

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 segments of the comments for their argumentativeness and to identify the claim segment.

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.

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.

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: Argumentativeness labelling (estimated timing: ~ 2hrs)

In the Excel file you received, all comments have been divided into segments. Your first task is to label each segment as either *argumentative unit (AU)* or *non-argumentative unit (NAU)*.

To do this, read the segments and determine whether they should be considered content segments important to the argumentation contained in the comment (AU), or whether they have a different function (NAU). A question you may ask to help determine the argumentativeness of a given segment is: *Do you think the commenter is using this segment to either (1) state his opinion (stance-bearing) and/or (2) supply the readers with reasons as to why they should accept his opinion?* If the answer to either one or both of these questions is ‘yes’, then you should add the AU label next to the segment. If not, then the segment should be labelled NAU.

Watch out for trickier cases such as the segment “Wat voor een uitleg is dat weer.”. This particular segment can be used to end a comment (playing a more structuring role), but it also plays an important argumentative role, since it is indicative of the opinion (stance) of the author. On the other hand, segments like “Maar” have an structuring role, but are not argumentative in themselves, and so should be labelled NAU.

Tip: argumentative segments expressing opinions like “Wat voor een uitleg is dat weer.” tend to be segments which appear to carry a higher degree of involvement from the author. For instance, a degree of annoyance or exasperation may be noted. An emoji (sequence) may also function as a stance-bearing AU segment, e.g., a series of thumbs up.

Example comment: *60.000 besmettingen en bijna 10.000 doden. Dat is 1 op de 6 mensen dat sterft. Draait het of keert het maar dat zijn de slechtste cijfers van heeeeeeeel de wereld.*

Comment	Segments	AU/NAU
60.000 bes	60.000 bes	AU
	Dat is 1 op	AU
	Draait het o	NAU
	maar	NAU
	dat zijn de s	AU

In this example, the segment “Draait het of keert het” has been labelled *NAU*, since this segment may be used in other arguments and does not add any topic-specific or stance-expressive content to the argument. “maar” has been labelled *NAU*, since it is a connective that helps structure the argumentation (in this case, it indicates that an argumentative segment will follow). The segments labelled *AU* all convey important content to understand the argumentation presented by the author.

Task 2: Claim detection (estimated timing: ~1hr)

Your second task is to identify the main claim of the argument. **Only segments marked as *AU* in the preceding task are eligible to be labelled as *claim*.**

The **claim** is the text segment(s) representing the arguer’s main point (his position towards a particular topic). It is usually a controversial statement which should not be accepted by the reader without further support or evidence.

In the following example, the claim is the segment in bold and may be summarized as: “I support a longer second lockdown.”

I hope we get a 2nd lock-down soon, preferably long enough for about 4 months then we'll be rid of this misery!

Always bear in mind that the claim represents *the arguer’s position*. Therefore, look for words signalling the personal position of the author (e.g., “I hope”, “preferably” in the above example). However, these indicators are not always present.

The claim may be restated by another segment in the same comment. In that case, label all the claim restatement segments as *claim* and add the label **restatement**. A restated claim has to express almost the exact same thing as the claim. In other words, you should be able to substitute the claim for its restatement without changing the meaning of the comment or its argumentation.

Add the *Restatement* label in the Excel file next to the claim segment:

AU/NAU	Claim	Restatement/Im
AU	Claim	Restatement
AU		
AU	Claim	
AU		
AU		

In some cases, the claim may not be explicitly stated in the comment. If this is the case, add the label *implicit claim* on the same row of the first segment of the comment. This claim can only be inferred by the reader, it is impossible to link one (or more) segments pointing directly to it.

E.g.:

Comment	Comment	Segments	AU/NAU	Claim	Restatement/Im
		☹️👤	NAU		
23	Laat ze alle	Laat ze alle	AU		
24	Er zitten hi	Er zitten hi	AU		Implicit claim
		is het niet o	AU		
		zal aan mij	NAU		
		ben dan oo	AU		

The claim may be split over multiple segments or be interrupted by another (non-argumentative) segment. In such cases, make sure to label all the segments representing the arguer's main claim. As a rule of thumb, try as much as possible to look for a single segment identifying the claim. If you can't find one, first determine whether the claim is implicit or whether there is a restatement. If none of these are the case, only then you can consider adding the *claim* label to all segments you think represent the claim.