

*Pieter Delobelle, Kristen Scott,  
Bettina Berendt*

# **Code book**

## Corona Sentiment on Twitter

APRIL 19, 2021

*KU Leuven*



# Contents

<i>Introduction</i>	5
<i>Coding procedure</i>	7
<i>Guidelines and examples</i>	9
<i>Missing context</i>	13
<i>Overview of codes</i>	15



# *Introduction*

This code book is meant for the *Corona Sentiment on Twitter* project. The goal is to gauge if people support specific measures taken by the Belgian government during the COVID-19 pandemic.

*Goal: Quantify support for Belgian corona-measures*

The tweets that are collected are in Dutch, French and English and passed through a rough location filter to filter for tweets from Belgium.

The goal of the project is to gauge if people support specific measures that are either established by the Belgian government, or that might be established, as well as to gauge the level of support in the government's handling of COVID-19 in general.



# Coding procedure

## Procedure

*Step 1.* Determine if the tweet is relevant. The following aspects are taken into account:

- Is the tweet about measures proposed or established against the spread of COVID-19?
- Is the tweet about the Belgian government's handling of the COVID-19 pandemic, even if no specific measure is mentioned?<sup>1</sup>

If neither of the above is true, then the tweet can be considered irrelevant. Label it with the 'irrelevant' tag that is provided, and move on to the next tweet. If at least one of the above is true, a tweet is relevant and move on to Step 2.

*Step 2.* If a tweet is relevant, than you need to label it with at least three labels, one from each of three categories (*topic, approval, government*). Definitions of these terms of labeling instructions are given below.

<sup>1</sup> It helps to be aware of dialects, e.g. "mondkapje" (the Netherlands) versus "mondmasker" (Belgium).





# *Guidelines and examples*

## *Codes*

There are three categories of codes: topics, approval, and government. All relevant tweets must be labeled with at least one tag from each of these categories. In some cases only one or two categories will actually be covered in the tweet - please use the *[category]:not-applicable* tag for categories not covered in a tweet.

## *Topics: coding specific measures*

Relevant topics that need to be coded are references to any specific measures (mask, lockdown, etc.) that the Belgian government has implemented, discussed, or could implement in order to mitigate COVID-19. If the tweet is referencing any of these topics, it can be coded with the appropriate topic codes (see fig.1 for an example). If you find a topic that you believe fits this criteria but is not on the list of labels, be sure to use the *topic:other-measure* label.

- `topic:masks`
- `topic:curfew`
- `topic:quarantine`
- `topic:lockdown`
- `topic:schools`
- `topic:testing`
- `topic:closing-horeca`
- `topic:vaccine`
- `topic:other-measure`

If no specific measure is mentioned, label the tweet with:

- `topic:not-applicable`

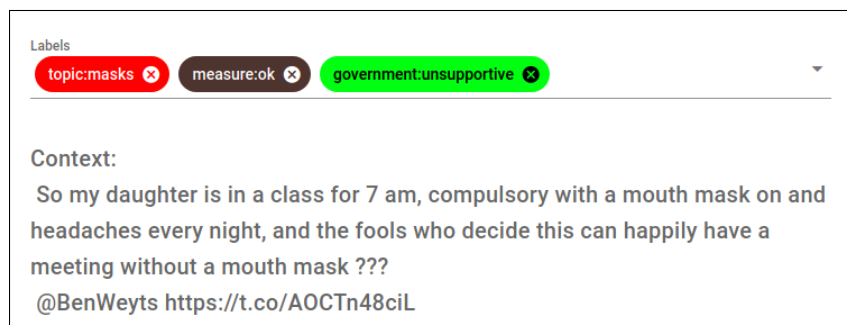


Figure 1: Screenshot of labeling tool Doccano with a tweet labeled with a 'topic' label. This tweet clearly mentions the policy of mask wearing, and therefore is labeled with *topic:masks*. The other labels seen in the image will be described below. Note that the referred user is Ben Weyts, a Belgian politician.

### *Measure: coding approval of specific measures*

This category of codes is used to indicate the tweet author's approval of any COVID-19 measure (whether in effect or not); specifically, do they find it to be too-strict, a good level of strictness, or not strict enough (see fig. 2 for an example). Use this category of codes even if a specific measure is not mentioned, but an opinion about the strictness level of government measures in general is expressed .

- `measure:too-strict`
- `measure:ok`
- `measure:too-loose`

If the tweet is not about the Belgian situation, label the tweet with:

- `measure:not-applicable`.

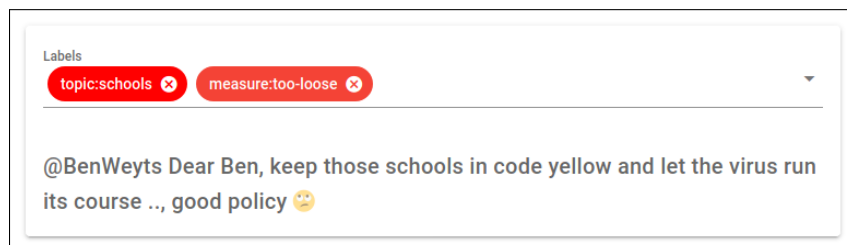


Figure 2: Screenshot of labeling tool Doccano with a tweet labeled with a 'measure' label. The 'eye rolling' emoji and other language indicates that this person finds schools remaining in code yellow (low alert) to not be a strict enough measure. *note, Ben Weyts is a public figure.*

### *Government: coding overall support of the government's approach*

This category of codes is used to indicate the general sense of support, approval of, or trust in, the Belgian government in handling the COVID-19 pandemic <sup>2</sup>. A government code should be used if the tweet

<sup>2</sup> This includes the Belgian, Flemish, Walloon, and Brussels government.

gives an indication of the author's answer to a questions like "Do you trust the Belgian government to properly handle the COVID-19 pandemic?" or "Do you support how the Belgian government has been handling the COVID-19 pandemic?". This category differs from the *approval* category in that it is meant to indicate a sentiment towards the government and their handling of the pandemic, rather than towards a specific measure. Additionally, this category does not identify opinions regarding strictness levels, but rather captures positive or negative (e.g. supportive or unsupportive) sentiments. A tweet can be labeled with codes from both the government and the approval category, if appropriate (see fig. 3 for an example).

- government:supportive
- government:unsupportive

If the tweet is not about a governmental organisation of Belgium, label this as:

- government:not-applicable.

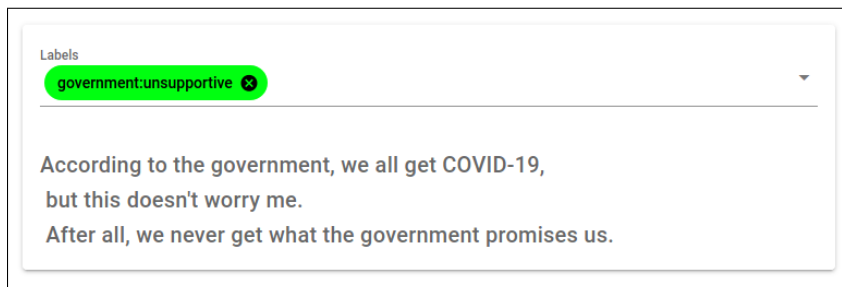


Figure 3: Screenshot of labeling tool Doccano with a tweet labeled with a 'government' label. While this tweet does not indicate the author's opinion towards any particular measures, it does clearly express a general lack of trust in the government.

### *Irrelevant*

- irrelevant

### *How many codes can I use?*

You can label tweets with as many tweets as you deem relevant. Topic, approval, and measure labels can all exist together. The only label that excludes other labels is 'irrelevant', which, if used, must be the only label used.



## Missing context

Some tweets are part of a conversation thread and others are retweets<sup>3</sup>. The content of a tweet is provided. More specifically, if the tweet is a reply or a retweet (with RT @. . .), the original content is provided as shown in Figure 4.

<sup>3</sup> Retweets are represented with RT @. . ., even if the Twitter interface used the retweet button. This differs from liking a tweet, since this also adds the tweet to the user's timeline.



Figure 4: Screenshot of labeling tool Doccano with a Dutch tweet where context is provided. The tweet that is labeled is shown in the lower part prefaced by Tweet:.

## Ambiguous tweets

The topic or context of a tweet is often ambiguous; it may be part of a conversation that is not included, or in reference to some topic that is no longer clear when reading it in a different time or place, or they may contain slang, hashtags or 'dog whistles' that the coder may not be familiar with. Some level of research into the context may be required: this includes clicking on any included links, and googling any terms or hashtags that you are not sure of. If this a few minutes of this kind of research does not make clear what the topic or sentiment of the tweet is, please move on: mark the tweet as 'irrelevant' (or leave blank).

Many of the tweets are 'retweets' of some other person's tweet (indicated by 'RT'). Code retweets as if they are the original tweet. The exception is if commentary about the original tweet is included in the

retweet (this text will appear before the 'RT'), use this commentary to determine what is being expressed. For example, if the original tweet expresses support of the government, but the included commentary states "This person is wrong! The government has been horrible!", then code as *government:unsupportive*.

### *Location*

In some cases, the location filter is not that accurate. It can also happen that policies or situations in other countries—or no particular country—are discussed in a tweet.

- For coding the topics, *the location does not matter*.
- But for coding the support and government's trust, tweets can only be labeled if they are *from Belgium and in reference to the Belgian government, with reasonable likelihood*.

## Overview of codes

Topic	Approval	Government	Relevance
topic:masks	measure:too-strict	government:supportive	irrelevant
topic:curfew	measure:ok	government:unsupportive	
topic:quarantine	measure:too-loose		
topic:lockdown			
topic:schools			
topic:testing			
topic:closing-horeca			
topic:other-measure			
topic:not-applicable	measure:not-applicable	government:not-applicable	

Table 1: Reference of all the code groups mentioned in this document.