

# Malta Government Budget Annotation Guidelines

## What is the goal of the project?

To annotate online posts (i.e. comments from newspaper articles and tweets from Twitter) written in Maltese and/or English about Malta's annual Government Budget for multiple social opinion dimensions, namely subjectivity/objectivity, sentiment polarity, emotion, sarcasm, and irony.

## What is each tag called and how is it used?

### Annotation Types

Each online post shall be annotated with the following 8 annotation types i.e., 5 different social opinion dimensions (*subjectivity*, *sentiment polarity*, *emotion*, *sarcasm*, and *irony*), and 3 others (*negation*, *off-topic* and *language*):

1. **Subjectivity:** binary value, with 1 referring to subjective posts and 0 to objective posts.

### Subjectivity Analysis

#### Subjectivity

Determines whether a piece of text expresses an opinion, in terms of personal feelings or beliefs



#### Objectivity

Refers to sentences that express some factual information about the world



2. **Sentiment Polarity:** categorical value (3-levels) for the sentiment polarity of the online post (negative, neutral, positive).

### Sentiment Analysis

Discovery of subjective expressions that describe people's sentiments, views or attitudes

toward entities, events, topics and their properties



POSITIVE



NEUTRAL

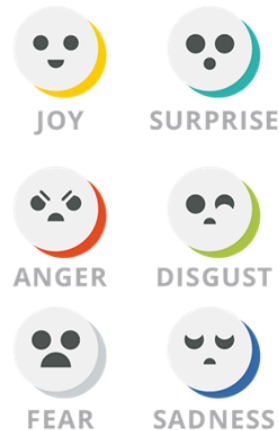


NEGATIVE

3. **Emotion:** categorical value for the emotion of the online post based on Robert Plutchik's 8 basic emotions (joy, trust, fear, surprise, sadness, anticipation, anger, and disgust)<sup>1</sup>.

## Emotion Analysis

Identification of human emotion from facial expressions or expressed subjective feelings and thoughts, such as anger, fear, sadness, disgust and surprise



4. **Sarcasm:** binary value, with 1 referring to sarcasm in online posts.
5. **Irony:** binary value, with 1 referring to irony in online posts.

## Sarcasm and Irony Detection

### Irony

Usually used to convey, the opposite meaning of the actual things you say, but its purpose is not intended to hurt the other person

### Sarcasm

Holds the “characteristic” of meaning the opposite of what you say, but unlike irony, it is used to hurt the other person

6. **Negation:** binary value, with 1 referring to negated online posts. A negated post refers to the opposite of what is conveyed due to certain grammatical operations/terms such as ‘not’.
7. **Off-topic:** binary value, with 1 referring to off-topic online posts that are not related to the Government Budget, even if they are of a political nature.
8. **Language:** numerical value, with 0 referring to online posts in English, 1 referring to posts in Maltese, and 2 referring to Maltese-English (Maltenglish) code-switched posts and 3 referring to any other language/code-switched posts. Note that '2' refers to an online post that contains switching (at least one normal/slang word) from one language to another (even multiple times) at a sentence/clause/phrase level.

## Annotation Examples

Refer to the Malta Budget 2019 dataset template, which contains 5 online posts annotated from each data source i.e., Times of Malta, MaltaToday, The Malta Independent and Twitter.

<sup>1</sup> <https://www.6seconds.org/2020/08/11/plutchik-wheel-emotions>

## Further Information

The following is further information in relation to the grey areas that the annotator may encounter when annotating the online posts:

- Retweets of **objective** posts are classified as still being **objective** in our case, even though they may reflect support to a particular budget measure.
- Retweets of **subjective** posts are classified as being **subjective**, that is, reflecting the opinion of the original post.
- If additional text is added to an objective retweet, this signifies an opinion on top of the original online post, hence is classified as being **subjective**.
- In the context of **objective** tweets, emotions are somewhat different than those for **subjective** tweets (which reflect the person's emotions) where *joy* is annotated to posts which shall bring a better quality of life and/or concrete support for certain budget measures.
- An online post having a *positive sentiment polarity* does not necessarily mean that the **emotion** reflected is a "*positive*" one i.e., *joy* or *trust*, since certain emotions reflect the user's (i.e. the user making the post) reaction to certain online posts (comments/tweets) made by other users. For example, online post with ID: 20190211 ("Then WHAT, exactly do you want?! Spending power and disposable income has been on the increase these last few years, so whatever business you're in, you MUST be better off; unless you are dealing with Stamperija Falluta, because then you have cause to be worried about your business/cash flow!") conveys a *positive sentiment polarity* and a *surprise emotion*.
- Certain online posts that convey **irony/sarcasm** may influence the respective **sentiment polarity** and/or **emotion**. However, this is not always the case. For example, online post with ID: 20190363 ("Min jaf x'jingibed mill inieher bhallissa - tant xoghol, tant srum barranin li ntlaghu buex isdir hsara lil Malta, u xorta johorgu figuri b'sahhithom hekk."), conveys a *positive sentiment polarity*, *joy emotion* and **sarcasm**.
- It is important to note for certain special characters, such as an exclamation mark (!), and/or emoticons, such as 😊, since these might affect the overall online post value of a particular annotation type e.g., **sentiment polarity** or **emotion**.
- Any reference to other political events/occurrences (e.g., references to a particular opposition Minister or measure introduced in a previous legislature) should be annotated as being **off-topic**, unless a direct comparison with the existing government budget is made, in which case it would be classified as being **on-topic**.
- Any online post that refers directly and/or is related to a particular budget measure announced by the Government for the respective year, should be annotated as being **on-topic**. For example, online post with ID: 20190014 ("Stqsi lil dawk li int ukoll kont tircievi kontijiet tad-dawl u l-ilma gholjien, staqsi lil dawk il-pensonanti li tahtkom ghal 25 sema ma hadux zieda, staqsi lil dawk l-eluf li kienu jirregistraw ghax-xoghol tahtkom u llum qed jaqighu x'jieklu, staqsi lil istudenti li llum mhux qed ihallsu ghal ezamijiet u ghat-trasport... Nahseb li m'ghandikx mera id-dar.....") is **on-topic** due to the sentence before last, which refers to two budget measures i.e., "Free public transport extended to 14- and 15-year-olds and all full-time students aged 20 and over." and "All Matsec exam fees to be scrapped", even though the first part of the post is totally off-topic.
- If an online post is referring/mentioning a general subject/domain, such as, rent, electricity and/or water, and is not related to any budget measure, this should be annotated as being **off-topic**. For example, online post with ID: 20190206 ("u ejja nahseb qieghad tmurt bil guh self employed") is **off-topic** since this has nothing to do with a work related budget measure for self-employed i.e., "Self-employed forced to shut up shop will be eligible for unemployment benefit until they find work".
- In the case of **off-topic** online posts, the overall general context of the said post (composed of one or more sentences) should be taken in consideration when carrying out all the other

annotations i.e., subjectivity, sentiment polarity, emotion, sarcasm, irony, negation and language.

- Online posts that are written in Maltese with the exception of one word e.g., budget, are still classified as being written in Maltese-English i.e., for the **language** annotation.
- If a link is provided in an online post, this should be ignored and not followed further to carry out the annotations. Therefore, all annotations should be based on the existing text provided in the online post, including any link text that might be relevant.
- If the name of an entity e.g., Person/Organisation within an online post is in a language that is different to the post's main language, this shall not be taken as being code-switched (2) or in another language (3), since it reflects the entity's original name. However, if the entity has both a Maltese and English name, the entity's written language is considered for the **language** annotation, e.g., Teatru Rjal (Maltese) vs. National Opera House (English).
- Should an online post contain multiple word/phrases that indicate **negation**, focus should only be given to those sentences/clauses/phrases that are **on-topic**, i.e., refers directly and/or is related to a particular budget measure announced by the Government for the respective year. Otherwise, the negations should be ignored.

## What parts of the text do you want annotated, and what should be left alone?

Online posts from four different data sources have been collected, namely three Maltese newswires in the Times of Malta, MaltaToday and The Malta Independent and one social networking service in Twitter. Each online post will be given a single value (label) for each of the 8 annotation types above. Note that one online post can be composed of more than one sentence. In such a case, the annotations should reflect the overall understanding of the entire online post by the annotator. For example:

- If an online post is composed of three sentences that signify a *negative* **sentiment polarity** in one sentence, and *positive* **sentiment polarity** in two sentences, the overall **sentiment polarity** should be *positive*.
- If an online post contains multiple sentences that convey different annotation representations e.g., experiencing both the *joy* and *surprise* emotions for the **emotion** annotation type, the primary **emotion** that is closest to the Government Budget topic (especially in cases where multiple topics are referenced in the online post), should be chosen.

## How will the annotation be created?

A spreadsheet that is prepopulated with all the online posts per annual Government Budget shall be used by each annotator. The values (labels) for each annotation type shall be in the form of a drop-down list.

## Other information

- The annotation types above are based on the first dataset that was annotated for the Malta Government Budget 2018. The annotation guidelines above were updated based on the feedback received from the annotators that carried out the task for this dataset and other observations. More information can be found [here](#).
- A summary of the Government Budget measures for each year should be primarily read by the annotator (i.e., before starting with the annotation process) to be familiar with the upcoming year's national measures. This shall help the annotator in understanding the opinions written by the general public, and determine which ones are relevant, valid and/or otherwise. The following are the suggested articles for each year's Government Budget:
  - Budget 2018: <https://timesofmalta.com/articles/view/budget-2018-at-a-glance.660017>

- Budget 2019: <https://timesofmalta.com/articles/view/budget-2019-at-a-glance.692338>
- Budget 2020: <https://timesofmalta.com/articles/view/budget-2020-at-a-glance-whats-in-it-for-you.742272>
- The annotation guidelines shall be evaluated by each annotator and task owner after the first 25 online posts, and then after the first 100 online posts from each different data source are annotated, to check whether any clarifications are required to the guidelines. In such a case, the guidelines will be updated accordingly after a common understanding is reached.
- During the annotation process any interesting observations and feedback should be written down by the annotator, so that these will be discussed further with the task owner.