

## Argumentation structure identification in Dutch user comments

### Overview procedure

You will be given 100 Dutch user comments to annotate. Each comment was made in response to a newspaper article from a Flemish newspaper. The comments were sourced from Facebook, specifically the official page of the Het Laatste Nieuws newspaper (<https://www.facebook.com/hln.be>).

The articles deal with a broad range of political topics, so the comments will likely reflect the opinions of users on those political issues. Most of the comments you will read reflect a user's opinion and the reasons they give for that opinion. In other words, they are *argumentative*.

The aim of these annotations is to discover the way in which people argue their particular position (*stance*) on a political issue online:

- Which specific position do they express towards the topic at hand? (*claim detection*)
- Which reasons do they give to justify their opinion? (*premise detection*)
- How do the different reasons interconnect? (*support, attack relations and subtypes*)

This round is the *first annotation round*, which is limited to topic, stance, segmentation and claim labelling (the first question in the above list). The second round will take place at a later date and will include premise and relation labelling.

**Your task is to annotate the comments for the topics, topic aspects and stances they contain.**

For more information about the concepts of argumentation underlying this annotation procedure, please read the following sections. Each section will introduce a few concepts and give some examples of actual annotations.

All the labels you will have to assign are marked in the following formatting: **label**. You will see the names of the labels correspond to the column names in the Excel file you have been given.

The estimated timing for each annotation task is added next to each section name. Summed up, the whole annotation procedure will take about 3 hrs in total. Please feel free to plan your annotations throughout the assigned period as you see fit based on this estimated timing. You do not need to complete the entire thing in one day. I would recommend trying to finish one task completely each time, for at least 50 comments, before taking a break. The two tasks need to be completed in the order given in this document.

A massive thank you to the participants! A short survey will be shared with you near the end of the annotation round to ask for any feedback you may have, which will be used to improve this annotation procedure.

Any questions or issues during the annotation process? Please send me a message via Slack at any time, so I can help resolve the issue.

### A note on the data

First, we want to give a word of warning on some of the characteristics of online user-generated comments which may make the annotation tricky at times.

As you may expect, online comments are *noisy* data. This means they very often do not contain standardized language use. The noisiness of user comments is for instance apparent in the occurrences of, formally:

- Character flooding: unnecessary repetitions of letters in a word or punctuation marks, for emphasis
- All caps for emphasis
- Irregular use of punctuation marks and spacing (for example, commas used in place of a full stop to signify the end of sentences or the absence of punctuation marks where they are needed)
- Missing verbs and incomplete sentences (try to infer the meaning of the incomplete sentence from the context, e.g. “België is tot ondergang” means “België is gedoemd tot ondergang”).
- Dialect words or other nonstandard language use
- Misspellings and word variations
- Emojis and gifs to illustrate a certain emotion

In terms of the content, the following aspects are typical of this type of data and may cause difficulties while annotating.

- Irony, sarcasm and other forms of humour
- References to the context of the discussion (could be: broader discussion context, immediate context of surrounding comments or post, context of the article text the comment is reacting to)
- Use of metaphors and other figures of speech
- Use of idiomatic expressions

Most of the challenges listed above will also figure in the example comments we give in the following.

#### Task 1: Topic detection (estimated timing: ~2hr)

Your first task is to identify the *topic(s)* in the comments. Some comments discuss topics of a more general nature, while others touch upon various subtopics and topic aspects of a broader issue. Your task is to identify both the topics and aspects in each comment. For some comments, you will only find a topic(s). For others, you will find multiple topic(s) and aspects.

**Please add all topic and aspect labels in Dutch.**

The difference between *topics* and *aspects* is made clear by the following examples:

*Wat had je gedacht, de tickets waren al vele jaren veel te goedkoop om nog deftig winst te maken.*

Comment with (interactional)<sup>1</sup> topic = luchtvaartmaatschappijen; aspect = economie

*Uw gezondheid zal hun een worst wezen. Wat telt is de economie om hun gat in de begroting te dichten met uw zweet. Niet meer of minder. En het niet dragen van een mondmasker tegen het advies van alle virologen in, is een ingeving van unizo die vrezen dat we dan minder geld gaan uitgeven in de winkels. Wiens brood men eet wiens woord men spreekt dus. Voor de rest sterkte wie morgen ziek wordt of vervolgens iemand dierbaar zal verliezen.*

---

<sup>1</sup> See below for explanation of “interactional topics”

Comment with (interactional) topic = corona; aspect = economie, mondkemperplicht, overlijden

These comments have different topics, but share the same *aspect* (economie).

The topics contained in the comments are called *interactional topic(s)*, which need to be labelled *along with the aspects for those topics*. Interactional topics are the topics which are touched upon by the individual commenter.

To annotate the topics

- Add the topic labels to the **InteractionalTopic(s)** column. Add all corresponding aspect labels to the **IT-Aspects** column.
- **A maximum of 5 topics and 10 aspects** is allowed per comment.

Since the comment is reacting to a Facebook post sharing an online newspaper article, the Excel file also contains the topic labels for both the post and the article. These are the *structuring topics and associated aspects*, and they represent the topics discussed in the wider context. Multiple comments will have the same structuring topic labels, since they are reacting to the same article. Given the fact that many users tend to not read the newspaper article shared via Facebook (for various reasons, including the paid subscription barriers), we make a distinction between structuring topics identified from a reading of the article (*StructuringTopicArt*) and those identified on the basis of the Facebook post (*StructuringTopicFB*).

The interactional topics and aspects which you will have to identify may in some degree overlap with the structuring topic(s), since the comment is reacting to the article, after all. However, many comments will touch upon different topics or aspects of topics than those presented by the article or in addition to those presented in the article.

For the annotation of the interactional topic(s), you are asked to base your decisions on the structuring topics associated with each comment. For example, the structuring topics and aspects labelled for the article and FB post associated with a comment include: *topics* – corona; *aspects* – measures, punishment, responsibility. Then the interactional topic + aspect pair of a comment will probably be similar. However, always read the comment carefully, because many users will comment on different *aspects* for the interactional topic-level than the ones presented on the structuring topic-level.

You will notice names of people occur as aspects of a structuring topic label which describes the category of the person (for example, *Maggie De Block* can be an aspect of the *politiek* label).

In the topics column, the various topics identified are separated by a comma. In the aspects column, the format is *(topic label)aspect label*. Multiple aspects are also separated by a comma. **Please use the same format to identify the interactional topics and aspects.**

#### **Example of interactional topic annotation:**

Comment: *Open VLD... t zijn de juiste om commentaar te geven op wat er allemaal gebeurt....t is net dat spreekwoord van die splinter en die balk .....zouden er toch al gauw nieuwe verkiezingen komen , misschien ....????*

**Structuring topic(s) FB:** racisme,betoging,politiek,politieker

**Structuring topic aspects FB (STFB-aspects):**

(betoging)veiligheid,(politieker)Sophie Wilmès,(politieker)eerste minister,(politiek)Open Vld

**Structuring topic(s) Article:** racism, betoging, politiek, politieke, corona, gezondheidszorg

**Structuring topic aspects Article (STA-aspects):**

(betoging) veiligheid, (politieke) Sophie Wilmès, (politieke) eerste minister, (politieke) Jean-Marie Dedecker, (politieke) Tim Vandepuut, (politieke) Brussels minister-president, (politieke) Rudi Vervoort, (politieke) burgemeester van Brussel, (politieke) Philippe Close, (politieke) Pieter De Crem, (racisme) discriminatie, (corona) maatregelen, (gezondheidszorg) personeel, (corona) ziekteverspreiding

**Interactional Topic(s) and Interactional Topic Aspects (IT-aspects):**

(politiek)Open VLD, (politiek)verkiezingen			
	Interactional Topic(s)	IT-Aspects	L
4	politiek	(politiek)Open VLD, (politiek	

You will notice in the above example, that the interactional topic and aspect list is very limited compared to the structuring topic and aspect list. Take care to always carefully read the comment first, since this example demonstrates the topic and aspect information presented in the comment can be quite different to the topic information extracted from the Facebook post and the article text.

Follow this procedure to decide the labels for the interactional topics of this comment:

1. Read through the whole comment in the Excel file
2. Look at the comment and the associated structuring topic label(s) and aspects side by side
3. Decide for each structuring topic label whether it may be considered an interactional topic in this specific comment (*Is the structuring topic explicitly discussed or implicitly alluded to in the comment?*)
4. If this is the case, add the same structuring topic label as an interactional topic
5. If the structuring topic label has an aspect label associated with it, determine whether the same aspect is discussed on an interactional level in the comment. If this is the case, add the same aspect label to the associated interactional topic label. Identify the remaining interactional topics in the comment, if there are any. Focus especially on the aspects which may be touched upon in the comment and add these as interactional topic aspect labels.

In deciding the topic and aspect labels, please **be as consistent and concise as possible in the formulation of the labels**. First of all, make sure your labelling is consistent to the structuring topic and aspect labels. In other words, if you encounter the aspect label *economie* in the structuring topic aspect column, make sure to maintain the same terminology when you encounter another comment which

talks about this aspect. The same goes for topic and aspect labels which did not occur in the structuring topic/aspect columns yet, try to be as consistent as possible.

## Task 2: Stance detection (estimated timing: 1hr)

The argumentative nature of a comment, especially if it deals with a controversial topic, entails that a *stance* or argumentative position is being argued. A commenter can be *for* or *against* a certain topic (aspect).

Similarly to the distinction between “structuring” and “interactional” topics, it is also possible to identify the stance in a broader and more narrow sense. Therefore, a distinction is made between the *debate stance* which is the stance towards the target of the whole debate and the *explicit stance* which is the stance presented by specific parts of the comment. This debate stance can be inferred from *explicit stance(s)* which rely on textual evidence from the utterance.

You **only** need to identify **the explicit stance** for the comments!

For the annotation of the stance for a comment, you will follow this procedure:

1. Carefully read through the entire comment
2. Check the list of interactional topic labels identified for this comment
3. For each interactional topic on the list, determine whether you can find an explicit indication of the stance of the author towards this topic in the comment.
4. If you can find an explicit stance indication, determine whether it expresses a pro/contra stance towards the topic and add “pro” or “contra” as the explicit stance label as well as link it to the interactional topic label it corresponds to in the following way: for Interactional Topic Label *mondmaskerplicht*, the stance label will be *(mondmaskerplicht)pro*. If you find more than one stance, make sure to add a comma separating the different labels. It is possible to find explicit stances for topics or for aspects. If you find one for an aspect, write the stance label as follows: *(aspect)pro*. If you find one for an interactional topic, write it as *(topic)pro*.
5. It is possible a topic label does not have an explicit stance label you can attach to it. In that case, simply don’t add a stance label for that topic. Some comments will not have clearly identifiable explicit stances, so don’t worry if you cannot assign any stance labels for that comment, simply write *NONE* in the stance label column.

You are not given a limit as to how many stance labels you identify per comment.