their attributes in the annotation scheme

# Chapter 6

## Data analysis

In this chapter I will present an in-depth analysis of data from the TWITTIRÒ corpus. In section 6.1 I will describe specific characteristics and features of the tweets related to the way in which irony is exhibited. Secondly, in section 6.2 I will discuss some concerns that have emerged among annotators during the annotation process. I will lastly, draw some conclusions and suggest new directions for further work.

### 6.1 Features and patterns

During the task of manually annotating tweets, we have been able to recognize some recurrent characteristics through which irony or general humor is expressed in social media texts. Some of these characteristics are exhibited through features and some of them create more complex patterns.

Generally speaking, CMC is highly informal, and this “informality effect” is reinforced by superficial punctuation and orthography. As Cerruti and Onesti describe in [10], *Netspeak* (i.e. language on social media or CMC) presents features such as certain lexical items, spelling variants, letters-plus-number combinations, multicapitalization, *leetspeak*<sup>1</sup> that are features, which seem to pertain the domains of lexicon and graphematics.

In this section will investigate tweets from the point of view of their char-

---

<sup>1</sup>Leet (or “1337”), is an alternative alphabet for many languages that is used primarily on the Internet. It uses some characters to replace others in ways that play on the similarity of their glyphs via reflection or other resemblance.

acteristics. In particular I have drawn a coarse subdivision for features in SYNTACTIC, GRAPHIC, SEMANTIC and PRAGMATIC devices, which we will examine in the next sections. Furthermore, ironic tweets have a general structure which is always common. According to the multi-layered annotation scheme (see also Appendix), in fact, ironic tweets are always composed by two different propositions, namely P1 and P2. When the two propositions are read together, a clash or contradiction arises and thus irony is generated.

As we have seen in chapter 5, the TWITTIRÒ corpus is composed by tweets, which belong to three different sub-corpora. In particular, we refer to the data extracted from TW-SPINO as being “controlled and supervised”, in the sense that before getting published on the Twitter account owned by the satirical blog, they undergo a selection and a grammatical revision. The same does not happen for the data from TW-BS and SENTIPOLC, which belong to the “free speech” part of the Internet.

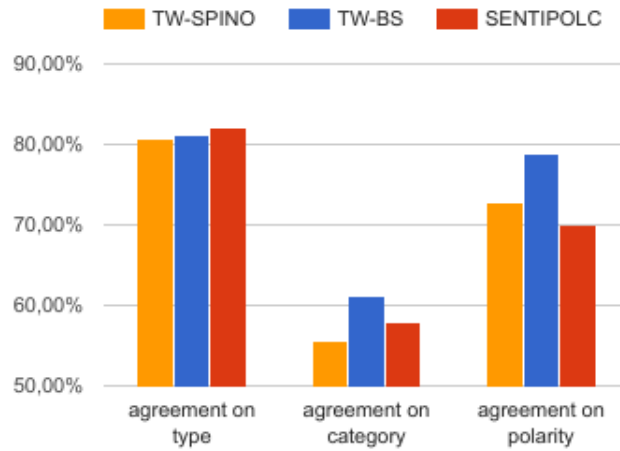


Figure 6.1: Inter-annotator agreement on type, category and polarity in the three sub-corpora.

Being the data collected from *Spinoza.it*, always grammatical and written in a high register, we assumed it would have been easier to annotate them and that inter-annotator agreement would have been relatively high. Our expectations leaned on the fact that language exploited in TW-SPINO is grammatical and easier to understand, but we forgot to consider that, thus, irony is more subtle

and that the overlapping of multiple one rhetorical devices or categories occurs quite often.

As we can see from figure 6.1 the inter-annotator agreement is lower in TW-SPINO (orange) than in TW-BS (blue) or SENTIPOLC (red) both in TYPE and CATEGORY. In POLARITY the lowest agreement is reached in SENTIPOLC. I am not completely sure on why this happens, but it might be interesting, if these percentages hold in the future also with an expansion of the corpus to return on this point.

As we cited in chapter 2, according to Fink [15], when Krippendorff's  $\alpha$  «is greater than 0.8 indicates superb agreement, whereas between 0.667 and 0.8 indicates good agreement». From figure 6.1 we notice that we have obtained a “superb agreement” on the definition of TYPE. A “good agreement” on POLARITY, and a really low score on the definition of CATEGORY.

One theory in explanation to this, can be that actually, language and figurative devices in TW-SPINO are always “controlled” and carefully exploited thus creating a subtler type of irony than the one that is generated in the two other sub-corpora.

### 6.1.1 Syntactic devices

As we have mentioned in chapter 4, tweets can have the maximum length of 140 characters, showing that complex structures are usually avoided by users in this text genre. As studied by Bazzanella in [6], and Chiusaroli in [12], the maximum size of a tweet induces indeed the users to prefer short nuclear sentences.

#### **TW-SPINO STYLE**

In particular, as it was just mentioned above, the structure of tweets retrieved from TW-SPINO is highly regular and repetitive. It always consists of two short sentences separated by a full stop (P1 and P2). Irony is created because the link between the two sentences is absurd, not coherent or paradoxical. In (6.1) an absurd analogy is depicted between the comedian duo “Soliti Idioti”

(usual idiots) and the people who work in the government. The tweet in (6.2) is about the freedom of press, but then the second sentence “we can not tell anything else” ironically disproves the former one. In example (6.3), it is drawn a parallelism between the catching of a well known member of the mafia and the success at the beginning of the football transfer market.

- (6.1) I Soliti Idiotti in scena a Sanremo. Ieri erano alla Camera. [@maurizioneri79] #dopofestival
- (6.2) Pubblicata la classifica mondiale della libertà di stampa. Non possiamo dirvi altro. [giga]
- (6.3) Alfano: “Stiamo per prendere Messina Denaro”. È il primo colpo del mercato invernale. [@maurofodaroni]

Often, such as in the examples (6.4), (6.5) and in (6.6) almost half of the 140 available characters is occupied by URLs not leaving much other space for real lexical content.

- (6.4) Sequestrate immagini di Wojtyla palesemente false. Erano perfettamente a fuoco. [CONTINUA su <http://t.co/oDPUtxkMK3>]
- (6.5) I pm di Potenza incontrano Maria Elena Boschi. L'inchiesta era partita con questo obiettivo. [CONTINUA su <https://t.co/oDPUtx2DvV>]
- (6.6) Ferrara, Gasparri e Salvini dichiarano guerra all'Isis. Ok, andate. [CONTINUA su <http://t.co/oDPUtxTqU7>]

#### QUOTATION (OR QUESTION) AND ANSWER

Another frequent syntactic pattern is to exploit the combination of two sentences, such as in the one above, where the P1 (underlined) is a direct quotation or a question, often rhetorical, and the P2 (bold) is a reply, given by users themselves.

In particular in (6.7) and (6.8) are underlined two direct quotations supposedly uttered by Grillo and Giannini, while the part in bold is an ironical reply given by the user.

In (6.9) the user expresses his thoughts through the combination of a rhetorical question (P1, underlined) and gives himself

- (6.7) Grillo: “Papa Francesco ci ha copiato il programma”<sub>P1</sub>. **Tranne la parte del libero arbitrio.**<sub>P2</sub> [@andr3\_21]
- (6.8) #labuonascuola Giannini: “I sindacati della scuola non sono interlocutori privilegiati”<sub>P1</sub>.. **Certo, meglio la pseudoconsultazione on line :)** <sub>P2</sub>
- (6.9) Da che dipende?<sub>P1</sub>... **Chiediamolo alla “buona scuola” del salvator della patria di Renzi!**<sub>P2</sub> <http://t.co/xb4JKX3BYM>

#### TRUE SENTENCE. FALSE SENTENCE.

Another similar pattern is the one that exploits the combination of P1 as a true sentence TS (something true, a world-wide known fact or a direct quotation), followed by P2 representing a false sentence FS.

Irony is produced from the clash of meaning between TS and FS (or as the common frame calls them P1 against P2), but that occurs only because FS is *false* or *deviated* from interpretation of TS. Most of the time, this pattern is found in IMPLICIT CONTRADICTIONS because if the reader does not perceive the illogical clash, also does not perceive the humorous content.

This type of pattern is present in 81 tweets out of 393 (20,61% of TWITTIRÒ). Among them, 65 are in TW-SPINO, which is almost a third of the sub-corpus (32,83% of TW-SPINO).

In example (6.10) the true sentence (P1 = TS) is a real news fact, while the false sentence (P2 = FS) is a hasty ironic conclusion drawn by the user.

More frequently the P1 is a direct quotation, such as in the following examples. Respectively, direct speech in example (6.11) is uttered by the secretary of the political union UIL, while the (6.12) and (6.13) are uttered by two politicians of the Italian Democratic Party (PD), Moretti and Renzi.

In all four the examples of this brief section, P2, i.e. the false sentence, is represented by a false, hurried inference.

- (6.10) [Terrorismo, studente universitario espulso dall'Italia]<sub>TS</sub>. [Ma sarebbe bastato aspettare che si laureasse]<sub>FS</sub>. [@pirata\_21]
- (6.11) [Il segretario Uil Barbagallo: “Oggi fermiamo l'Italia per farla ripartire”]<sub>TS</sub>. [Deve aver studiato informatica.]<sub>FS</sub> [@7semola]
- (6.12) [Alessandra Moretti: “Rosy Bindi ha mortificato la bellezza”]<sub>TS</sub>. [Così lei si è vendicata sull'intelligenza]<sub>FS</sub>. [@batduccio]

- (6.13) [Renzi: “Senza Napolitano non si possono fare le riforme”]<sub>TS</sub>. [Ecco perché tutta questa fretta]<sub>FS</sub>. [CONTINUA su <http://t.co/oDPUtx2DvV>]

## JOKES: ONE-LINERS

Mihalcea and Pulman, in their paper, [28], discuss the features of “*one-liners*”. A one-liner is a short sentence with comic effects and an interesting linguistic structure: simple syntax, deliberate use of rhetoric devices and frequent use of creative language constructions meant to attract the readers’ attention. While longer jokes can have a relatively complex narrative structure, a one-liner must produce the humorous effect “in one shot”, with very few words.

One frequent pattern that I found out to be often exploited in in TWITTIRÒ and, more precisely, in TW-SPINO, is something for which I propose the name “**TWO-LINERS**”, because it has the same characteristics of what is described above, but instead of soliciting a laugh in “one-shot” – irony is more subtle than common jokes – “two shots” are needed.

In (6.14) we see a pattern of *question + answer*, while in (6.15) and (6.16) we can see more common two-liners.

- (6.14) cos’è quella cosa che fai in pochi minuti ma che ti frega per 9 mesi? L’iscrizione a scuola. #labuonascuolauncazzo

- (6.15) Alfano chiede ai prefetti di annullare i matrimoni gay. Sembra l’inizio dei Promessi Sposi. [@frandiben]

- (6.16) Arrestato Scajola. Stavolta l’alloggio lo offriamo noi. [@ABkualcosa]

## 6.1.2 Graphic devices

### PUNCTUATION

One other device highly exploited in CMC is the broad use of unattended punctuation. It is by any means a *graphical* device, through which users can exploit a single punctuation character or the repetition of a string of characters. It is immediately visually perceivable, which has proved to be essential in social media communication – often even more than content itself.

The reason to exploit punctuation are various, for example, in (6.17), the user exploits the opening and closing quotation marks more than one time,



as it would be acceptable for common writing norms, as to underline the fact that “Buona Scuola” is not good at all, according to him.

In (6.18) the user exploits three full stops at the end of the sentence to express his uncertainty on the matter discussed.

Instead, in (6.19) are used different punctuation signs in a rather confused way. First the word *pacco* is enclosed in two apostrophes, instead of quotations, to express that the word is used in a figurative meaning (here: meaning “defection” instead of “package”). After that, the word *italiani* is shortened by the use of a triple-dot glyph, meaning continuation. The rhetorical question *La scuola?* is accompanied also by an exclamation mark, instead of just one question mark.

In (6.20), something really original happens: the users represents an opinion through a mathematical representation, namely a proportionality equation. The user places words and phrases in the equation instead of numbers.

(6.17)      ““““ buona scuola ””””

(6.18)      @SteGiannini @MiurSocial Intanto per #labuonascuola vorrei che all’Ambito Territoriale di #Roma rispondessero al telefono...

(6.19)      La buona scuola di Renzi è un ‘pacco’ per gli ita...: <http://t.co/iKWpHqDzxV..>  
La scuola! ?....ci vuole un dottore! <http://t.co/mmToHn5Mzo>

(6.20)      Proporzione della “buona scuola”: La “buona scuola” : Renzi-Giannini = la gallina innamorata : Banderas. (Noncelapossofà Zan zan)

## HASHTAGS

In Twitter hashtags are broadly exploited to convey the attention of the user on a particular topic or entity. Users create and use hashtags by placing the hash character # in front of a single word or unspaced phrase. Sometimes are precisely hashtags that encode irony or funny expressions. For example, in (6.21) the hashtag is an entire phrase “*effetto sotto zero con gelato*”, that is: below zero (degrees) with ice-cream. The reference is probably a connection between financial terms and meteorological words.

In (6.22) the user writes an hashtag containing a youth-slang expression “*bullo di provincia*”, literally meaning “provincial bully” and describing some-

one who comes from the suburbs and shows a really aggressive demeanor. The irony implies here, that the reform *Buona Scuola* would transform public schools in places suitable for bullies.

Example (6.23) shows also how users can be creative and not following linearly grammatical rules in hashtags. In the sentence “*Non ci arrenziamo*”, we firstly have to mention that “*arrenziamo*” is not even an existent word in Italian. It is phonetically similar to “*arrendiamo*” (we do not give up), but contains the name Renzi, the Italian minister who proposed the reform of education.

In the hashtag present in example (6.24) the word “school” is written with a voluntary spelling error, i.e. “*squola*” instead of the correct form “*scuola*”, probably to make fun of the institution or make fun of the reform.

- (6.21) #Renzie promette di investire su #labuonascuola, ma vuole tagliare il 3% di spesa da tutti Ministeri (anche MIUR) #EffettoSottoZeroConGelato
- (6.22) la buona scuola del #bullodiprovincia <http://t.co/eb9fCe40r0>
- (6.23) @matteoreenzi #labuonascuola?Difficile crederlo,dopo tante bugie #Quota96Scuola e la mancata soluzione#Nonciarrenziamo <http://t.co/Iu3IRyRGGo>
- (6.24) @nonleggerlo @repubblicait alla faccia de #labuonascuola,a questo punto #labuonascuola inizia a destare qualche dubbio...

## CAPS LOCK

In the corpus are present two different ways of exploiting the use of capital letters. First of all, the tweet can be all written in capital letters such as in (6.25), (6.26), and (6.27) or only some words may be highlighted through this device such as in (6.28) and (6.29). Some theories say that writing something in caps lock is to express yelling or shouting [6, 11, 12].

- (6.25) i MINISTRI DEL GOVERNO MONTI: SONO GIOVANI, MA SI FARANNO?
- (6.26) TAGLI AI POLITICI? MA DOVE? GOVERNO MONTI, SPARATI!
- (6.27) CARO MARIO MONTI, VAI DA VESPA A FIRMARE ANCHE TU'?
- (6.28) In quanto al merito: cosa ce ne facciamo delle specializzazioni e abilitazioni già acquisite fino ad oggi? CARTA STRACCIA? #labuonascuola
- (6.29) SCANDALOSO! A capodanno un marò su due era a casa in malattia

## EMOTICONS

An emoticon is a graphical representation of a facial expression using punctuation marks, numbers and letters, usually written to express user's feelings or mood. Conventionally, in Western countries, emoticons are usually written at right angle to the direction of the text.

In all three the following examples are used smiley faces, which according to [19] are the most widely exploited in Twitter. In (6.30) and (6.31) the smiley face is created through the juxtaposition of a colon (eyes) and a closed bracket (lips/mouth), while in (6.32), between eyes and mouth there is also a dash that represents the nose.

(6.30) Leggo #labuonascuola e rispondo nuovamente al questionario. Anche questo si fa, pensa un po' :) #passodopopasso

(6.31) #labuonascuola Giannini: I sindacati della scuola non sono interlocutori privilegiati".. Certo, meglio la pseudoconsultazione on line :)

(6.32) PURE SCUOLA ZOO FA POLITICA!!! :-) Oi oi Basta che cos è un comizio!!! Ke scandalo, parliamo della buona scuola,... <http://t.co/djEtyTIyX3>

### 6.1.3 Semantic devices

#### NEGATION

Quoting Karoui in [22], «tweets contain explicit negation markers such as “*ne... pas*” (not) or negative polarity items like “*jamais*” (never) or “*personne*” (nobody). Negation seems thus to be an important clue in ironic statements, at least in French»; it is interesting to see how negation is exploited in Italian ironic tweets.

Italian is a language that exploits the device of double negation (or negative concordance), in fact, the presence of more than one negation word in a sentence does not entail that the sentence has an affirmative reading (as it would be in mathematical logic).

Furthermore, characteristics of negation in Italian can be described on multiple levels: lexical, syntactic, morphological, semantic and pragmatic [26].

In the present thesis, we mainly exploit negation considering its semantic point of view, sharing Klein’s belief [24] according to whom, negation highlights «and differentiates each sentence from its correspondent with a positive polarity».

In the TWITTIRÒ corpus, the word “**non**” (not) occurs 68 times out of 393 (17,30%), as reported in the examples (6.33) and (6.34). The word “**mica**” (not at all) just 2 times out of 393 (0,51%), as it may be seen in (6.35) and in (6.36). The word “**nessun**” (nothing/no one) appears only 1 time in the whole corpus (0,25%). Another negative word such “**mai**” (never) appears 7 times (1,78%); one time, as we can see in (6.37), co-occurs with “nessun” and one time with “non”. In figure 6.2 it is shown a distribution of negative words in the corpus.

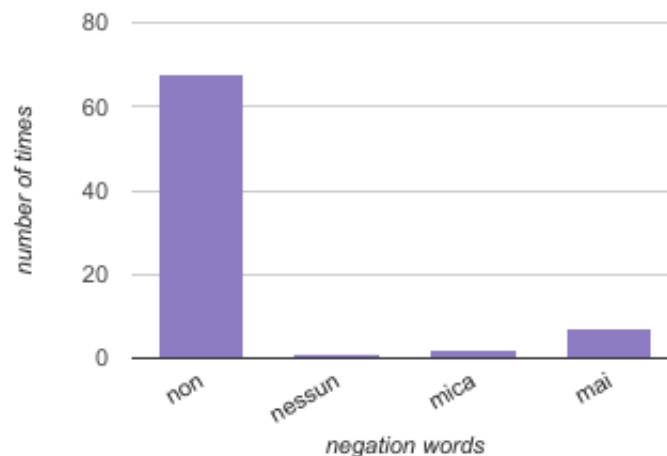


Figure 6.2: Occurrences of negation words in the TWITTIRÒ corpus

- (6.33) governo monti con gianni letta, pd perplesso. Ma che vogliono? Non esagerate sennò ci mettiamo @matteorenzi
- (6.34) E' davvero scandaloso che nel nuovo governo Monti non ci sia nemmeno una mignotta o qualche indagato per mafia! #ministri #monti #colle
- (6.35) Avvincente la nomina dei sottosegretari del governo Monti. Che tanto mica abbiamo problemi più seri...noooo, noooooo...
- (6.36) Thohir esonera Mazzarri e chiama Mancini. “Conosci mica un buon allenatore?” [@mordicchio90]

- (6.37) Simona Ventura: “Mai stata a letto con nessuno per lavorare in tv”. Allora è davvero grave. [CONTINUA su <http://t.co/oDPUtx2DvV>]

## HOMONYMY

Homonymy is the relation through which a group of words that share the same pronunciation and the same spelling (homography) have different meanings. Usually the items share the same name, independent of how close they are or how close they are related in terms of their meaning or etymology.

It is exploited in the generation of irony because it creates an unexpected turn of meaning.

In example (6.38) the word *Shanghai* is ambiguous because it refers firstly to the stock market of the Chinese city and in the second sentence “risollevarla” (lift it up) makes an implicit reference to a common board game also called *shanghai*.

In example (6.39), homonymy is exhibited through the surname of a famous Italian magistrate *Borsellino* and a common word to name a wallet: *borsellino*.

Less explicitly shown, in (6.40) there is a double layered type of irony. The first level implies that the Russian capital *Moscow* and the insect *fly* are, in Italian, both named *Mosca*. The second level of understanding needed implies that a fly usually flutters around excrements.

In both examples (6.41) and (6.42) are exploited cases of homonymy in which one of the two meanings of a word is a sexual reference or connected to something vulgar. In the first, namely *pene*, is in English, both penis and the plural form of punishment. In the second case the surname of the minister *Passera* can be also seen as referring to female sex organs (in non standard Italian).

- (6.38) Crolla la borsa di **Shanghai**. Ora bisogna risollevarla senza muovere le altre. [ @blogstark ]
- (6.39) Alemanno twitta una foto che lo ritrae con **Borsellino**. Il mio. [ @lowerome ]
- (6.40) Salvini in volo verso **Mosca**. Di solito avviene il contrario. [ a.mazed ]
- (6.41) L’arcivescovo pedofilo rischia **pene** dai 6 ai 7 anni. Ma nel dubbio spiegategli bene il senso della frase. [ @leomorabito ]

- (6.42) Indiscrezioni Governo Monti: **Passera** per lo sviluppo pare sia una richiesta particolare di Berlusconi.

## WORDPLAY

A wordplay is a rhetorical technique where words are used to become the main subject of the work, chiefly for the purpose of amusement. A largely exploited sub-category is the one of **PARONOMASIA** (pun), that is a particular situation in which in wordplays, two or more meanings are suggested, by exploiting multiple meanings of words, or of similar-sounding words (pseudo-homophones). This kind of rhetorical device is explicitly described in tweets in the sense that, usually, all the parts that compose the joke are written. At the same time, knowledge of idiomatic expressions, outer world or other additional pragmatic context are usually needed to understand this kind of multi-layered irony.

For example in (6.43) the joke meets the fact Renzi, being a minister, often wears a shirt “camicia”, and that *uova in camicia* is a particular way of cooking eggs.

Both (6.44) and (6.45) are puns. The first one exploits the assonance of *io* (I) and *Dio* (God), to show that Renzi feels like a divinity. The second one is implicit, because it states that Pd and Pdl, its just differ for “one letter”. Through the same pattern, an analogy is drawn between the lexicalized word *dighe* and a non mentioned word – *fighe* – that the reader should be able to infer to understand ironical content.

Also example (6.46) exhibits a pun, namely the expression *dieci e lode*, meaning “really good” (such as getting an *A+* in school) is converted to **loden**, which is a piece of clothing, that Monti usually wears.

- (6.43) **Uova** contro Renzi. Ora sono **in camicia**. [MisterDonnie13]
- (6.44) Renzi pubblica un selfie intitolato “**io**” ma lo rimuove subito. Mancava una **D**. [violadelpensiero]
- (6.45) Venezia, Pd sotto inchiesta per i finanziamenti alle **dighe**. Ancora una volta è **una lettera** a distinguerli dal Pdl. [blogstark]
- (6.46) @EdoardoCamurri Avrai apprezzato il titolo di Agorà di questa mattina. Governo Monti: dieci e **loden**.

### 6.1.4 Pragmatic devices

#### REGISTER CHANGE

The most frequent pragmatic device exploited in tweets is the register change, chiefly, a variation on the diafasic axis.

Generally this device leads to the creation of a context shift. It can occur by using exaggerated politeness in a situation where this is inappropriate as in example (6.47), where the writer is too polite for a normal conversation among friends (this is known as *hyperformality*, i.e. the use of extreme politeness).

Or, on the contrary, when the writer is using a very informal expression, which sounds odd in the context of a formal interaction as in (6.48), (6.49) and (6.50).

- (6.47) La buona scuola 2: **delle italiche emergenze**. <http://t.co/0XyeLk-QCDw> via @wordpressdotcom
- (6.48) Alemanno: “È in corso un tentativo di delegittimarmi”. **Ah, nevica di nuovo?** [CONTINUA su <http://t.co/oDPUtxTqU7>]
- (6.49) Napolitano saluta gli italiani. **“Buona continuazione!”** [@m\_a\_s\_s\_s]
- (6.50) Paziente mummificato ritrovato in ospedale dopo tre anni. **“Come andiamo oggi?”** [xanax]

#### ECHOIC MENTIONS

According to Sperber [40], many ironies are merely implicit echoes or ECHOIC MENTIONS of popular wisdom or received opinion but it does not describe any criteria for deciding what is a possible implicit echo and what is not.

In TWITTIRÒ, I was able to find different types of so called echoic mentions, such as **BIBLICAL QUOTATIONS** in (6.51) and (6.52), where respectively we encounter a quotation from the gospel of Mark “new wine into old bottles”, and a semi-citation the *Lord’s Prayer* (Our Father).

In examples (6.53) and (6.54) there are two direct quotations. The first one is historically attributed to Philip II of Macedon and has nowadays become a common sentence, almost proverbial; while the second is taken from a dialog scene with Don Abbondio in Manzoni’s novel *The Betrothed*.

Additionally, there is a small exploitation of popular **SONGS** and **FOOTBALL CHANTS** as in (6.55) and (6.56), where the tune of a commonly known refrain is suggested and the lyrics are changed. This happens also in (6.57), but this time the reference is to the opening theme-song of the popular cartoon *Heidi*, which was broadcasted in Italy.

It is also common to find **COMMONPLACES** that can also have become **PROVERBS** as in (6.58), where the user himself declares that it is a citation, through the word “*semicit.*”.

In (6.59) the common expression in bold, would be only “*Che Dio ce la mandi buona*”, but the user adds *tecnicamente* (technically) to refer to the technical government held by Monti, thus generating irony.

Lastly, in (6.60) the hyperbolical expression “Se non ci fossi bisognerebbe inventarti” (If you would not exist someone should invent you) is both euphemism and a false assertion.

expression

- (6.51) RENZITE e sogni. **Vino nuovo in otri vecchi** Il documento di Renzi “La buona scuola” ha creato non poche... <http://t.co/S7yUAE00s4>
- (6.52) @matteorenzi: Nostri 12 punti per #labuonascuola. Dal 15 set al 15 nov saremo in ogni scuola! **INCIELO, INTERRA, INOGNISCUOLA** @simonefurlan1
- (6.53) **Divide et impera** la buona scuola di Renzi <http://t.co/DqWHf7qwHS>
- (6.54) Riguardo #labuonascuola Don Abbondio avrebbe sicuram detto: “**Questa riforma sà da fare!**” #idoneiconcorsone2012 @FusacchiA @SteGiannini
- (6.55) #labuonascuola “Sarà bellissimo ascoltare la voce di tutti,...” @matteorenzi **la senti questa voce?....**(cit. coro da stadio )
- (6.56) @SBartezzaghi: #anagr #comedire MARIO MONTI **Ampio è lo spread, imperioso è il richiamo: / vada al governo, così RIMONTIAMO.**
- (6.57) **Merkel, Sarkozy, ti sorridono i #Mario Monti... Merkel, Sarkozy, le caprette ti fanno ciao**
- (6.58) **L’unica scuola buona è una scuola morta.** semicit. #labuonascuola
- (6.59) formato il nuovo governo Monti. **Che Dio ce la mandi tecnicamente buona.**
- (6.60) E vai adesso con Mario Monti/Superman, crisi finita, stipendi in aumento, e riforme. Grazie Stato! **Se non ci fossi bisognerebbe inventarti!**



### 6.1.5 Mixed devices

LAUGHTER

One other graphical device which is also immediately recognizable is the one of laughter. It is, most of the time, used throughout a great repetition of “laughing” characters. In CMC we find repetitions of the string “eh” or “he”, but mainly the string of “ah” and “ha”. In TWITTIRÒ there are only two occurrences (0,5%), and both of them are in all caps.

- (6.61)     **AHAHAHA** la buona scuola finanziata dai tagli su universita' e ricerca ahaha-  
hah <http://t.co/FcPfXOSIJy>
- (6.62)     @tuttoprof Solo sentir parlare di “buona scuola” mi fa ridere a crepapelle, non  
posso resistere. **AHAHAHAHAHAHAHAHAHAHAHAHAHAHA ...**

## ONOMATOPOEIC WORDS

Onomatopoeic words evoke an idea of a sound, a vivid impression of certain sensations or sensory perception, movements, colors, shapes, or actions. It is not a grammatical word class in the traditional sense of the word (like “verb” or “noun”). It is not broadly used in TWITTIRÒ, where we find only one occurrence, but is nonetheless worth mentioning.

In example (6.63) the onomatopoeia “KABOOM” mimes the explosion of a bomb. To convey the idea of an even bigger blast, the word is written in caps lock, thus exploiting both a phonetic and a graphical device.

- (6.63) Un manifestante #NoTav chiede a Cicchitto cosa farebbe il #PDL se il governo #Monti toccasse le TV di Berlusconi... **KABOOM!** #PiazzaPulita

In the following section I will introduce some examples, similar to the one just commented, in which multiple devices are exploited together to create irony.

## OVERLAPPING OF DEVICES

In several cases not only categories overlap, but also devices do. For example, in example (6.64) are exploited both the use of caps lock (bigger letters) for the word *cosmiche* (cosmic), which actually signifies something of big size.

Also, between brackets, the word *eufemismo* (euphemism), written by the user himself, as if he wanted to auto-denounce that he was being sarcastic.

In (6.65) we are able to see a lot of mixed features. A voluntary orthography error in *sQuola*, instead of *scuola* (school). Furthermore, letter *Q* is written in caps lock, inside the word, as even to accentuate it. Another time, school is defined as *good* between quotations, as if to suggest school is actually not good at all. Other random punctuation and digits is used in the text of the tweet, but it is not really explainable.

(6.64) @SteGiannini ma quale buona scuola, se avete fatto una figuraccia di dimensioni COSMICHE (**eufemismo**) al concorso di medicina?

(6.65) @SteGiannini ne #labuonascuola ci sono anche i -6 -9 dei test/i 60 del liceo, futuri medici(...)? La **sQuola** sarà sempre “**buona**”, all’italiana

## 6.2 Discussion: difficulties in applying the common frame

After having presented a broad overview of tweets and their relative comments or explanations, it is time for us to present also the hidden part of our annotation task, that is: doubts and difficulties in applying the multi-layered annotation scheme to data. Some proposals for extending the guidelines and the scheme in order to overcome such difficulties will be also presented and discussed.

As we have seen in chapter 5, a common frame was exploited to annotate the tweets of the TWITTIRÒ corpus. The common frame is a *Multi-layered annotation scheme for irony* (available in the Appendix) and it was first developed and proposed in [22], and later perfected and exploited in [23].

Thanks to this annotation scheme, which collocates this thesis in the broader context of a multilingual study on irony in social media texts, the annotators were able to apply annotation tags on tweets. In particular, as we have seen in the diagram in figure 5.1 in chapter 5 annotators had to select tags both concerning the TYPE OF CONTRADICTION and tags concerning the CATEGORY that was exploited to create irony.

As I mentioned before, supervised annotation tasks handled by humans, often see the participation of at least three people. That happens because every language presents ambiguities and it is better to rule them out through confrontation sessions among annotators. Furthermore, figurative language adds even more problems because, as discussed by many [15, 17, 18, 33, 34, 37, 40], its understanding is intrinsically connected to individual experience, individual sense of humor and situational context.

Confrontation proves to be essential to reflect upon different nuances of language that might have been perceived from one annotator but naively ignored from another. Of course, as we have seen in chapter 2 there is a statistical measure to calculate agreement among the different annotations. Once that number is good enough to proceed with the study, thus reliability of data is confirmed, it is also advisable to engage confrontation sessions to discuss data face-to-face.

In the following subsections I will discuss the main reasons of obscurity and doubts emerged while completing the annotation task.

#### **IRONIC OR NOT IRONIC?**

For example, after several meetings with the other two annotators, once the individual annotation task was complete, we decided to discard 7 tweets (1,75%) from the corpus TWITTIRÒ. None of us was completely sure about them being ironic (even if they all should have been ironic, because they had been previously retrieved from pre-existent ironic corpora) [7, 8, 39].

#### **EXPLICIT OR IMPLICIT?**

Also the distinction between implicit and explicit contradiction types may be not always clear (i.e. level 2 of annotation, according to the multi-layered scheme). The contrast, that has to be annotated is referred to the type of activation of irony.

Even if the most of cases requires some knowledge about the context to be understood, not necessarily the activation of irony depends on this knowledge.

The distinction between EXPLICIT and IMPLICIT relies on the fact that both propositions P1 and P2 are internal to the tweet, or they are not. It is a mere consideration about what is present as lexicalized element in the tweet and what is not, which is something totally different from having knowledge of the situational context or of other additional references to the outer world. Of course though they are still reliable and helpful hints to understand irony even better.

Our opinion is that must be marked as IMPLICIT only tweets where the activation of irony happens only if the reader knows the context, EXPLICIT all the other cases. The inter-annotator agreement on whether a tweet is explicitly ironic or implicitly ironic is decisively high (81,2%)<sup>2</sup>.

#### POSITIVE OR NEGATIVE?

The second struggle we had to face, deals with the level of polarity. First of all it has to be really clear what we mean when we talk about it. There are, in fact, different ways to handle this matter.

A distinction has to be done *a priori*, between SENTIMENT POLARITY of the tweet and author's STANCE, which represent two different levels of analysis, even if they both can be described in terms of **positive** or **negative** values (i.e. we can refer to positive and negative sentiment polarity of an opinion expressed in a tweet, or to positive or negative stance taken from a tweeter towards a given target). We have noticed that there is a certain degree of uncertainty and confusion between the two levels.

On the one hand, as we have seen in chapter 2, there are many exploitable lexical resources that can help an automatic tool to establish if a tweet is either positive or negative, just based on list of words or some more sophisticated linguistic knowledge (state-of-the-art technology is still unreliable on Italian).

STANCE DETECTION, is the task of determining whether the author of the text is in favor of, against, or neutral towards a given proposition or target well identified.

---

<sup>2</sup>All the calculations are from here on in the text, rely on a sample of 393 ironic tweets.

This detection can also be improved automatically as can be seen in Sohbani [36].

On the other hand we know that there are many devices that are “polarity reverser” that is: a tweet containing many positive words and just one reverser can actually have a negative stance.

Instead, we annotated **SIMPLE POLARITY** of tweets, regardless the target entity or event on which the opinions are expressed, mainly because polarity detection has not been our prime focus within the present thesis.

When we confronted among three annotators, in one of the meetings, it emerged that there was disagreement and that among annotators it was only partially understood what they had to look to “capture” whether a tweet was positive or negative. The low agreement on annotation, can be motivated by misunderstanding or bias of annotators themselves.

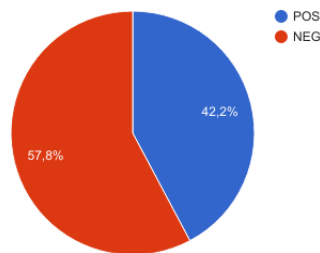


Figure 6.3: Positive and negative tweets in the corpus

Furthermore, we inquired whether a fine-grained annotation of irony, such as the one we proposed to handle could really rely on only a boolean value of positive or negative. There were cases such as in (6.66) where nothing lexical seems to point towards a negative polarity. When we examine the hidden layer that is that of irony, we infer that Minister Boschi, a good looking woman, is being diminished through the commonplace that women use sex to lure men. The tweet is thus annotated with a **negative** polarity.

(6.66) La Boschi riceve un biglietto da Casini. “Sali da me, ti mostro la mia collezione di partiti”. [CONTINUA su <http://t.co/oDPUtxkMK3>]

There has been also previous work, which inspired this thesis, such as [7, 8, 39], that expressed polarity via five different labels. What we can reflect upon is that, in the context of irony, tweets are for the most part negative, as we can see in figure 6.3, so according to the final aim of the present thesis it is not the most important level to study.

### **CATEGORY-TAGS**

Among annotators, the most discussed topic was certainly the one pertaining the annotation of level 3, i.e. categories. First of all, categories exploited in the multi-layered scheme are eight different, so the choice is harder to make (analogy, hyperbole, rhetorical question, false assertion, oxymoron/paradox, euphemism, context shift, and other).

What emerged from the confrontation sessions, is the need to mark the presence of overlapping categories that we have examined in the previous section. That is, if a tweet presents two or more different devices, that both activate irony, the annotator should be allowed to select multiple tags. In the annotation task described in this thesis, instead, for sake of simplicity, the annotator was forced to select just one category-tag from the list of the available ones. Indeed, the recommendation was to choose the category-tag, which was the most relevant for the tweet at issue, and then to add the others in the related comments.

Maybe also because of this, as it is shown in graphic in figure 6.4 the inter-annotator agreement for level 3 (categories) is not as high as in level 2 (contradiction type) or polarity, which have proven to be an easier annotation tasks.

As we can see from the chart in 6.4, concerning level 3, there is concordance only on 226 out of 393 tweets, that is a disagreement of 42,5% among annotators. A further qualitative analysis of data did not shed light on any particular repetitive pattern. It also seems that both annotators (the third one only discussed the cases of disagreement providing a way out) widely used all tags without particular biases on a favorite tag.

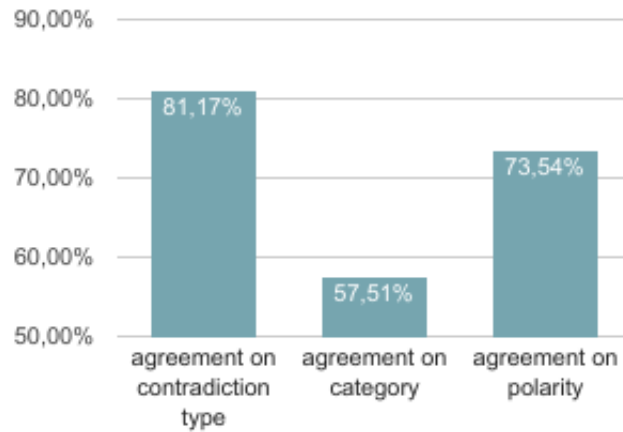


Figure 6.4: Three different levels of inter-annotator agreement

One thing worth mentioning is the relationship that exists between the category FALSE ASSERTION (only implicit) and the category OXYMORON\PARADOX (only explicit). An in-depth analysis of the disagreement between the two annotators shows that 17 times out of 393, when an annotator chose to tag false assertion, the other chose oxymoron\paradox, instead (4,33%). It is not a highly relevant percentage in TWITTIRÒ but indecision on whether of these two categories has been, according to all annotators, the moment of highest uncertainty.

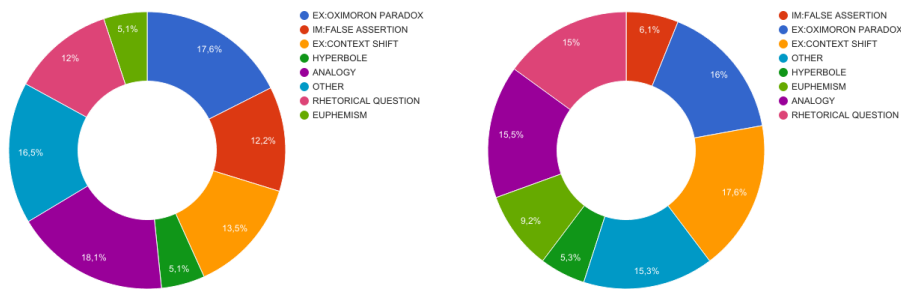


Figure 6.5: Category-tags exploited by two annotators

Furthermore, according to the multi-layered scheme for irony presented in the Appendix, those two categories just mentioned, basically cover the same type of irony. Often the decision of level 3 (categories) is triggered from the previous decision of level 2 (contradiction type).

There are 49 cases (12,47%) in which one annotator chose OTHER, because

of indecision while the other decided to tag a specific category. We believe that in this case, it should be counted as partial disagreement. For instance, in example 6.67, one annotator tagged FALSE ASSERTION while the second tagged the category as OTHER.

(6.67) Attivisti del Movimento 5 Stelle fanno irruzione nella sede del Secolo XIX. Non capivano l’emoticon. [@lowerome]

The second annotator, as written in the additional comment section, believes that XX could be interpreted as emoticon, while XIX (roman number for 19th) can not. Seeing the tweet as decoding irony through a graphical device, the second annotator was not completely convinced to choose a specific category and, instead, decided for the more generic category OTHER.

In figure 6.6, it is shown for each category of level 3 the distribution of *Annotator 1* and *Annotator 2*, respectively when one had chosen that category and the other had chosen the tag OTHER. *Annotator 1* annotated 65 times the tag OTHER and *Annotator 2* tagged a tweet with it 60 times.

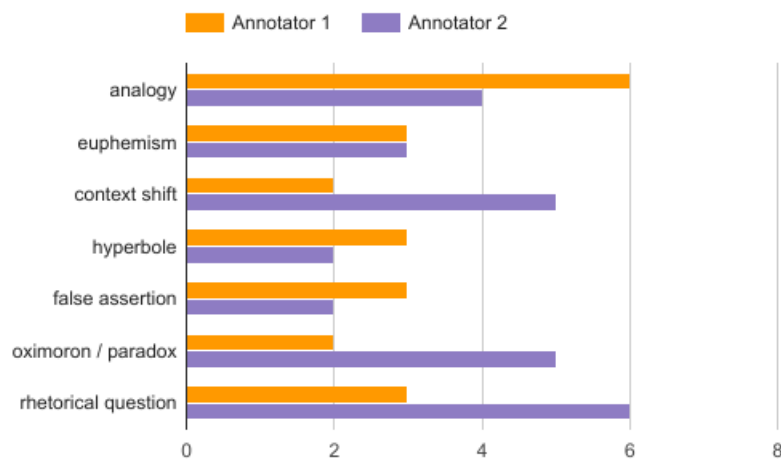


Figure 6.6: Distribution of category tags in opposition to OTHER

We believe that in future studies, disagreement calculations should give differentiated weights for the disagreement between different categories, that is:



if two annotators, for example, disagree on the tags for ANALOGY or for EUPHEMISM, that should be taken into account as a major disagreement; definitely more than a disagreement between FALSE ASSERTION and OXYMORON \ PARADOX.

Another point of discussion among annotators is that concerning REGISTER CHANGING. Namely, register changing is not a category pertaining level 3. In TWITTIRÒ, it was detected and reported in the comment section in 16 out of 393 tweets (4,07%). We observed that tweets where register changing was reported were tagged as CONTEXT SHIFT (11), ANALOGY (1), FALSE ASSERTION (1) or OTHER (3). Therefore, we decided, also according to the guidelines of the common scheme, that cases of register change are to be annotated as EX: CONTEXT SHIFT. However, in the future, we might consider to create a new category just for register change.

#### A NEW CATEGORY?

During the annotation task I have also noticed the exploitation of a common pattern, which I believe should constitute a new category on its own. I named it **FALSE LOGICAL CONCLUSION**, most of the time is an EXPLICIT CONTRADICTION, and it expresses which kind of relationship exists between a *P1* and *P2*.

According to my analysis, in 26 out of 393 (6,62%) tweets, a specific type of device is exploited, that is: statement *P1* and statement *P2* are connected through an evident **logical conclusion**, typical of philosophical reasoning, which is however, in our situation, always **false**. Irony is generated precisely because of this particular kind of logically unrelated connection.

What I have noticed is that in 15 out of 26 cases, when a false logical conclusion was present (57,7%), the category was tagged as OTHER. We can interpret this as a signal of unsatisfaction of annotators towards the available seven applicable category-tags. I believe this is the perfect example where one might reconsider the exploitability of the common scheme and try to revise it proposing enhancements.

In the present chapter we have been through an extensive and detailed analysis of tweets, and I we described the TWITTIRÒ corpus, grouping examples in different classes.

As every research can always be improved and further carried on, besides the data I have presented in this section, there are still several patterns and features that deserve further investigations.

My attention has been primarily dedicated to the most interesting patterns and schemes that I personally regarded as relevant for the present thesis and regarding the exploited corpus, such as the description of features typically used in CMC.

One first point of inquiry has been that of underlying the distinctive style and register that characterize the communication on social media, Twitter in particular. Furthermore, our data are retrieved from three different sources which present also among themselves different peculiarities. Among them: the correctness of grammatical forms, the use of abbreviations, the devious exploitation of punctuation, the presence of direct quotations, URLs, hashtags and mentions, and so on.

We have also paid attention and analyzed the situations and patterns that were most likely to occur in the TWITTIRÒ corpus, such as the majority of tweets marked as negative. This fact, as we have shown before, is supported by many linguistic theories on figurative language [17, 18, 37, 40], according to which ironic devices are most of the time used to express feelings towards unpleasant situations.

Another discussion that can be handled, linked to the point I have just mentioned, is referred to the high presence of negative words in the whole corpus. Negation is a complex device which can be seen under many points of view (lexical, syntactic, semantic and pragmatic). In our study, we merely counted the occurrences of negative words in the corpus, to have an idea of their impact on general polarity.

The same procedure, also with a deeper attention, can be further extended to a whole new set of patterns and cue words which entail intrinsic lexical

meaning and can therefore help to decode certain ironical patterns.

For example, the category of ANALOGY, is often found in concordance with words such as “**come**” (like) or expressions such as “**allo stesso modo**” (in the same way).

A broader study in this direction, can also have an important impact on automatic recognition of irony, considering that machines can well improve on tagged corpora, and having list of cue words as inputs. When irony, or a particular category that triggers it, can be described through certain peculiar words, it is also easier to design an automatic systems recognizing it.

In all other cases, where irony is generated via different kind of devices, and also an overlapping of different categories is present, the automatic recognition of irony through machines is still rather difficult to guide.

# Chapter 7

## Conclusions

The present thesis collocates in the context of a multilingual project for studying irony and for developing resources to be exploited in training NLP tools for sentiment analysis.

My work has concerned the development of a fine-grained annotation scheme for pragmatic phenomena to be exploited within the context of automatic sentiment analysis, but also to shed some light on these linguistic phenomena. In particular, it has been used to annotate the rhetorical device of irony in texts from social media. Our prime area of investigation has been, in fact, the microblogging platform *Twitter*.

From previous research the recognition of irony and the identification of pragmatic and linguistic devices that activate it are known as very challenging tasks to be performed by both humans or automatic tools. Our goal, has been to create an annotated Italian corpus of social media texts, through which we would be able to resolve some issues concerning irony formalization and automatic detection.

In chapter 2 (Sentiment Analysis and Figurative Language), I have provided an introductory overview of sentiment analysis, a sub-discipline in the field of NLP that has been gaining more and more attention in the past decades.

In this section, I have introduced the reader to the main areas in which sentiment analysis extends its domain, I have mentioned some of the most important tools and resources that are used in this field. Lastly I have focused on the main topic concerning the present research: irony detection, and I

briefly mentioned the panorama of similar studies that inspired my work.

In chapter 3 (Linguistic Theories), I have quoted and described many theories of Pragmatics, that have been and still are a great source of inspiration for computational studies on natural languages. In particular, I have focused on the theories by Austin and Searle, reporting the *Speech Acts Theory* through which we reached Grice and his studies on conversation.

The present thesis in computational linguistics is also highly in debt with the theories of Clark, Gerrig, Sperber and Wilson. The last two theoreticians, in particular, have provided the sufficient linguistic background that we have exploit to study figurative language from a computational perspective, as we have seen in chapters 5 and 6.

In chapter 4 (Twitter), has been given a description of the microblogging platform. Our main concern has been, mainly, to make the reader conscientious that the present work is an inquiry of irony in the context of Italian social media texts. Chiefly, the data studied here are Twitter messages, therefore making the social networking site our prime source for the creation of a corpus.

With chapter 5 (Methods for Data Annotation) we have entered the empirical part of the present master thesis. In this chapter I have presented the process that lead to the creation of a dedicated corpus for studying irony in social media text: TWITTIRÒ. I have described the corpus, its composition and its characteristics in relation to pre-existing corpora exploited for other studies. After that, I have introduced through many explicative examples, a detailed overview of the multi-layered annotation scheme for irony that we used to complete our task, which is also available in the Appendix in its full length.

In chapter 6 (Data Analysis), for the most part, I have discussed the tweets from the dataset and I have pointed out the existence of some recurrent features and patterns. The idea, in this chapter, has been to look for connections between linguistic theories and their exploitation in sentiment analysis. Further on, I have discussed the doubts that have come up during the annotation work and during the analysis of data.

In the Appendix we provide the original text of the exploited *Multi-layered Annotation Scheme for Irony* developed in [22] and largely exploited in [23].

## 7.1 Future work

The work on this project can be further developed in many different directions, including the following and many more.

First of all the creation of the dataset has not been an easy task. The difficulty of the present thesis lies mostly in the fact that we created the annotated corpus, TWITTIRÒ, to be used for the training of automated systems of sentiment analysis and thus, data must be clean and thoroughly checked. For this reason, numerous revisions and confrontation sessions among annotators of the annotated tweets were required.

The fact that the present master thesis collocates in a broader multilingual project offers very easy access to the shared resources, such as the common frame: *Multi-layered Annotation Scheme for Irony* (available in the Appendix). It had been clear from the beginning, that a common frame had to be followed and directions were made even clearer by a wide exploration of similar researches [7, 8, 22, 23, 39].

For my work, using a common frame resulted to have both a positive and negative aspects. It was decisively good because, from the first moment, I treated the data I had a pre-existent resource to rely on, and I did not have to create a scheme by myself. Already from the first approach to the data, I could focus on annotation, following a scheme.

At the same time, having a precise scheme to follow sometimes has lead to great confusion and has made doubts and misunderstanding emerge. A fixed scheme, forces the annotators to reflect systematically on data and on language phenomena.

Another aspect worth mentioning is that the common annotation scheme itself is quite new and it has not been exploited and tried so many times to be considered perfect and in its final shape. Possibly it will soon undergo some

modifications.

Working with different types of data is certainly the best way to prove if a common resource is valid across different domains. Our area of interest still remains the detection of irony in Italian social media text, but as we have seen in our three sub-corpora, also ironic tweets retrieved from different sources can be much different among each others.

Working on TWITTIRÒ I had the opportunity to test the *Multi-layered Annotation Scheme for Irony*, and to experience first hand that a pure theoretical vision of irony, as the ones expressed in the theories of Grice, Sperber and Wilson [17, 37, 40] are not sufficient to describe linguistic phenomena. The theory is complete only once there has been a confrontation with actual data and examples.

Of course linguistic theories on figurative language, as we referred to them in our whole work, have been fundamental to the development of the present thesis and extremely relevant in all those occasions in which it was hard to find an agreement between annotators.

Future work will entail a deeper and broader analysis of data, as the one seen in chapters 5 and 6, and a new exploitation of the common frame.

The first steps towards a new research will be conducted through the expansion of the TWITTIRÒ corpus. Our aim is to reach soon the ideal amount of 1,400 tweets. The expectation, is to reach within few months a number of annotated tweets that would make the corpus created for this research comparable to the French and English corpora mentioned in Karoui [23] and described in section 5.1.1 of the present thesis.

The annotation of the corpus has also been relevant in the wider frame of a study concerning the ways exploited in different languages to express irony. Besides French, English and Italian this study includes also Arabic. The final aim is precisely that of inquiring irony in different languages and cultures.

Furthermore, the next step leading to the completion of the annotation

task, concerns the exploitation of the dedicated software **GLOZZ**, through which a more precise annotation of CLUE WORDS is possible. In GLOZZ, every tweet has to be annotated in terms of units and, when applicable, relations between units. The annotation is performed according to a GLOZZ annotation scheme which thoroughly follows the elements presented in the sections of this thesis and in the manual available in the Appendix.

GLOZZ allows a fine-grained annotation of irony on word-level, which would be the best way to provide a clean and carefully annotated dataset for developing resources to be exploited in training NLP tools for sentiment analysis.

Through the interface of this dedicated tool, it would be easier to annotate, and later automatically detect, special cue words that have a high impact on the exploitation of irony categories. It would be easier to see, for example, if some words trigger one specific category more than another, or even if there are words that often occur in the same lexical patterns. Our research work will continue on the paths above mentioned hopefully providing new interesting discoveries to discuss.



# Appendix

In the following pages it is provided, as full-text appendix, the *Multi-layered Annotation Scheme for Irony*. The original text was developed in [22] and largely exploited in [23].

The complete TWITTIRÒ corpus will soon be available for consultation at the web page <https://github.com/IronyAndTweets/Scheme>.

## Introduction

This document presents an annotation schema for irony in tweets (Section ) and the associated annotation procedure using the Glozz tool (Section ). Besides commonly used annotation at a global level (ironic/non-ironic or ironic / positive / negative), to the best of our knowledge, no study has investigated the annotation of different types of irony as defined in linguistic theories. Our aim is to answer the following questions:

- Can the types of irony studied in linguistic state of the art be found in social media such as Twitter?
- If yes, what are the most frequent types?
- Are these types explicitly marked? What are the correlations between irony types and these markers?
- How can we exploit these correlations in a purpose of automatic detection?

To answer these questions, we propose multi-layered annotation scheme that includes four annotation levels, as shown in Figure 1. We detail in the next section each layer. All examples are extracted from the FRiC (the French Irony Corpus) or Spinoza (the italian irony corpus).

## The annotation scheme

### Level 1: Class

The first layer involves classifying tweets into ***Ironic*** and ***Not ironic***. In case of indecision (annotators do not understand the tweet because of cultural references or lack of background knowledge), a tweet can be classified into the ***No decision*** class.

### Level 2: Contradiction types

In social media such as Twitter, incongruity in ironic tweets often consists of at least two propositions (or words)  $P_1$  and  $P_2$  which are in contradiction to each other (i.e.  $P_2 = \text{Contradiction}(P_1)$ ). The presence of this contradiction activates the irony.

This contradiction can be at a semantic, veracity or intention level.  $P_1$  and  $P_2$  can be both part of the internal context of an utterance (that is explicitly lexicalized), or one is present and the other one implied. There are thus two ways to infer irony from tweets: Rely exclusively on the lexical clues internal to the utterance, or combine these clues

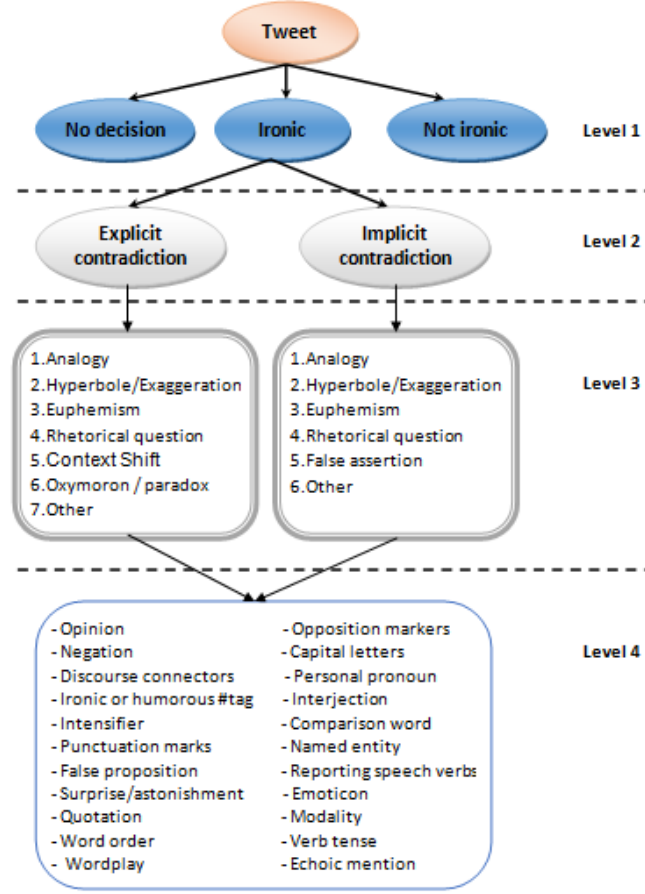


Figure 1: Annotation schema for irony.

with an additional pragmatic context external to the utterance. We define two types of contradiction: EXPLICIT and IMPLICIT.

Note that inferring irony in both types of contradiction requires some pragmatic knowledge. However, in case of IMPLICIT contradiction, the activation of irony happens only if the reader knows the context. However, in case of IMPLICIT, the activation of irony happens *only* if the reader knows the context. To identify irony activation type, the following rule can be applied: if  $P_1$  and  $P_2$  can be found in the tweet, then EXPLICIT, otherwise IMPLICIT.

In some cases, the contrast is both implicit and explicit, but, if a component of implicitness is involved in the contrast, the tweet must be annotated as implicit. We provide annotation guidelines for this case in section .

## Explicit contradiction

*Explicit contradiction* can involve a contradiction between words in  $P_1$  and words in  $P_2$  that have opposite polarities, as in (1) or, are semantically unrelated, as in (2) and (4). In particular in (4) there is a contrast between  $P_1$  that contains no opinion words, and  $P_2$  which refers to a situation which is commonly judged as being negative, but in a communicative context which is clearly unsuitable w.r.t. to the one expressed in  $P_1$  (communication between countries). Explicit opposition can also arise from an explicit positive/negative contrast between a subjective proposition  $P_1$  and a situation  $P_2$  that describes an undesirable activity or state. The irony is inferred from the assumption that the writer and the reader share common knowledge about this situation which is judged as being negative through cultural or social norms. For instance, (3) assumes that every one expects its cell phone to ring loud enough to be heard. Raining on summer holidays or growing older are other typical examples of such situations.  $P_2$  can be either *explicit negative/positive situation* containing a negative/positive sentiment (cf. (1)), or *implicit negative/positive situation* that contains no opinion words, as in (2).

- (1) [J'aime] $_{P_1}$  quand mon téléphone [**tombe en panne**] $_{P_2}$  lorsque j'en ai besoin.  
([I love] $_{P_1}$  when my phone [**fails**] $_{P_2}$  when I need it.)
- (2) [**The Voice**] $_{P_1}$  est plus important que [**Fukushima**] $_{P_2}$  ce soir.  
([**The Voice**] $_{P_1}$  is more important than [**Fukushima**] $_{P_2}$  tonight.)
- (3) [J'aime] $_{P_1}$  quand mon téléphone [**baisse le volume automatiquement**] $_{P_2}$ .  
([I love] $_{P_1}$  when my phone [**turns the volume down automatically**] $_{P_2}$ .)
- (4) L'Italia [**attende spiegazioni**] $_{P_1}$  da così tanti paesi che comincio a pensare che le nostre richieste [**finiscano nello spam**] $_{P_2}$ . [@ilmagodifloz]  
(Italy is [**waiting for explanations**] $_{P_1}$  from so many countries that I suspect our requests are being [**labeled as spam**] $_{P_2}$ .)

## Implicit contradictions

*Implicit contradiction* on the other hand, arises from a contradiction between a lexicalized proposition  $P_1$  describing an event or state and a pragmatic context external to the utterance in which  $P_1$  is false or is not likely to happen. In other words, the writer asserts or affirms  $P_1$  while he intends to convey  $P_2 = \text{Contradiction}(P_1)$  or more generally  $P_2 \neq P_1$ .

The irony occurs because the writer believes that his audience can detect the disparity between  $P_1$  and  $P_2$  on the basis of contextual knowledge or common background shared with the writer.

The proposition  $P_1$  can be either subjective (cf. (5)) or objective.  $P_1$  can occur on the left or on the right side of some subjective elements (cf. (6)) or be surrounded by objective statements. For example, in (6), the negative opinion word (in bold font) combined with the negated fact (in italic font) helps to recognize that the tweet is ironic.

In (7), the irony occurs because the situation in *italic* is absurd or not valid regarding the reality (Ayrault has really been the French Prime Minister). (8) is ironic because everybody in France expected a Hollande vs. Sarkozy duel in the 2012 presidential election. Finally, in (9) the contrast is determined by the fact that the referendum is 17 April and not May.

- (5) [#Hollande est vraiment un bon diplomate #Algérie.]<sub>P1</sub>  
 → *P*<sub>2</sub>: Hollande n'est pas un bon dipolamate selon les algériens.  
 ([#Hollande is really a good diplomat #Algeria.]<sub>P1</sub>)  
 → *P*<sub>2</sub>: Hollande isn't a good diplomat according to Algerian.
- (6) La #NSA a mis sur écoutes un pays entier. Pas d'inquiétude pour la #Belgique  
 : [*ce n'est pas un pays entier*]<sub>P1</sub>.  
 → *P*<sub>2</sub>: la Belgique est un pays.  
 (The #NSA wiretapped a whole country. No worries for #Belgium: [*it is not a whole country*].<sub>P1</sub>)  
 → *P*<sub>2</sub>:Belgium is a country.
- (7) #Ayrault a admis qu'il savait pour les écoute téléphonique de #Sarkozy. Ce-  
 pendant, [*il n'a pas précisé s'il savait qu'il était un premier ministre*]<sub>P1</sub>.  
 → *P*<sub>2</sub>: Ayrault doit savoir qu'il est premier ministre.  
 (#Ayrault admitted that he knew for #Sarkozy wiretapping. However, [*he didn't say if he knew that he was a Prime Minister*].<sub>P1</sub>)  
 → *P*<sub>2</sub>: Ayrault must know that he is Prime Minister.
- (8) Hollande et Sarkozy au second tours, [*on s'attendait pas à ça*]<sub>P1</sub>.  
 → *P*<sub>2</sub>: tout le monde s'attendait à voir Sarkozy et Hollande dans le deuxième  
 tour des élections.  
 (Holland and Sarkosy in the second tour. [*We did not expect it*]<sub>P1</sub>).  
 → *P*<sub>2</sub>: everyone expects to see Sarkozy and Hollande in the second round of  
 elections.
- (9) [Vedo che c'è molta disinformazione [*sul referendum del 17 maggio*]<sub>P1</sub>. [@Mi-  
 sterDonnie13]  
 ([I see that there a lot of misinformation [*about the referendum of the 17  
 may*]<sub>P1</sub>. [@MisterDonnie13])  
 → *P*<sub>2</sub>: the referendum is 17 April and not May

### Both implicit and explicit contradiction

In this case, the tweet has to be annotated as IRONIC WITH IMPLICIT CONTRADICTION. Since the activation of irony is also due to the presence of words or cues that are in contrast, these words have to be annotated as irony indices using the Level 4 of the annotation schema (cf. section ). Then, these words have to be linked using the *Explicit contradiction* relation (cf. section ).

(10) is a case of explicit and implicit contrast, taken from the Spinoza corpus. The irony here depends on the contrast between the words *legge* (law) and *amore* (love) and by the exploitation of the adverbs *in fondo* (after all) and *solo* (just) with the verb *dirottare* (deflect); this can make it an example of explicit irony; nevertheless also the knowledge of the context (where this female minister has caused the approval of a law favorable to the activity of her ex-husband) is necessary for the activation of irony and this makes it an example of implicit irony.

- (10) In fondo il ministro Guidi ha solo dirottato una legge per amore. [ @pirata\_21 ]  
(After all, the minister Guidi has just deflected a law for love)

### Level 3: Categories

Both explicit and implicit contradictions can be expressed in different ways which we call irony categories. Linguistic literature defines several categories. We use 8 categories: five are more likely to be found in both types of contradiction (marked *Both*) while three may occur exclusively in a specific type (marked *Expl* for explicit or *Impl* for implicit). Note that these categories are not mutually exclusive. The categories are as follows.

#### Analogy<sup>Both</sup>

We use analogy as a broader term that covers analogy, simile, and metaphor which are tools that involve a similarity between two things that have different ontological concepts, domains, or classes, on which a comparison may be based. Roughly, analogies are comparisons (cf. (11)(14)), while metaphors and similes are mechanisms for making those comparisons using figures of speech, as in (12) and (13).

- (11) Le dimanche c'est comme Benzema en équipe de France: il ne sert à rien... :D  
*Sunday is like Benzema in the French team. He is of no use... :D*
- (12) Pour une fois que je regarde la télé, c'est pour **voir Depardieu en député communiste**. #Savoureux.  
For once I watch TV, it is to see **Depardieu as a communist deputy**. #Tasty.
- (13) On n'avait qu'à écouter ses déclarations des dernières années pour savoir que **Depardieu était en fait très belge**  
Just listen to his declarations of recent years to know that **Depardieu is very Belgian**.
- (14) La Pausini farà il tour degli stadi. **Un po' come l'Inter**. [ @straccameriggi\_ ]  
Pausini will tour stadiums. **A little like Inter**.

## Hyperbole/Exaggeration<sup>Both</sup>

It is a figure of speech which consists in expressing an idea or a feeling with an exaggerated way. It is often used to make a strong impression or to emphasize a point, like in (15) and (16). Often, there is a clear distinction between exaggeration and hyperbole: look at the difference between *I feel very very thirsty* and *I'm dying of thirst*. However, we do not distinguish between them when annotating tweets. In (17) an example from the Italian Sentipolc corpus, where a literally negative feeling is expressed by using the hyperbolic expression *E' davvero scandaloso* (It is truly outrageous).

- (15) Le PS a **tellement bien** r'euissi que tt va mois bien: pollution, logement, s'ecurit'e #PARISledebat #Paris2014  
*The PS was **so successful** that all is less well: pollution, housing, security*  
*#PARISledebat #Paris2014*
- (16) @morandiniblog C'est vrai que **c'est un saint** #Berlusconi, il ne mérite vraiment pas tout cet acharnement...  
 @morandiniblog it's true that #Berlusconi **is a saint**, he does not deserve all this fury.....
- (17) E' **davvero scandaloso** che nel nuovo governo Monti non ci sia nemmeno una mignotta o qualche indagato per mafia! #ministri #monti #colle  
*It is **truly outrageous** that in the new Monti government there is not even a whore or somebody which is involved in a mafia trial! #ministri #monti #colle*

## Euphemism<sup>Both</sup>

It is a figure of speech which is used to reduce the facts of an expression or an idea considered unpleasant in order to soften the reality (like the use of *less well* instead of *worse* in (15)). Other examples are the use of *reach the stars* or *not be* to mean *a die*, or *Not seeing* to mean *blind*. (18) is an example from the French corpus. Note the use of the adverbs *peu* (little) and *pas assez* (not enough). In 19, the word *respect* is used to express a negative opinion. (20) is, instead, an example from the Spinoza's corpus. Note the use of the adverbs in *fondo* (after all) and *solo* (only).

- (18) #JO2012 **un peu** amer... au point que les gens s'entichent d'un simple caleçon de course. Surement **pas assez** "Sport" pour eux #tunisie  
 #JO2012 **a little** bitter ..... to the point that people become cray about a simple racing short. Probably **not enough** "Sport" for them #tunisie
- (19) Ca c'est un homme politique ! Savoir faire un discours sans tout le temps lire sa feuille, **respect** ! #HenriGuaino #mariagepourtous  
 This is a politician ! Knowing how to make a speech without looking his paper, **respect** ! #HenriGuaino #mariagepourtous

- (20) **In fondo** il ministro Guidi ha **solo** dirottato una legge per amore. [ @pirata\_21]  
 (**After all**, the minister Guidi has **just** deflected a law for love)

### Rhetorical question<sup>Both</sup>

It is a figure of speech in the form of a question asked in order to make a point rather than to elicit an answer, as in (21) from the French corpus and in (22) from the Italian Sentipolc corpus).

- (21) “Miss France c’est une compétition” **Non sérieux?** parce que je ne savais pas!  
*“Miss France is a competition” **No seriously?** because I didn’t know!*
- (22) Fa schifo il posto fisso. Meglio 5 anni di parlamento ed avere la pensione d’oro a vita. **Vero signor Mario?** #monti Permanent job sucks. Better five years of parliament and, then, to have a golden retirement for life. **Isn’t it, Mr. Mario?** #monti

### Context Shift<sup>Expl</sup>

It occurs by the sudden change of the topic/frame in tweet, as in (25) where the first sentence is about Duflot’s departure from the gouvernement while the second is about Lent.

Context shift can also occur by using exaggerated politeness in a situation where this is inappropriate, as (25), where the writer is too polite for a normal conversation between friends (this is known as hyper-formality, i.e., the use of extreme politeness). Or, on the contrary, when the writer is using a very informal idiomatic expression, e.g., “Chi non muore si rivede!” (So you didn’t fall off the face of the earth!) which sounds odd in the context of a formal interaction as the one described in the example (26) from the Italian Spinoza corpus.

Context shift can also happen by the use of polysemous words, like in (23) where the irony is activated by the contrast between a context where *to meet* has the meaning of to interrogate someone involved in an investigation and another when *to meet* has the meaning to spend time with a nice woman (since Boschi is known for her beauty).

- (23) I pm di Potenza incontrano Maria Elena Boschi. L’inchiesta era partita con questo obiettivo.  
 (The public prosecutors meet Maria Elena Boschi. The inquiry started with this purpose.)
- (24) You may grant me the honor of listening to another one of your finepredictions.
- (25) Duflot quitterait le gouvernement. En plein carême, on ne peut même pas le fêter. Décidément, elle embête jusqu’au bout... \*soupon\*



Duflot left the gouvernement. In the middle of Lent, we can not even celebrate it. Really, she bothers until the end ... \*sigh\*

- (26) Gentiloni accoglie i due ostaggi italiani sopravvissuti: **“Chi non muore si rivede!”** [@montales1]  
Gentiloni welcomes the two surviving Italian hostages: **“So you didn’t fall off the face of the earth!”**

### False assertion<sup>Impl</sup>

It indicates that a proposition, fact or an assertion fails to make sense against the reality. The speaker expresses the opposite of what he thinks or something wrong with respect to a context. External knowledge is needed to understand the irony. For example, the tweets in the examples (27) to (29) are ironic because the situations in bold font are absurd or not possible in reality. Note that the tweet in (29) is also an example of the RHETORICAL QUESTION category. Also in the example 30 the situation in bold is absurd (Divino Otelma is an Italian magician, a television star with no competence on Economy).

- (27) The #NSA wiretapped a whole country. No worries for #Belgium: **it is not a whole country.**
- (28) @Vince75015 Les agences de notation **ne font pas de la politique**  
The notation agencies **do not make politics any more.**
- (29) @infos140 @mediapart Serge Dassault ? Corruption ? Non ! Il doit y avoir une erreur. **C’est l’image même de la probité en politique.**  
@infos140 @mediapart Serge Dassault ? Corruption ? No ! There is an error. **He is the perfect image of probity in politics.**
- (30) #Governo #Monti: il divino Otelma #ministro dell’ #Economia  
#Government #Monti: **Divino Otelma #Economy minister**

### Oxymoron/paradox<sup>Expl</sup>

This category is equivalent to the category “False assertion” except that the contradiction is explicit, like the use of two antonyms in the first sentence in (15) (*so successful* vs. *less well*), and the use of two opposite events in (31) (assuming that hitting is not a violent action).

In (32), there is a paradox because whistlers were not in Paris to celebrate November 11th but to bother the president.

- (31) Ben non ! Matraquer et crever des yeux, ce n'est pas violent et ça respecte les droits !!! #assnat #polqc #ggi  
No ! **Clubbing and putting up eyes**, is **not violent** and **it does obey the laws**. !!! #assnat #polqc #ggi
- (32) Il est evident que chaque siffleur etait la pour le 11 novembre, et non pour siffler la politique de François Hollande.  
It is obvious that **every whistler was here for November 11th** and **not to whistle François Hollande's politics**.
- (33) Sulla morte di Giulio Regeni l'Italia chiede all'Egitto **indagini congiunte**. Metti che alla fine **i depistaggi non coincidano**. On the death of Giulio Regeni Italy asks Egypt joint investigations. Just in case in the end the false tracks will not converge. [@\_edelman]

### Other<sup>Both</sup>

It represents ironic tweets with explicit or implicit contradiction which can not be classified under one of the seven categories above. This can occur for example in case of humor or situational irony. Here are some examples:

- (34) Alerte à la pollution de l'air : il est déconseillé de prendre son vélo pour aller au travail à 9h... mais pas sa voiture diesel !  
Polution alert: it is not recommended to take your bike to go work at 9am...but not your diesel car !
- (35) Merci Hollande d avoir sauvé le monde ! Sans toi, la terre serait actuellement entrée en 3ème guerre mondiale  
Thanks Hollande for saving the world! Without you, the earth would now into the third world war.
- (36) Sallusti e Santanchè sono per le unioni naturali. Anche se non si direbbe.  
[CONTINUA su <https://t.co/oDPUtx2DvV>]  
Sallusti and Santanchè are in favor of natural partnerships. Even if you would not say it.

### Level 4: Clues

Clues represent words that can help annotators to decide in which category belongs a given ironic tweet, such as *like* for Analogy, *very* for Hyperbole/Exaggeration. Clues include also negation word, emoticon, punctuation marks, interjection, opinion, named entity, etc, as depicted in Figure 1. Examples 37 and 38 show the clues annotation (named entities are highlighted in blue, opinion expressions in yellow, intensifier in olive

color, comparison word in green, negation in red, punctuation in magenta and emoticons in pink).

(37) #Hollande est vraiment un bon diplomate #Algérie.  
 ( #Hollande is really a good diplomat #Algeria.)

(38) Le dimanche c'est comme Benzema en équipe de France: Il ne sert à rien ... :D  
 ( Sunday is like Benzema in the French team. He is of no use ... :D )

Annotators are asked to annotate clues even if the tweet is considered as non ironic or classified into the no decision class. This is very important to analyze the correlation between clues and irony classes/contradiction types/irony categories at respectively the first, second and third level of our schema.

## The annotation procedure

### Overview of the Glozz tool

Each tweet has to be annotated in the Glozz tool, in terms of units and, when applicable, relations between units. The annotation is performed according to a Glozz annotation scheme which thoroughly follows the elements presented in the previous sections of this manual.

### Main interface

The main interface of Glozz comes with 6 main frames, as shown in Figure 2:

- 1: main view, where we can see the annotated text, and directly add or edit annotations.
- 2: macro view, it's a view on the same annotated text as the main view does text but in macro mode, enabling to have a global view on the annotated text, and to navigate quickly through it.
- 3: mode buttons, in order to set the current mode (adding units, editing units, and so on).
- 4: annotation model, where we can see the list of all available types (one column for units, one for relations, and one for schemas).
- 5: feature sets table, which shows the features values of the selected element.
- 6: annotation as text table, where each element is shown in a logical predicat

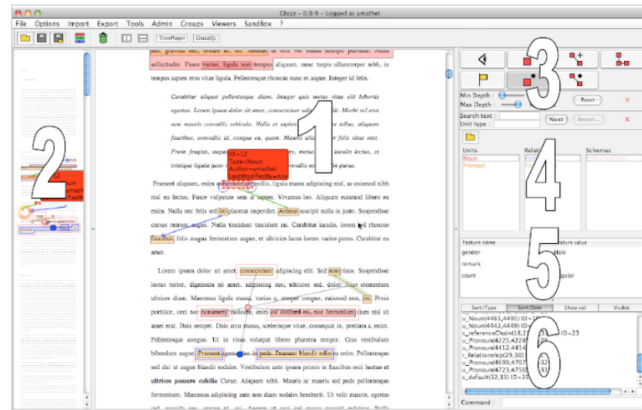


Figure 2: The Glozz Interface.

### Before annotating: corpus and scheme loading

Before annotating a tweet, you need to open a corpus. This means that you have to upload an .ac file and an .aa file via the File menu or the shortcut button (open a corpus), as shown in Figure 3.

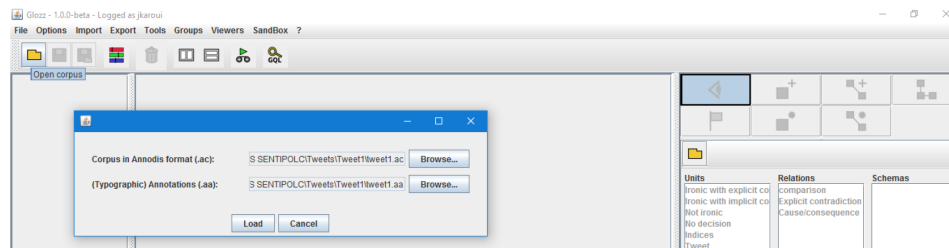


Figure 3: The Glozz Interface.

Once a corpus is opened, before creating or editing annotations, you may load an annotation model (an .aam file), as shown in Figure 4.

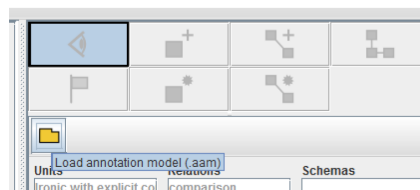


Figure 4: Loading an annotation model.

Since no style file is loaded, all the style names appear with the same color. So you may load the style file (an .as file), as shown in Figure 6.

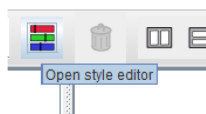


Figura 5: Opening a style file.

### Annotating how-to: adding, editing and saving annotations

A toolbar is provided in frame 3 as follows, which enables to choose the current mode of annotation.

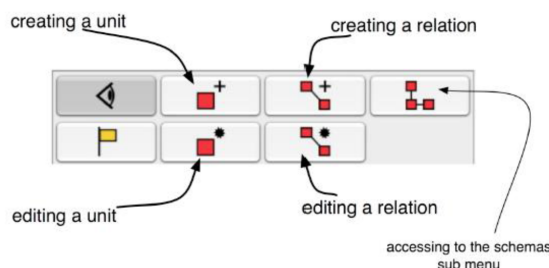


Figura 6: Annotation toolbar.

To create a unit, first, click on the unit creation mode button. This results in highlighting this button (showing that this mode is now active), but also in activating the Units part of the annotation model selector in the left. Then, select in the Units model the type you want to give to the next created unit(s).

To create a relation, first, click on the « create a new relation » button mode, which activates this mode and also the related part of the annotation model selector. Then, the procedure is very simple:

- Reminder: a relation is always a link between two annotation elements (and cannot point on a part of text not having annotations), these annotations being units, relations or schemas.
- put the mouse over the start element, then click.
- put the mouse over the end element (it becomes red), then click.

When the annotation is done, do not forget to save the updated annotation files.

## Annotating a tweet in Glozz following the annotation scheme

### Annotating a unit

Each unit has a set of attributes as shown in Figure 7. The easiest way is to start by annotating textual elements and then annotate the whole segments. The following table summarizes the various units that you will handle, and the associated attributes.

For each tweet, a Glozz pre-annotation is made available beforehand to the annotators. Such a pre-annotation corresponds to the Tweet unit and contains:

- Id: tweet identifier as given by the Tweeter API,
- Internal Id: the tweet internal identifier
- Date and Topic: the date of the tweet and the topic (if these information are not available, the default values are *null*).
- Global opinion: the tweet global opinion among positive, negative, objective and subjective, as given by SentiTuT annotations.
- Irony (yes/no): Indicates if the tweet is ironic or not according to the SentiTuT annotations. This is a pre-annotation and you MUST in addition specify the type of the irony. Hence:
  - If the tweet is not ironic, you need to annotate the tweet with the Non ironic unit.
  - If the tweet is ironic, you need to decide if it is an irony with explicit contradiction or an irony with an implicit contradiction.

### Annotating a relation

Relations are used to connect units appearing in a tweet. We distinguish three types of relationship:

- Comparison relation: Connects two units that are compared (subjects of comparison) (Figure 8).
- Explicit contradiction relation: Connect two units that are opposed (Figure 9).
- Cause/consequence relation: Connect two units such as the first unit is the cause and the second one its consequence (Figure 10).

### Annotation constraints

During the annotation, make sure to follow the constraints below:

Units	Attributes
<b>Irony with explicit contradiction</b>	Analogy = {No, Yes}
	Hyperbole/Exaggeration = {No, Yes}
	Euphemism = {No, Yes}
	Rhetorical question = {No, Yes}
	Register changing = {No, Yes}
	Oxymoron/paradox = {No, Yes}
	Other = {No, Yes}
<b>Irony with implicit contradiction</b>	Analogy = {No, Yes}
	Hyperbole/Exaggeration = {No, Yes}
	Euphemism = {No, Yes}
	Rhetorical question = {No, Yes}
	False assertion = {No, Yes}
	Other = {No, Yes}
<b>Not ironic</b>	No attributes
<b>No decision</b>	In case of indecision (annotators do not understand the tweet because of cultural references or lack of background knowledge)
<b>Clues</b>	Discourse connectors = {No, Yes}
	Punctuation = {No, Yes}
	Opinion = {+, -, neutral, No}
	Emoticon = {+, -, neutral, No}
	Opposition markers = {No, Yes}
	Capital letters = {No, Yes}
	Intensifier = {No, Yes}
	Comparison word = {No, Yes}
	Modality = {No, Yes}
	Negation = {No, Yes}
	Quotation = {No, Yes}
	Interjection = {No, Yes}
	Personal pronoun = {No, Yes}
	Reporting speech verb = {No, Yes}
	Surprise/astonishment = {No, Yes}
	Named entity = {Person, Object, Place, Animal, Event/Fact, Function, Organization, Age, Duration, Date, Period, Day, Weight, Temperature, Percentage, Length, No}
	False proposition = {No, Yes}
	Irony or humorous #tag = {No, Yes}
	Word order = {No, Yes}
	Wordplay = {No, Yes}
	Verb tense = {No, Yes}
	Echoic mention (proverbs, songs, poem)={No, Yes}
<b>Tweet</b>	Tweet Id → already defined
	Internal Id of the tweet → already defined
	Global opinion of the tweet = {+, -, neutral, No}

Figura 7: Units and their attributes in the annotation scheme.

1. Whatever the type of irony (among irony with explicit contradiction and irony

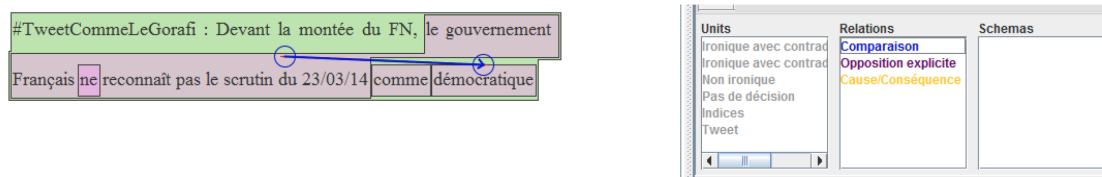


Figure 8: A comparison relation between two units. Here the cue “comme” (*like*) triggers a comparison relation.

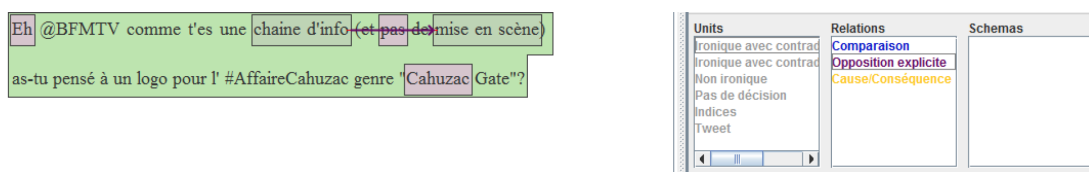


Figure 9: An explicit opposition relation between two units. Here the cue “pas” (*not*) triggers a contradiction relation.

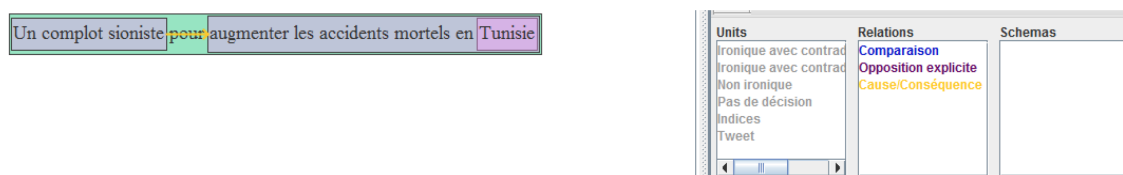


Figure 10: A cause/consequence relation between two units. Here the cue “pour” (*in order to*) triggers a causal relation.

with implicit contradiction), you must annotate all clues.

2. For ironic tweets with explicit or implicit contradiction, you must classify it in one of the categories belonging to each type of irony. The default value is *Undefined*. Hence, at least one attribute should have the value *yes*.
3. For ironic tweets with explicit contradiction, you must connect some parts of the tweet with the “Explicit contradiction” relation.
4. In some case of Ironic tweets with implicit contradiction, you can use “Explicit contradiction” relation to connect some parts of the tweets in contradiction.
5. If the tweet contains a comparison, you must connect the units that are compared with the “Comparison” relation.
6. DO NOT delete the tweet pre-annotation, i.e., the tweet unit.



# Bibliography

- [1] ATTARDO, S. *Linguistic theories of humor*, vol. 1. Walter de Gruyter, 1994.
- [2] AUSTIN, J. L. *How to do things with words*. Oxford university press, 1975.
- [3] BACCIANELLA, S., ESULI, A., AND SEBASTIANI, F. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *LREC* (2010), vol. 10, pp. 2200–2204.
- [4] BARBIERI, F., BASILE, V., CROCE, D., NISSIM, M., NOVELLI, N., AND PATTI, V. Overview of the evalita 2016 sentiment polarity classification task. In *Proceedings of Third Italian Conference on Computational Linguistics (CLiC-it 2016) & Fifth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2016)* (2016).
- [5] BASILE, V., BOLIOLI, A., NISSIM, M., PATTI, V., AND ROSSO, P. Overview of the evalita 2014 sentiment polarity classification task. In *Proceedings of the 4th evaluation campaign of Natural Language Processing and Speech tools for Italian (EVALITA '14)* (2014).
- [6] BAZZANELLA, C. Oscillazioni di informalità e formalità: scritto, parlato e rete. *Formale e informale. La variazione di registro nella comunicazione elettronica. Roma: Carocci* (2011), 68–83.

- [7] BOSCO, C., LAI, M., PATTI, V., AND VIRONE, D. Tweeting and being ironic in the debate about a political reform: the french annotated corpus TWitter-MariagePourTous.
- [8] BOSCO, C., PATTI, V., AND BOLIOLI, A. Developing corpora for sentiment analysis: The case of irony and senti-tut. *IEEE Intelligent Systems* 28, 2 (2013), 55–63.
- [9] BOSCO, C., TAMBURINI, F., BOLIOLI, A., AND MAZZEI, A. Overview of the evalita 2016 part of speech on twitter for italian task. In *Proceedings of Third Italian Conference on Computational Linguistics (CLiC-it 2016) & Fifth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2016). Associazione Italiana di Linguistica Computazionale (AILC)* (2016).
- [10] CERRUTI, M., AND ONESTI, C. Netspeak: a language variety? some remarks from an italian sociolinguistic perspective. *Languages go web: Standard and non-standard languages on the Internet* (2013), 23–39.
- [11] CHIUSAROLI, F. Sintassi e semantica dell’hashtag: studio preliminare di una forma di scritture brevi. In *The First Italian Conference on Computational Linguistics, CLiC-it 2014–Proceedings* (2014), vol. 9.
- [12] CHIUSAROLI, F., AND ZANZOTTO, F. M. Scritture brevi di oggi, 2012.
- [13] CLARK, H. H., AND GERRIG, R. J. On the pretense theory of irony.
- [14] FATIGANTE, M., MARIOTTINI, L., SCIUBBA, M. E., ET AL. *Lingua e società. Scritti in onore di Franca Orletti: Scritti in onore di Franca Orletti*. FrancoAngeli, 2009.
- [15] FINK, C. R., CHOU, D. S., KOPECKY, J. J., AND LLORENS, A. J. Coarse- and fine-grained sentiment analysis of social media text. *Johns Hopkins APL Technical Digest* 30, 1 (2011), 22–30.

- [16] GHOSH, A., LI, G., VEALE, T., ROSSO, P., SHUTOVA, E., BARNDEN, J., AND REYES, A. Semeval-2015 task 11: Sentiment analysis of figurative language in twitter. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)* (2015), pp. 470–478.
- [17] GRICE, H. P. Logic and conversation. *1975* (1975), 41–58.
- [18] GRICE, H. P. Further notes on logic and conversation. *1978 1* (1978), 13–128.
- [19] GROOT, R. D. Data mining for tweet sentiment classification.
- [20] HU, M., AND LIU, B. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (2004), ACM, pp. 168–177.
- [21] JORGENSEN, J., MILLER, G. A., AND SPERBER, D. Test of the mention theory of irony. *Journal of Experimental Psychology: General* *113*, 1 (1984), 112.
- [22] KAROUI, J., BENAMARA, F., MORICEAU, V., AUSSÉNAC-GILLES, N., AND BELGUITH, L. H. Towards a contextual pragmatic model to detect irony in tweets. In *53rd Annual Meeting of the Association for Computational Linguistics (ACL 2015)* (2015), pp. PP–644.
- [23] KAROUI, J., BENAMARA, F., MORICEAU, V., PATTI, V., AND CRISTINA, B. Exploring the impact of pragmatic phenomena on irony detection in tweets: A multilingual corpus study. In *Proceedings of the European Chapter of the Association for Computational Linguistics – 3/7 April 2017, Valencia (Spain)* (2017), p. in press.
- [24] KLEIN, W. A simple analysis of sentential negation in german. *Manuscript, Max-Planck-Institute for Psycholinguistics, Nijmegen* (2007).
- [25] KOULOUMPIS, E., WILSON, T., AND MOORE, J. D. Twitter sentiment analysis: The good the bad and the omg! *Icwsm 11* (2011), 538–541.

- [26] MANZOTTI, E., AND RIGAMONTI, A. La negazione. *Grande grammatica italiana di consultazione 2* (1991), 245–317.
- [27] MAYNARD, D., AND GREENWOOD, M. A. Who cares about sarcastic tweets? investigating the impact of sarcasm on sentiment analysis. In *LREC* (2014), pp. 4238–4243.
- [28] MIHALCEA, R., AND PULMAN, S. Characterizing humour: An exploration of features in humorous texts. In *International Conference on Intelligent Text Processing and Computational Linguistics* (2007), Springer, pp. 337–347.
- [29] MIHALCEA, R., AND STRAPPARAVA, C. Making computers laugh: Investigations in automatic humor recognition. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing* (2005), Association for Computational Linguistics, pp. 531–538.
- [30] PAK, A., AND PAROUBEK, P. Twitter as a corpus for sentiment analysis and opinion mining. In *LREc* (2010), vol. 10, pp. 1320–1326.
- [31] PANG, B., AND LEE, L. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval 2*, 1-2 (2008), 1–135.
- [32] PANG, B., LEE, L., AND VAITHYANATHAN, S. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10* (2002), Association for Computational Linguistics, pp. 79–86.
- [33] REYES, A., ROSSO, P., AND BUSCALDI, D. Finding humour in the blogosphere: the role of wordnet resources. In *Proceedings of the 5th Global WordNet Conference* (2010), pp. 56–61.
- [34] REYES, A., ROSSO, P., AND BUSCALDI, D. From humor recognition to irony detection: The figurative language of social media. *Data & Knowledge Engineering 74* (2012), 1–12.

- [35] SEARLE, J. R. *Speech acts: An essay in the philosophy of language*, vol. 626. Cambridge university press, 1969.
- [36] SOBHANI, P., MOHAMMAD, S. M., AND KIRITCHENKO, S. Detecting stance in tweets and analyzing its interaction with sentiment. *The SEM 2016 Organizing Committee. All papers 2016 their respective authors. This proceedings volume and all papers therein are licensed under a Creative Commons Attribution 4.0 International License.* (2016), 159.
- [37] SPERBER, D., AND WILSON, D. Irony and the use-mention distinction. *Philosophy* 3 (1981), 143–184.
- [38] SPERBER, D., AND WILSON, D. *Relevance: Communication and cognition*, vol. 142. Citeseer, 1986.
- [39] STRANISCI, M., BOSCO, C., PATTI, V., AND FARIAS, D. I. H. Analyzing and annotating for sentiment analysis the socio-political debate on #labuonascuola. *CLiC it* (2015), 274.
- [40] WILSON, D., AND SPERBER, D. On verbal irony. *Irony in language and thought* (2007), 35–56.