

ANNOTATION OF CYBERBULLYING MESSAGES

1. DEFINITION	1
2. OBJECTIVE	1
3. WHAT TO ANNOTATE	2
4. ANNOTATION STEPS	2
5. TEXT SPAN OF HARMFUL MESSAGES	2
6. CYBERBULLYING ROLES	3
7. CYBERBULLYING TYPES	3
8. SARCASM	6
9. NON-OFFENSIVE	6
10. COMMENT	6
11. HOW TO USE CAT	6

These guidelines are an adaptation of the “Guidelines for the Fine-Grained Analysis of Cyberbullying” developed by the “Language and Translation Technology Team” of Ghent University (<https://repository.uantwerpen.be/docman/irua/623d87/e1409538.pdf>).

1. DEFINITION

Cyberbullying definition (Smith et al., 2008):

“Cyberbullying is an aggressive, intentional act carried out by a group or individual, using electronic forms of contact, repeatedly and over time against a victim who cannot easily defend him or herself.”

2. OBJECTIVE

The objective of the annotation is to indicate which messages in the conversations are (potentially) harmful and thus are examples of cyberbullying. For the annotation, all messages presented within the conversation should be considered thus it is important to read the whole conversation to understand the context.

3. WHAT TO ANNOTATE

The annotator should identify all the harmful messages in a conversation.

For each harmful message, the annotator should annotate:

1. the cyberbullying role of the message's author
2. the cyberbullying type of the message
3. the presence of sarcasm in a message
4. whether the message containing insults is not really offensive but a joke.

Please note that:

- Each file corresponds to a single conversation.
- Each line of a file corresponds to a comment posted by an author who has a specific role.
- Each comment can contain more than one harmful message, each having a different cyberbullying type.
- Only the messages written in Italian should be taken into accounts.
Languages other than Italian are not to be considered in this annotation.

4. ANNOTATION STEPS

- Read the whole conversation
- Identify (potentially) harmful messages within the posted comments
- Annotate the text span corresponding to a (potentially) harmful message
- Annotate the cyberbullying role
- Annotate the cyberbullying type
- Mark the presence of sarcasm (of leave it blank)
- Mark the presence of not offensive messages containing insults (of leave it blank)
- Add a comment in case of doubts related to a single annotated message

5. TEXT SPAN OF HARMFUL MESSAGES

A text span is the part of the text that relates to the action under investigation, in our case cyberbullying.

Usually a text span is a clause containing a noun and a verb (e.g. “*sei un idiota*”) but it can also be a single word (e.g. a curse word or insult like “*ciccione*” or an hashtag like “#ammazzati”) or an entire sentence if it expresses as a whole an offense (e.g. a threat like “*appena ti vedo ti spacco*”).

As a convention, the subject and verb of a sentence part to be categorized (when available)

should be included in the annotated text span. Punctuation marks at the end of the text span should be included in the annotated text span. Emoji are to be included in

the text span if they help in understanding the meaning of the message: this is particularly important in the case of sarcasm. Indeed, in many cases smiles and hearts are added to insults to make them non-offensive.

Please note that a comment can contain messages to be annotated with different cyberbullying types thus different text span should be annotated, one for each cyberbullying type.

Example:

- *Che fallito, vatti a chiudere nella fogna, finocchio!*
[che fallito,] Insult: General Insult [vatti a chiudere nella fogna,] Curse or Exclusion
[finocchio!] Insult-Discrimination: Sexism

Annotation cannot span more than one sentence and less than one word. This means that, if a comment consists of more than one sentence, different messages should be annotated separately taking into consideration the sentence boundary (a period, a semi-colon, a colon, a question mark or an exclamation mark):

- *Ma perché mi trattate così? Sono una brava persona!*
[Ma perché mi trattate così?] Defense [Sono una brava persona!] Defense

6. CYBERBULLYING ROLES

We distinguish four cyberbullying roles:

1. Harasser: Person who initiates the harassment.
2. Victim: Person who is harassed. Sometimes the victim responds to the harasser with insults and threats; however, she/he must still be annotated as a victim.
3. Bystander-defender: Person who helps the victim and discourages the harasser from continuing his actions.
4. Bystander-assistant: Person who does not initiate, but takes part in the actions of the harasser (e.g. by encouraging the harasser).

Please note that no role is to be assigned to non-offensive messages.

Examples:

Daniele: *Ma vai a fare calcio [femminuccia che non sei altro]* [Harasser]

Enrico: *Ma se non mi avete mai visto ballare come potete giudicare?* [Victim]

Alice: *[Lasciatelo stare]* [Bystander-defender]

Giulio: *[Grande Daniele!] 😊* [Bystander-assistant]

7. CYBERBULLYING TYPES

A type should be assigned to all annotated messages, including sarcastic and non-offensive messages.

Threat or blackmail

This category contains expressions of physical or psychological threats towards the addressee and expressions indicating blackmail.

Examples:

- *[Ti tiro un pugno in faccia.]*
- *[Fai come ti ho chiesto o posto una foto di te nuda.]*

Insult: General Insult

Expressions that insult or offend the victim and not belonging to the following specific types.

Examples:

- *[Sei solo una testa di cazzo]*
- *[Spione!]*

Insult: Body Shame

Expressions that criticize someone based on the shape, size, or appearance of their body.

Examples:

- *[Sei un ciccone!]*
- *[Hai la grazia di un elefante]*

Insult-Discrimination: Sexism

Expressions with a sexist nature, such as prejudice or discrimination based on the victim's sex, gender or sexual orientation.

Examples:

- *[Faresti meglio a tacere e tornare nella tua cucina]*
- *[Frocio!]*

Insult-Discrimination: Racism

Expressions of discrimination that are based on the victim's race, skin color, ethnicity, nationality, or religion.

Examples:

- *[Sei uno stronzo ebreo]*
- *[Negro di merda]*
- *[Terrone!]*

Insult: Attacking Relatives

Expressions that insult relatives or friends of the victim.

Examples:

- *[Tua madre è talmente grassa che non entrerebbe nel Grand Canyon]*

Please note that the above insults are about the appearance of the mother however, it should be annotated as a case of "Insult: Attacking Relatives" and not as "Insult: Body Shame".

Curse or Exclusion

Expressions of a wish that some form of adversity or misfortune will befall the victim and expressions that exclude the victim from a conversation or a social group.

Examples:

- *[Spero che tu marcisca all'inferno]*
- *[Ucciditi], [non piaci a nessuno]*
- *[Non ti vogliamo più nella squadra]*
- *[Scarto della società]*
- *[Zitta!]*
- *[È tutta colpa tua]*

Defamation

Defamations are expressions that reveal confident, embarrassing or defamatory information about the victim to a large public or expressions that ridicule the victim in public.

Examples:

- *[Ho sentito che suo padre ha perso il lavoro ed è diventato un alcolizzato]*

Sexual Talk: Harmless Sexual Talk

Expressions with a sexual meaning, such as sexting between equals.

Examples:

- *[Vorrei baciarti]*

Sexual Talk: Sexual Harassment

Expressions with a sexual meaning that have a compelling character and that are considered undesirable (e.g. unwanted requests to talk about sex or to do something sexual).

Examples:

- *[Mandami una tua foto nuda!]*

Defense

Expressions in support of the victim. These can be uttered by the victim himself or by a bystander. The following 4 subtypes are all included in this category:

1. general expressions in support of the victim,
2. expressions that specify positive characteristics of the victim,
3. expressions of disapproval,
4. expressions showing the victim's indignation and helplessness.

Examples:

- *[Stai zitto su mia sorella,] [non è una troia!] (subtype 1)*
- *[Smettila,] [è una brava ragazza!] (subtype 2)*

- *[Ti rendi conto di quanto facciano male le parole?] [Lasciami in pace!]*
(subtype 3)
- *[Perché te la prendi con me?] [Cosa ti ho fatto?]* (subtype 4)

Harmful messages written by the victim should be annotated with the corresponding type of insult.

Encouragement to the Harassment

Expressions in support of the harassment from the bully or the bystander-assistant.

Examples:

- *[Ahah hai ragione!]*

Other

Expressions that contain any other harmful utterance than the ones described above.

- *[Sei uno 04 di merda]*

8. SARCASM

Whenever sarcasm is contained by one of the aforementioned categories, the annotator should mark it.

Examples:

- *[Come siete simpatici!] = come siete antipatici* (type=generic insult)
- *[Sei leggera come una piuma!!] = sei grassa* (type=body_shame)

9. NON-OFFENSIVE

In some contexts, messages containing insults are used jokingly and therefore they are not really offensive. Expressions that are not offensive in the context in which they are used, should be annotated by marking the *non-offensive* attribute.

In these cases, the cyberbullying role is not to be annotated whereas the annotator should mark the cyberbullying type.

Examples:

- *[Zoccolone mio!!]*
- *[Sei proprio una zoccola! ❤️❤️❤️❤️❤️]*

10. COMMENT

The annotator can use the field “comment” to add specific doubts related to a single message.

11. HOW TO USE CAT

Annotating a text span (markable)

1. Select the portion of text (i.e. one or more tokens) to be annotated by pressing **shift** and clicking on the relevant token/s. To select non-contiguous tokens, press **ctrl+shift** while clicking on the token/s. After the selection, selected tokens are highlighted in green. To remove the token selection, press **ctrl + z** or click on icon “Delete Selection” in the Text Panel menu bar or.



2. Confirm the annotation by pressing **ctrl+a** or clicking the icon “Confirm Markable” in the Text Panel menu bar. After confirming the annotation, selected tokens are highlighted in pink. To delete or modify the extent of a text span, **right-click** on it and chose the option you need (i.e. “Delete Markable” or “Change Markable Extent”).



Annotating attributes of a markable

1. Clicking on a markable, the Markable Attributes pop-up window appears.
2. Fill in the form with correct values.
3. Click “Save” otherwise the annotation will be lost.

Please note that the Markable Attributes window does not close automatically after saving. When finished, close the window clicking on the icon.

Suggestion: annotators can first annotate all the text spans and then concentrate on the annotation of the attributes. Following this procedure, we suggest to click on one text span after the other without closing the pop-up window: the window updates itself with information related to the new text span (check the text span you are annotating by looking at the “Current Extent” information at the top of the Markable Attributes window). Is however important to remember to **save the annotation of each single markable before passing to a new one.**

Markable Attributes

Current Extent:	#sfigato
Markable type:	Bullism_instance
role_type:	<input checked="" type="radio"/> Harasser <input type="radio"/> Victim <input type="radio"/> Bystander_defender <input type="radio"/> Bystander_assistant <input type="radio"/> Threat_or_B blackmail <input type="radio"/> Insult-General_Insult <input type="radio"/> Insult-Attacking relatives <input type="radio"/> Insult-Discrimination-Sexism <input type="radio"/> Insult-Discrimination-Racism <input type="radio"/> Curse_or_Exclusion <input type="radio"/> Defamation <input type="radio"/> Sexual_talk-Harmless_sexual_talk <input type="radio"/> Sexual_talk-Sexual_harassment <input type="radio"/> Defense <input type="radio"/> Encouragement_to_the_Harasser <input type="radio"/> Other <input type="radio"/> Insult-BodyShame
sarcasm:	<input type="checkbox"/>
comment:	<input type="text"/>
<i>*unsaved values</i>	
Save	