

# Reported Speech Guidelines:

## Sentence-based annotation on Social Media

### 1 Theoretical background

Reported speech takes place when one speaker reports what another speaker said.

Cases of reported speech are often enclosed in quotation marks, as in the example below.

(1) TAPPER: *Senator Rubio, last October, you said that you're "generally very much in favour of free trade."*

Such cases should be annotated differently in terms of ethotic appeals in our language resources as speakers might just report ethotic expressions stated by others without claiming support or attack on the targeted entity. In example (1) the speaker (Tapper) explicitly mentions other speaker (Senator Rubio), and directly recalls his words inside quotation marks ("you said that...").

### 2 Data

The data has been automatically separated into text units (sentences), which is a highly accurate but not a perfect method, thus the split of social media posts might be inadequate in some cases. Therefore, annotators are asked to look at the full posts provided in the annotation file to better understand the context of individual sentences.

Annotation of reported speech accompanies the annotation of ethotic appeals, thus when you open the file, you will see rows with text spans and descriptions of their properties in corresponding columns;

- *full\_text\_id* consisted of an individual, chronological number for given text unit;
- *comment\_url* consisted of an individual id for a given text unit;
- *conversation\_id* consisted of an individual id of a comment starting the Reddit sub-thread in which a given each text unit was responding to;
- *source* consisted of a name of the user who produced a given text unit;
- *full\_text* consisted of a given text unit;
- *full\_text* consisted of a full sentence a given text unit was a part of;
- *No\_ethos* to be filled with 1 if a given text unit did not consists of an ethotic move;
- *Contains\_ethos* to be filled with 1 if a given text unit consisted of an ethotic move;
- *Support (1)* to be filled with 1 if a given text unit consisted of an ethotic support;

- *Attack (2)* to be filled with 1 if given text unit consisted of an ethotic attack;
- **Reported Speech** to be filled with 1 if given text unit consists of quotation of another speaker;
- *Source* to be filled with the name of the author of a given text unit;
- *Target* to be filled with the name of the target of ethotic attack or support in a given text unit;

### 3 Guidelines

As reported speech we regard cases of quotation of another speaker. An example of reported speech in social media is illustrated in Figure 1 and in example (2), which is a sentence automatically extracted from the post. In this example the user reports an ethotic attack performed by other individual on Twitter. S/he however, does not perform an ethotic move.

(2) den0us: *So you originally say "someone didn't listen ins school" as an insult.*

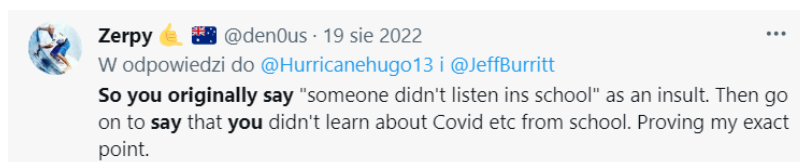


Figure 1: Reported speech in social media messages.

Although reported words are often enclosed in quotation marks, we find also cases of platform-specific notations of such cases. In Figure 2 we present a sentence where a user refers to the words of others and the style of notation of such cases on the Reddit site (top screenshot) and in a spreadsheet file in a text format downloaded with the use of the Reddit API (bottom screenshot in Figure 2).

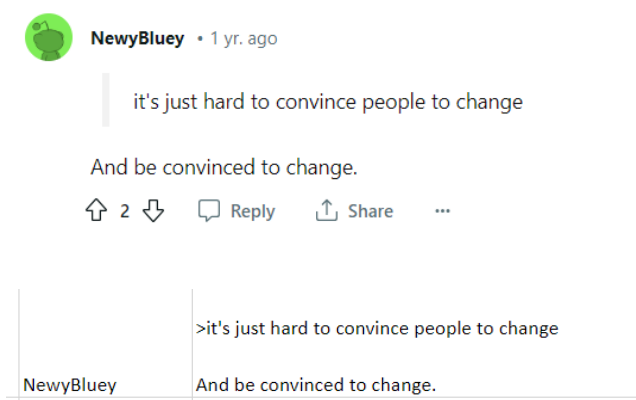


Figure 2: Comparison of notation of reported speech on the platform site (top screenshot) and in the obtained text file from the platform (bottom screenshot).

♣ **Rule 1.1 As reported speech we consider cases of nested speech or direct quotations enclosed in quotation marks.**

(3) Bob said Amy said that x.

Sentences such as in Ex. (3) are annotated additionally with reported speech besides rhetorical strategies of ethos and pathos. Ethotic appeals in such cases might be performed by the reporting speaker (Bob in Ex. (3)) or by the reported speaker (Amy in Ex. (3)). The latter are given an additional label of reported speech in our language resources.

**♣ Rule 2.1 Decide if a speaker reports what other speaker said.**

- Yes - mark “1” in the column “Reported Speech”;
- No - leave the column “Reported Speech” empty.

Annotation of reported speech in the PolarIs corpus needs to be adjusted to the specificities of the informal language and communication in the social media environment. Annotation guidelines need to account also for affordances of specific media platforms, in this case Reddit and Twitter. Thus, the language an annotator will encounter is very often different from ‘school English’, filled with colloquialisms and abbreviations.