

# Annotation Guidelines for Labeling Topics, Aspects, Stance, Argumentativeness and Claims in Dutch social media comments

**Version 1.0**

*LT3 Technical Report – LT3 20-11*

Nina Bauwelinck and Els Lefever

LT3 – Language and Translation Technology Team  
Department of Translation, Interpreting and Communication  
Ghent University

All rights reserved.

LT3, Faculty of Arts, Humanities and Law, Ghent University, Belgium.

December 23, 2020

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Procedure . . . . .	1
<b>2</b>	<b>Topics and Aspects in articles and screenshots</b>	<b>2</b>
2.1	What is a topic? . . . . .	2
2.2	What is an aspect? . . . . .	2
2.3	labeling topics and aspects . . . . .	2
<b>3</b>	<b>Topics and Aspects in comments</b>	<b>5</b>
3.1	What is a topic? . . . . .	5
3.2	What is an aspect? . . . . .	5
3.3	labeling topics and aspects . . . . .	5
<b>4</b>	<b>Stance in comments</b>	<b>7</b>
4.1	What is a stance? . . . . .	7
4.2	labeling topics and aspects . . . . .	7
<b>5</b>	<b>Argumentativeness of segments</b>	<b>9</b>
5.1	What is argumentativeness? . . . . .	9
5.2	labeling argumentativeness . . . . .	10
<b>6</b>	<b>Claim detection</b>	<b>11</b>
6.1	What is a claim? . . . . .	11
6.2	labeling claims on comments . . . . .	11

# Chapter 1

## Introduction

These annotation guidelines have been drawn in the framework of the PLATOS project. The PLATOS project aims at investigating stance and argument detection on automatically extracted topics in social media text by combining linguistic knowledge and machine learning. In this research, we want to build further on research in aspect-based sentiment analysis for the automatic extraction of topics, their related features and sentiment and the stance taken by the author. In addition, we will push the state-of-the-art in the research field of argumentation mining by modeling argument structures in social media text in order to detect why people have a certain viewpoint about a concerned issue. The proposed research will be evaluated for the use case of political social media text, which is a very relevant test domain because (1) it is important to have a flexible approach to detect emerging and trending political topics, (2) political posts have by nature a more explicit stance and persuasive goal, which makes them a good case for argumentation research and (3) political UGC shows a large variety of language use, ranging from very noisy (blogs, comments) to more standard text (political posts from representatives).

The current version of the guidelines contains detailed guidelines for labeling topics, aspects, stance, argumentativeness of segments and claims. A next version of the guidelines will also contain detailed annotation guidelines for labeling relations between arguments and premises.

### 1.1 Procedure

The annotation consists of five steps:

1. labeling of Topics and Aspects based on news articles and screenshots
2. labeling of Topics and Aspects based on online comments
3. labeling of Stance based on comments
4. labeling of argumentativeness of comments
5. labeling of the claim of the argument in the comments

## Chapter 2

# Topics and Aspects in articles and screenshots

### 2.1 What is a topic?

The topic of a text answers the question “What is this text about?” It is often the case that multiple topics can be assigned to one text. An example of a topic: *corona*.

### 2.2 What is an aspect?

An aspect is a specific component of a topic, and answers the question: “Which specific aspects of the topic are addressed in the text?” An aspect is thus always linked to a given topic, e.g., *corona\_measures*. As was the case for the topic labeling, it is also possible to assign various aspects to a single text. It is not necessary to assign specific aspects to every topic of the text, as some topics will not be addressed in much detail in the text.

### 2.3 labeling topics and aspects

The first task consisted of finding topics and aspects in a news article text and accompanying screenshots

Important guidelines for labeling topics and aspects are the following:

- Always start from the content of the text. If the article or screenshot deals with *corona*, use this term as the topic label, and not another term (e.g., *COVID*). If the text contains various synonyms for the same concept, choose one specific term and be **consistent** in using this term.
- Choose as much as possible **one-word-labels**. It is, for instance, better to replace *measures for corona* (Dutch: *maatregelen corona* by *corona measures* (Dutch: *coronamaatregelen*, or to replace *make face masks obligatory* (Dutch: *mondmasker verplichten* by *face mask requirement* (Dutch: *mondmaskerplicht*).

- When you have to assign multiple words as a label, try to be as **concise** as possible, e.g., *parliamentary agreement* instead of *agreement in the parliament*.
- Only label topics and aspects that frequently occur in the article. Not every term you encounter in the article is a good candidate topic/aspect. In case of doubt whether a topic/aspect is relevant, go back to the title and first two paragraphs of the article, that normally mention the **focus** of the article more clearly.
- You can identify a maximum of 5 topics per screenshot, and 15 topics per article. The number of aspects is not limited, but try not to exaggerate. Try to only label those topics and aspects that are **most relevant** for the article.

For labeling, you apply the following steps:

1. Read the entire article once (normal reading speed)
2. Define the topics (the main subjects the article addresses) and aspects (the topic components addressed) in the article.

Caveat:

- Only add topics that are generic enough, otherwise it is better to add them to the aspect labels, e.g., *corona* is sufficiently generic and can be a topic label, but *corona measures* is rather a component of a broader topic, and often functions better as an aspect label.
- The previous rule is no set rule, it often depends on the specific article text. If, for instance, an article only talks about corona measures, but does not comment on a specific measure, it is possible to assign the following labels:

Topics	Aspects (separted by comma)
corona	measures

However, if the article addresses a particular measure in more detail (e.g. *face mask*, you have to assign the following labels:

Topics	Aspects (separted by comma)
corona	measures
measures	face mask

This example shows that a given label can function both as a topic and as an aspect, depending on the article content.

3. Assign the list of topics to the article ID
4. Assign the list of aspects, and make sure you put the aspect next to the topic it belongs to. If you detect multiple aspects for one topic, separate them by commas.

Topics	Aspects (separted by comma)
politics	minister of public health Maggie De Block
corona	measures, contamination
economy	

Cave at: Persons/proper names can also be aspects. In case of politicians, it is important to first add the official job title or function (if mentioned in the article), and then the name of the person, e.g., *minister of public health Maggie De Block*, *virologist Marc Van Ranst*.

5. You follow the same procedure for the screenshots: determine the major topics and related aspects addressed in the screenshot. To this end, you receive both the text and the image. Add these labels as well. Make sure you link the correct aspects to the related topic and to separate multiple aspects by comma.

## Chapter 3

# Topics and Aspects in comments

### 3.1 What is a topic?

The topic of a text answers the question "What is this text about?" It is often the case that multiple topics can be assigned to one text. An example of a topic: *corona*.

### 3.2 What is an aspect?

An aspect is a specific component of a topic, and answers the question: "Which specific aspects of the topic are addressed in the text?" An aspect is thus always linked to a given topic, e.g., *corona\_measures*. It is again possible to assign various aspects to a given text. It is not necessary to assign specific aspects to every topic of the text.

### 3.3 labeling topics and aspects

This task consists of detecting topics and aspects in comments.

Important guidelines for labeling topics and aspects are the following:

- Given the fact that comments are smaller texts, you are limited in the number of topics that can be assigned to a comment (5 topics maximum). The number of aspects is not limited.
- **Very important:** base your labeling decisions only on the information that is explicitly available in the text. In other words, try to "switch off" your background knowledge as much as possible for this task. And, in cases of doubt, check if you can link the labels you are unsure of to specific words in the given comment. If you cannot connect the label to specific words, you are not allowed to assign it as a label for that comment. E.g., *She should have left when she let Corona in along with Beke and the rest of the 9 musketeers who manage public health*.

This comment refers to a person ("She"). The context (as evident from the news article this comment is reacting to) indicates this person to be "Maggie De Block", the former Minister of Public Health in Belgium who was one of the first politicians to be accused of not handling the corona crisis well. However, since her proper name is not explicitly

mentioned in this comment, it was not allowed as a label for this task. Labels which were appropriate for this comment included "Beke", "public health" and "corona", because all three are mentioned explicitly in the text.

Note that the label "Beke" is assigned as an aspect of the topic label "politics". This is because the official function title of this person is not mentioned in this comment, so it cannot be added from the annotators' background knowledge.

- As a consequence of the brevity of these texts, almost all of their content words are candidate topic/aspect labels. This is contrary to the article texts from the previous annotation task, where not every content word appeared relevant as a topic/aspect label.
- Try to limit yourself as much as possible to *one word labels*. In cases where this is not possible, try to keep the label as concise as possible.

For labeling, you apply the following steps:

1. Read the entire article once (normal reading speed).
2. Define the topics (the main subjects the comment addresses) and aspects (the topic components addressed) in the comment.

Caveat:

- Stay as close to the text as possible. Do not fill any context or background information in yourself to be able to add a specific topic/aspect label. If it's not explicitly mentioned in the text, it cannot become a label. E.g., *What does that dude know or claim to know ?? If there is to be a restriction then it needs to be possible to monitor it. This isn't even possible on the trams at the coast: you still see people without protection there.* Even if you can guess from your background knowledge that this comment most likely deals with corona (and the face mask requirement), you are not allowed to add these as labels. Instead, correct labels for this comment include:

Topics	Aspects (separated by comma)
requirement	
monitoring	
transportation	coastal tram
protection	

- The aspect label always need to be linked to the specific topic it belongs to. If you find multiple aspects for one topic, separate them by comma.



# Chapter 4

## Stance in comments

### 4.1 What is a stance?

The stance represents the viewpoint the user takes with regards to a specific topic and/or aspect. **Not every topic/aspect in the comment will have a stance linked to it.** Sometimes a topic is only mentioned, but not evaluated. This task requires the annotator to find only the topics/aspects the comment is clearly expressing an opinion on.

*E.g., I personally regret that it is not mandatory, especially now that we are allowed to shop with more people. I also believe that if you wear a mask, you also respect others. And, yes, I wear a mask myself. Best to everyone and keep it healthy!*

This user is clearly a proponent of the face mask requirement. He also touches upon other topics, like "health". The difference is that he does not express his stance or opinion about that.

### 4.2 labeling topics and aspects

This task consists in labeling the stances expressed in the comment. A stance label is always assigned towards a specific topic or aspect label. Four possible stance labels exist:

- **pro:** the author of the comment seems to be a proponent
- **con:** the author seems to be an opponent
- **neutral (neu):** the author does not appear to take sides, he may be both proponent and opponent
- **none:** no stance is being expressed towards the given topic/aspect label

Important guidelines for labeling stance are the following:

- The annotator is given topic and aspect labels which have been previously annotated. Their task is only to determine the presence and type of stance towards these labels.

- Not every topic/aspect label associated with a particular comment will receive a stance label. E.g., *What does that dude know or claim to know ?? If there is to be a restriction then it needs to be possible to monitor it. This isn't even possible on the trams at the coast: you still see people without protection there.* In this example, multiple topics and aspects are touched upon, but only a few of these have an associated stance. The labels for this comment might look like this:

Topics	Stance (pro/con/neu/none)
requirement	neu
monitoring	pro
transportation	none
protection	pro
coastal tram	none

For labeling, you apply the following steps:

1. Read the entire article once (normal reading speed).
2. For every label in the topic/aspect column, determine whether the comment expresses a stance towards that label and select one of the four labels from the list in the stance column next to the topic/aspect. Use the "none" label if you cannot find a stance for the topic/aspect.

## Chapter 5

# Argumentativeness of segments

### 5.1 What is argumentativeness?

A segment/sentence is argumentative, if it used to construct an argument or if it is in some way relevant to the argument. This argumentative relevance may be construed in the following ways, represented by different roles the segment plays within the comment:

- Making a claim about a subject or giving one's opinion on the subject;
- Giving evidence for a claim;
- Expressing a stance towards a given topic (see stance annotation subtask).

E.g., the following comment consists of three argumentative segments, each playing a different role within the comment. All three segments are relevant to the argumentation constructed by the author to support their viewpoint (in this case, the argument may be summarised as "Implementing the face mask requirement so late has stopped people from taking this measure seriously, as is evident from the example I have seen myself.").

*/Isn't it a bit too late?../stance /the people who do wear one do so as a fashion accessory [upside down face emoji]./claim/opinion /witnessed this Sunday [see no evil emoji].. a family of five.. only mommy wears 1.. matches her top [rolling eyes emoji]./evidence*

Segments can also be non-argumentative. This means that they are not relevant to the argument of the author of the comment. These are usually segments used to introduce another segment or to structure the text.

E.g., the following comment contains a "but" segment solely meant to structure the comment. It announces the second part of the contrasting expression ("high moral highground" vs "nothing else is high"). In other words, this segment is **in itself** not argumentatively relevant (NAU).

*/What does Mrs. Rutten think she will gain from this?/AU /Her moral highground remains high,/AU /but/NAU /that's the only thing./AU*

## 5.2 labeling argumentativeness

This task consists of labeling predefined segments of comments as argumentative (AU) or not (NAU).

Important guidelines for labeling argumentativeness are the following:

- Sometimes an author will express their opinion, without explicitly stating what exactly it is they are giving their opinion on. E.g., a segment like "This is ridiculous" does express a stance (even if it is not clear what exactly it is targeted at). Such segments also count as argumentative and should be labeled as such.
- Some segments can play both a structuring and a stance expressing role. E.g., "I get what you're trying to say": this segment expresses a stance (the author appears to be a proponent of or at least neutral towards something said previously in a discussion), but it is also used to structure the text (the segment predicts the following segment due to the specific structure that is typical of such expressions: "I get what you're trying to say, **but...**"). Segments like these are **argumentative**, since they are relevant to the argument (because of their stance expression in this case).

For labeling, you apply the following steps:

1. Read the entire article once (normal reading speed).
2. For every segment, determine whether it is argumentative (AU) or not argumentative (NAU).
3. In cases of doubt, always reread the guidelines for this task.

## Chapter 6

# Claim detection

### 6.1 What is a claim?

A claim is the conclusion to an argument. It is the sentence/part of the sentence representing the point of the comment. The claim answers the question "What is the central thesis of the author?"

### 6.2 labeling claims on comments

This task consists in identifying the central claim (conclusion) for each comment. The pre-segmented text is shown to the annotator, along with the argumentativeness labels for each segment (output of the previous task). **Only segments marked as argumentative (AU) are eligible claims.** Two possible labels can be assigned for this task: CLAIM for the claim segment and NOT for every other segment in the comment.

Important guidelines for labeling claims are the following:

- In some cases, the claim corresponds to the stance of the comment. E.g., *I've been wearing one since the very start. Used to admire Marc van Ranst but I'm starting to wonder whether even he knows what's going on.* It is not easy to indicate the claim of this comment. In this case, the central point the author is trying to make may be summarized as "I'm a proponent of something (in this case, face masks)". In cases like these, the claim corresponds to the stance of the author, even if the stance is not explicitly expressed here. Therefore, we added the label "Claim is stance" to accommodate those cases.
- Always try to find a segment representing the claim. Only when this turns out to be truly difficult for a specific comment, you are allowed to use the "Claim is stance" label.
- In most cases, a single segment will contain the central claim of the comment. In rare cases, you will have to indicate two segments as the claim.
- Expressions that indicate the presence of a stance (e.g., "I think that" or "Personally") are often contained within the claim segment. Segments expressing some sort of advice or council (e.g., "We should go do X") or containing an imperative "Just wear a mask when going to the shops!!!" can signal claim segments.

For labeling, you apply the following steps:

1. Read the entire article once (normal reading speed).
2. For every segment labeled "AU", decide which of them best represents the central message of the comment. Label this segment as "CLAIM". Label the rest of the segments as "NOT".
3. If you are unable to detect a claim segment, first try to determine whether the claim corresponds to the stance of the comment. The annotators are not given the stance labels, so they have to formulate a stance themselves, since it is not always explicitly expressed (and if it is, the annotator should be able to identify the stance in a single segment). In such cases, indicate the "Claim is stance" label for the given comment.

## References