

A Multi-layered Annotation Schema for Irony in Tweets

September 28, 2016

1 Introduction

This document presents an annotation schema for irony in tweets (Section 2) and the associated annotation procedure using the Glozz tool (Section 3). 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 Sponiza (the italian irony corpus). Category markers are in bold font.

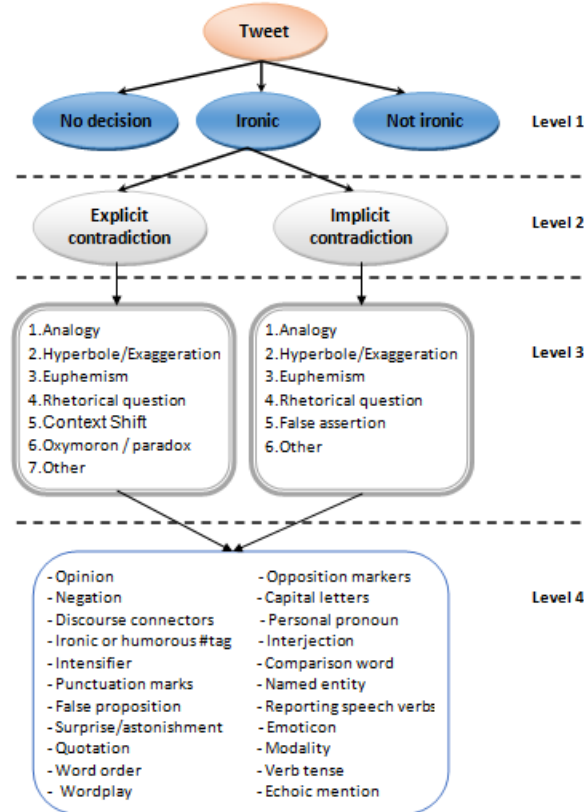


Figure 1: Our annotation schema.

2 The annotation scheme

2.1 Level 1: Class

The first layer involves classifying tweets into ***Ironi***c and ***Not ironi***c. 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.

2.2 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 opposed to each other. It is the presence of this opposition that activates the irony.

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 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, which is not the case for EXPLICIT contradictions.

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 2.2.3.

2.2.1 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) I **love** when my phone **fails** when I need it.
- (2) **The Voice** is more important than **Fukushima** tonight.
- (3) I **love** when my phone **turns the volume down automatically**.
- (4) L'Italia **attende spiegazioni** da così tanti paesi che comincio a pensare che le nostre richieste **finiscano nello spam**. [@ilmagodifloz]

(Italy is **waiting for explanations** from so many countries that I suspect our requests are being **labeled as spam**.)

2.2.2 Implicit contradictions

Implicit contradiction on the other hand, arises from an opposition 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'_1 such that $P'_1 = \text{Not}(P_1)$ or more generally $P'_1 \neq P_1$. The irony occurs because the writer believes that his audience can detect the disparity between P_1 and P'_1 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. In this latter case, 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 is really a good diplomat #Algeria.
- (6) The #NSA wiretapped a whole country. **No worries** for #Belgium: *it is not a whole country*.
- (7) #Ayrault admitted that he knew for #Sarkozy wiretapping. However, *he didn't say if he knew that he was a Prime Minister*.
- (8) Holland and Sarkosy in the second tour. *We did not expect it*.
- (9) Vedo che c'è molta disinformazione sul referendum del 17 maggio. [@MisterDonnie13]
(I see that there a lot of misinformation about the referendum of the 17 may)

2.2.3 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 2.4). Then, these words have to be linked using the *Explicit contradiction* relation (cf. section 3.2.2).

(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)

2.3 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.

2.3.1 Analogy^{Both}

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 litte like Inter**.

2.3.2 Hyperbole/Exaggeration^{Both}

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'eussi que tt va mois bien: pollution, logement, s'ecurit'e #PARISledebat #Paris2014
The PS was so successful that all is less well: polution, 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

2.3.3 Euphemism^{Both}

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)

2.3.4 Rhetorical question^{Both}

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 Sentipol 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

2.3.5 Context Shift^{Expl}

It occurs by the sudden change of the topic/frame in tweet, as in (25) where the first sentence is about Duflo's departure from the government 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 hyperformality, 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... *souponire*
DufLOT left the government. 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!**”

2.3.6 False assertion^{Impl}

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**

2.3.7 Oxymoron/paradox^{Expl}

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]

2.3.8 Other^{Both}

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 recommanded 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.

2.4 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 word, emoticon, punctuation marks, interjection, opinion, etc, as depicted in Figure 1. 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.

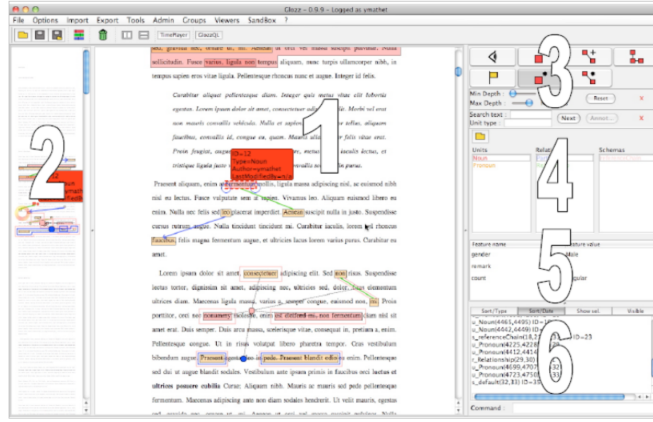


Figure 2: The Glozz Interface.

3 The annotation procedure

3.1 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.

3.1.1 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

3.1.2 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.

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

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 10.

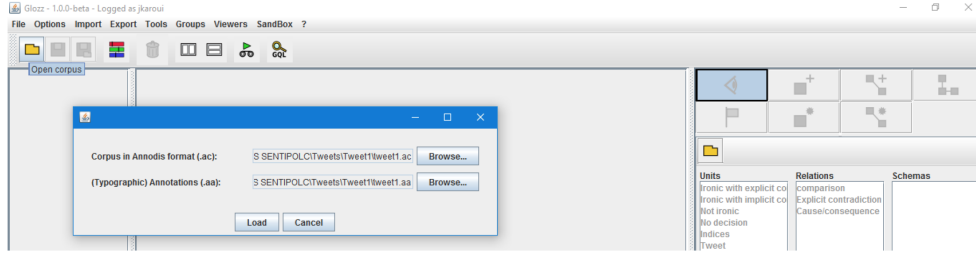


Figure 3: The Glozz Interface.

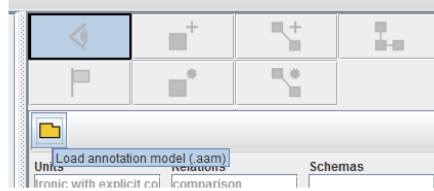


Figure 4: Loading an annotation model.

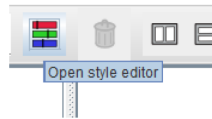


Figure 5: Opening a style file.

3.1.3 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.

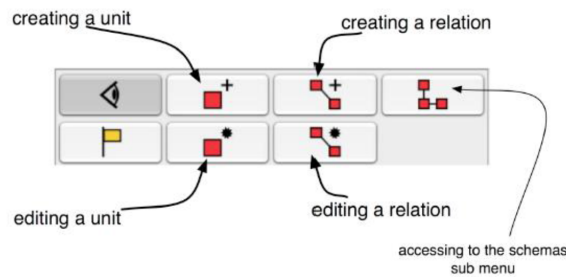


Figure 6: Annotation toolbar.

To create a unit, first, click on the unit creation mode button. This result 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.

3.2 Annotating a tweet in Glozz following the annotation scheme

3.2.1 Annotating a unit

Each unit has a set of attributes. 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.

3.2.2 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).
- Explicit contradiction relation: Connect two units that are opposed.
- Cause/consequence relation: Connect two units such as the first unit is the cause and the second one its consequence.

Units	Attributes
Irony with explicit contradiction	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}
Irony with implicit contradiction	Analogy = {No, Yes}
	Hyperbole/Exaggeration = {No, Yes}
	Euphemism = {No, Yes}
	Rhetorical question = {No, Yes}
	False assertion = {No, Yes}
	Other = {No, Yes}
Not ironic	No attributes
No decision	In case of indecision (annotators do not understand the tweet because of cultural references or lack of background knowledge)
Clues	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}
Tweet	Tweet Id → already defined
	Internal Id of the tweet → already defined
	Global opinion of the tweet = {+, -, neutral, No}

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

3.2.3 Annotation constraints

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

1. Whatever the type of irony (among irony with explicit contradiction and irony with implicit contradiction), you must annotate all clues.

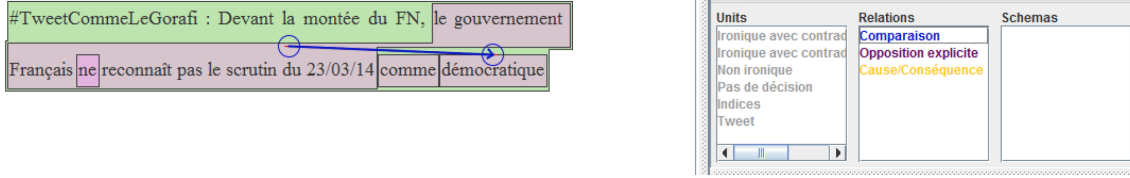


Figure 8: A comparison relation between two units. Here the cue “comme” (*like*) triggers a comparison relation.

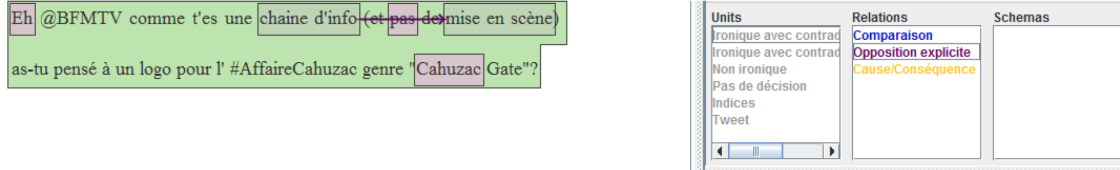


Figure 9: An explicit opposition relation between two units. Here the cue “pas” (*not*) triggers a contradiction relation.

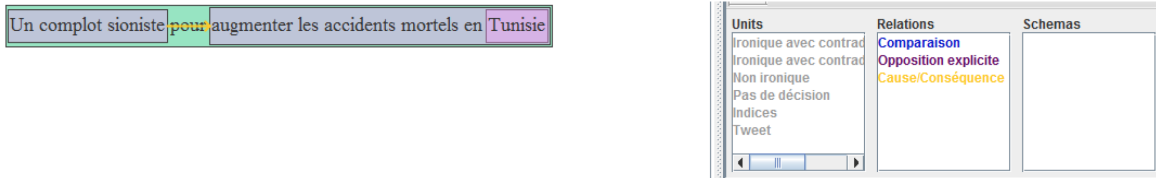


Figure 10: A cause/consequence relation between two units. Here the cue “pour” (*in order to*) triggers a causal relation.

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.