Labeling Guide

Overview

The goal of this project is to identify "influence campaigns" in social media and traditional news media. To do this, we will be marking up a corpus of *documents* to indicate whether each document in the corpus contains evidence of "agendas," "concerns," and "emotions." (See below.)

This document provides an overview of the scenario itself and the process you should take to label the documents; you should also familiarize yourself with the "ACE" document before beginning annotation.

Scenario

The current scenario that we are exploring is the 2017 French election campaign. This election had two rounds - during the first round (April 23, 2017), no candidate garnered enough votes to win, but Emmanuel Macron and Marine Le Pen proceeded to a run-off election, which took place on May 7, 2017. Ultimately, Emmanuel Macron won the election.

What is a "document"?

For purposes of this project, a **document** can be any sized body of written text. A document may be as short as a Twitter post, or it may be as long as a newspaper op-ed. We are currently only labeling written text; we will not be listening to audio files or watching videos, though we may need to label transcripts of audio files.

Examples of document types:
- Twitter post
- Newspaper article
- Newspaper editorial
- Reddit post
- Blog post
- Other types of social media posts
- Transcripts of audio or video recordings (e.g., a transcript of a political talk show)

Labeling task

Your task is to read through each of the "documents", one document at a time. For each document, you will need to indicate the agendas, concerns, and emotions that are found in that document. You may find that a document will have more than one label in each category; if that's the case, then you will select **each** label that is relevant. For each category, we have included a description of the category and a list of possible labels below. We have also included an "other" option where you can write in a label that you believe fits the document but that is not on our list.

When you are deciding how to label each document, please do not spend a lot of time debating about what to choose. Select the option(s) that fit best and then move on. Each document will be labeled by three separate labelers.

You should do your labeling without discussing your labels with anyone else and without asking for help. This is important, because we want three independent labelers to provide independent assessments of the posts, without being influenced by someone else's opinion. We will also be calculating some statistical measures of agreement (called "inter-annotator reliability") to better understand how much agreement the annotators have on what is present in a particular post.

If a post includes a link (e.g., to a news story or other web page), you may use the linked content to inform your annotation. However, you shouldn't spend much time examining the linked page - a quick check to see what the page is about is sufficient.

In the annotation spreadsheet, you will see three columns for each category (Agendas, Concerns, and Emotions). You should add a single label in each column. At the top of the spreadsheet, we have labeled a few tweets as examples.

What if there is insufficient information?
If a document does not contain enough information to provide a label for a category, you should label it with a 0. This indicates that a given category (agenda, concern, or emotion) is not exhibited in the dataset.

For some documents, you may find that only one or two of the three categories are exhibited. For some documents, you may find that there is not enough information to code any of the categories.

See, for example, the following tweet, which references candidate Benoit Hamon, but otherwise provides insufficient information for labeling:

@benoithamon super 👍

Do I have to label at least one agenda, concern, and emotion for every document? It is possible for a document to exhibit only a single category. In the following tweet, the concern "terrorism" is present, but there is no clear agenda or emotion expressed:

France determined to combat #terrorism everywhere in Europe, President @EmmanuelMacron says at UK Embassy in Paris

Similarly, this tweet expresses an emotion (love/admiration), but no clear agenda or concern:

b.r.a.v.o Benoit we love 💗💗💗💗💗💗💗💗💗💗💗💗💗💗💗💗

Finally, there are many concerns and agendas that are not included in our "core" list. For these tweets, if you think there is a clearly important concern or agenda, you can write it in the "other"

column in the spreadsheet. For example, for the tweet below, you might write "LGBTQ" in the "other" column:

@EmmanuelMacron Nous avons évoqué le cas des personnes LGBT en Tchétchénie. J'ai très précisément indiqué au président Poutine les attentes de la France.

Agendas:

An agenda is what the author wants the audience to believe or do as a result of their message.

There is an almost endless list of possible agendas. We will not be able to label all possible agendas, but will focus on a limited, core set of agendas that we believe are important for understanding influence campaigns in this dataset.

List of agendas

Agendas are listed below; you can find a complete description of each agenda, as well as example agendas, in the ACE document:

1: BELIEFS (that will have an impact on the election)
1.1 Believe that the election process is flawed and/or manipulated by ENTITY (including potential foreign interference)
1.2 Believe that the election process is fair and has not been manipulated.
1.3 Believe that ENTITY or GROUP is immoral/unethical/dishonest/harmful
1.4 Believe that ENTITY or GROUP is moral/ethical/honest/beneficial
1.5 Believe that you/GROUP are at risk
1.6 Believe that your actions can lead to a good outcome/hope for GROUP

2. ACTIONS (that will have an impact on the election)
2.1 Share information and opinions author endorses
2.2 Voting behaviors
    2.2.1 Vote for ENTITY
    2.2.2 Vote against ENTITY
    2.2.3 Vote
    2.2.4 Don't vote
2.3 Take action: protest/demonstrate/attend rally/volunteer/campaign

Agendas will often be **explicitly expressed** in a message or a hashtag, but in some cases they may be **implicitly expressed**. You should label both implicit and explicit agendas. Implicit agendas are ones where the belief or action is not explicitly mentioned, but the information in the tweet clearly is intended to guide the reader towards that belief or action. For example:

Explicit: Vote for Macron!
Implicit: It's vital that we elect Macron. (implies: you should vote for Macron)

<u>Explicit</u>: Come join me at a rally at 3pm at the Arc de Triomphe!
<u>Implicit</u>: There's a rally at 3pm at the Arc de Triomphe! (implies: you should attend)
<u>Not codeable</u>: Yesterday, I attended a rally.

<u>Explicit</u>: Macron is a terrible person who will relinquish France to the EU!
<u>Implicit</u>: There's no way I could vote for someone with Macron's lack of morals.

Concerns

Concerns are broader societal issues on which there are multiple, competing and often incompatible positions. You can also think of concerns as "topics".

You can usually identify a concern by the presence of particular terms or keywords. We've listed possible keywords for each concern below; note that this is not an exhaustive list - you may find relevant documents that do not have keywords in our list.

Also note that you should consider the keyword in context - just because a keyword appears in the tweet does not necessarily mean it is relevant for a given concern. (For example, the word "job" is listed in the economy concern, but a tweet that says "Yesterday, I started a new job" wouldn't be relevant.)

List of concerns:
3.1 Economy (incl. economic growth, unemployment)
This concern category will include messages that relate to the economy, including economic growth, unemployment, major investment projects, etc.
   Possible keywords: national debt, economy, government spending, jobs, inflation, banks, interest rates, prices

3.2 Terrorism and counterterrorism
Includes messages that relate to terrorism and counter-terrorism efforts. Other references to terrorism will also be included here (e.g., accusing candidates of being terrorists or supporting terrorists, etc.).
    Possible keywords: terrorism, terrorist, terrorist attack, counter-terrorism, Jihad, bomb, extremism

3.3 Religion
Includes content broadly related to religious topics - this may include commentary on specific religions or religious practices, religious beliefs of candidates or voters, future of religion in France, etc. Note that discussions of Islamic terrorism combine Religion (3.3) with Terrorism (3.2). These are two concerns that we have observed as frequently co-occurring in the dataset.
   Possible keywords: religious, religion, Christian, Jesus, Jews, Judaism, Muslim, Islam, Church, Mosque, Pope, Catholicism, atheism, God

3.4. Immigration and refugees
Includes content related to legal and illegal immigration (including immigration processes, asylum, etc) as well as commentary on refugees and immigrants themselves.
   Possible keywords: immigration, migration, refugees, asylum, illegal alien, foreigner, Syrian, Arab, Algerian, Tunisian, visa, citizenship, deport

3.5 International alliance organizations

Content related to the role of and membership in various international alliance organizations, including European Union, the Schengen Area, NATO, etc. This category is not designed to capture bilateral agreements or French relationships with individual countries (e.g., French-US relations would not be included here.) Possible keywords: EU, NATO, alliance, Frexit, Brexit

3.6 Relationship with Russia

Content related to France's relationship with Russia, Russian interference in French affairs, Russian financial institutions or other companies, and/or individual Russian political figures (e.g., Putin). This would also include concerns about Russian influence operations, hacking attributed to Russian entities, etc. (However, posts or commentary that simply discuss influence operations or election interference that do not mention Russia or Russian entities would not be considered relevant.)
   Possible keywords: Putin, Russia, Kremlin, Russian

3.7 National Identity and national pride

Expressions of French national identity and national pride; this also includes commentary on the value of French culture.
   Possible keywords: tradition, true French, proud (to be French), nationalism

3.8 Environment and climate change

Includes discussions on the environment, issues related to climate change, pollution, environmental regulation, etc. In some cases, this may include natural disasters (but only in cases where natural disasters are associated with climate change.)
   Possible keywords: climate, climate change, environment, pollution, greenhouse gas, fossil fuels

3.10 Fake news/misinformation

*This refers to **messages and commentary <u>about</u>** (supposed or actual) fake news and misleading information. This will also include accusations that someone is spreading fake news/misinformation or dismissal of another message or issue as "fake news." (This label is not intended to be used to identify fake news itself, but rather concerns about fake news.) Accusations that an author is not who they purport to be (e.g., that the author is a bot) will also fall in this category.*
   Possible keywords: fake news, disinformation, misinformation, malinformation, bot, troll farm, hoax, manipulation

3.11 Character of ENTITY (candidates, other key figures)

This includes commentary on the character of a particular entity. WIthin the context of the French election, this will primarily be the political candidates themselves and other key figures. Discussions of scandals will also fall under this category.
   Possible keywords: scandal, good, decent, honest, immoral, ethical….

Emotions

There are many different ways of defining and identifying emotions in text. For this project we will be focusing on emotion EXPRESSED in a text by the author (not necessarily the same as the emotion evoked in a reader). Again, humans are good at making very fine distinctions between different emotions but we need to focus on a smaller, identifiable list of emotions, and look for explicit features of the text which help to identify them.

When annotating for emotion, take into account the entire text, including any emojis, hashtags, etc. that may impact how the text is interpreted. For example, a hashtag or emoji may indicate that an explicitly expressed emotion is in fact being used ironically :

"Such a shame that I didn't blah blah blah. #sorrynotsorry"

Note that the emotions expressed should be the emotions of the author, not a third person they are referring to. For example: "People hate Taylor Swift so much. I don't know why." should NOT be coded as expressing hate.

List of emotions:
4.1 Anger, hate, contempt, disgust
Lexical items which indicate this cluster of emotions will often be used as descriptions of the entity which is hated: *vile, awful, nasty, sickening, revolting,* etc. The emotion may also be explicitly indicated in the text, e.g. *"I hate…" "I am disgusted by…" "I can feel only contempt for…".* Angry or disgusted emoji may also occur. (For example, "*Marine Le Pen makes me* 🤮").

4.2 Embarrassment, guilt, shame, sadness
In addition to lexical items which directly attribute the emotion to the author ("*I am sad/I am embarrassed/I am in mourning…*"), these emotions may be expressed through descriptions of an event or entity (*"a tragedy", "an embarrassing situation"*) or by an explicit acceptance of guilt or blame (*"I take full responsibility"/"I deeply apologize for my actions"*)

4.3 Admiration, love
This grouping includes messages that express admiration, love, attachment, respect, caring, etc. towards a particular entity. Emoji may also be used to express admiration and love, e.g., 😍, 💖, ❤️

4.4 Optimism, hope
This grouping includes message that express optimism and hope. As such, these messages are often future-oriented, and express a hope that good things are coming. Within the context of an election, optimism and hope is often expressed as tied to an election outcome (e.g., "Vote for Macron and he will open the door to a great future for France!")

Keywords that may indicate optimism and hope include hope, optimism, future, promising, can't wait,. 🤞, etc.

4.5 Joy, happiness
Joy and happiness generally refer to the author's current positive emotional state. *"I am thrilled", "This is wonderful", "Great news!", "I am so glad that…"*

Note that some positive emotions that often co-occur with happiness would by themselves be Positive-other, e.g. *relief, interest, satisfaction*.

There are also some lexical items that commonly co-occur with joy and happiness that do not directly express emotions, e.g. *congratulations.*

4.6 Pride, incl. national pride
This cluster includes individual, group, and/or national pride. The message may express pride in an individual's own accomplishments or a group to which the individual belongs, or it may express pride for the action of another individual. Within the context of an election, pride is often associated with national identity (e.g., "France is a great country with a magnificent history!")

Keywords that may indicate pride include pride, proud, patriot, patrie, fatherland, or phrases like "Liberté, égalité, fraternité," etc. You will also see general positive adjectives (e.g., great, magnificent) in combination with "France," "French people," etc.

Some emoji that may indicate expressions of French pride: , 🥐, 🥖, 🍷, etc. Note that the French flag emoji sometimes looks like a French flag or it may show up as the letters FR.

4.7 Fear, pessimism
This emotion grouping is the inverse of optimism and hope, and so tends to be similarly oriented towards the future and express concern that negative events are coming. In addition to the direct attribution of an emotion to the author (*I am afraid, frightened, concerned, worried*), fear and pessimism in this scenario may be indicated through negative predictions tied to an election outcome: *"It would be a disaster for all of us", "France cannot survive four years of his leadership"*, etc.

4.8 Amusement (often expressed with sarcasm)
Amusement in tweets will frequently be expressed with emoji (😂😂😂😂) as well as "*it's hilarious/ridiculous/funny*", etc. Sarcasm is not itself an expression of emotion, but in many cases may accompany amusement.

4.9 Positive-other
Any expression of a generally positive emotion which does not fit into another category above; for example: *relief, support, loyalty*; "*This is nice/good/enjoyable.*"

4.10 Negative-other
The negative-other cluster is intended to include any other emotions with a negative valence that aren't included in the categories above. This may include fatigue, general distaste, suspicion, jealousy, boredom, loneliness, and/or resignation. You can also use this category for messages that feel like they express a negative emotion, but for which it is difficult to tell what that emotion is.