The annotation guidelines

TODO: There are currently two sheets - the first one contains the current (full) version of the corpus, while the second one contains tweets that had been wrongly annotated at first (they contain up to 280 characters but the text is limited to 140). Metadata and several other categories remain valid and are left intact. Contributions are more than welcome.

GENERAL:
  A. Try to leave your own political / social / cultural opinions aside and aim for maximum objectivity, i.e., focus on the tweet itself (however, some political knowledge is required to detect sarcasm).
  B. If unsure about the verdict, highlight the tweet or just leave it unannotated. Anyway, it's only beneficial to have more than one opinion, so any input is more than welcome! This also applies to adding your own polarity verdict to already annotated entries.
  C. The first part (some 1500 tweets) needs heavy editing (copying and pasting the rest of the tweet) — it took me one week to notice that Twitter changed the max. tweet length back in 2017.
  D. To maintain some kind of order in the rapidly growing table, empty cells are padded with N/A.
  E. Any category can contain more than one label. For example, sometimes it is perfectly appropriate to assign the labels 'agitated' and 'sarcasm' to a given tweet.
      a. example: EAnd I'm sorry that you are and for ever will be an utter twat who thinks wheat is a bit naughty #TheresaMay #Bellend
  F. If you notice that an unannotated tweet is either completely unrelated to politics or is simply a quoted headline (or in any other way doesn't fit the database), delete it and leave an appropriate comment.
  G. The focus is on political tweets from the English-speaking countries (first official language, that is). Currently the corpus includes tweets related to political affairs in the USA, the UK, Australia, New Zealand, and Canada.
  H. To make the corpus clear and easy to navigate, please use the following formatting guidelines:
      a. font: Helvetica Neue
      b. size: 10
      c. borders: both horizontal and vertical
      d. horizontal alignment: centre
      e. vertical alignment: middle
      f. text wrapping: wrap

I. Finally, the guidelines are just that — guidelines. In practice, it is often quite difficult to make a decision based solely on these few rules.

POLARITY

A. To keep things simple(r), we agreed on discrete polarity, i.e., 1 — positive, 0 — neutral, -1 — negative. Bear in mind that polarity may (and often is) distinct from mood (see below).

B. If the tweet contains, more or less, an equal amount of positive and negative words, assign the value 0 to it.

    a. example: @GeorgeAylett Exactly why a Tory biased media try to keep him off our screens. Far more TV appearances needed. @jeremycorbyn always comes across as truly genuine & can answer all questions honestly & in a way we can understand, without tripping up like May does over her lies #JC4PM #Corbyn

MOOD

A. This is perhaps the most arbitrary category. We work with the following labels:

● neutral — self-explanatory, it most often mirrors the neutral polarity

    ○ example: Remove private and religious school federal funding, pour the money into public schools and Universities. Tax all religions and use the funds to improve our hospitals. #auspol

● agitated — the most popular label; signifies anything from slight dissatisfaction to positive, yet emotionally charged (e.g., punctuation, capitalisation, emoticons, etc.) tweets

    ○ example: @chuckwoolery Trudeau is a disgrace... #TrudeauMustGo #MCGA

● anger — somewhat rare; describes tweets that convey definite anger, much stronger than 'agitated'

    ○ example: You bitch Bishop! We didn't sign the Paris Accord, you and Talkbull signed it for your own glory. Now we want OUT, do you understand that word. God, what a dimwit. @LiberalAus #auspol

● joy — describes any sort of a positive message; perhaps slightly inaccurate because 'positive' doesn't always imply 'joy', but we want to avoid introducing too many labels

    ○ example: @theresa_may #TheresaMay may not give a crap but I am so so proud of #England #ENGLANDVSCOLOMBIA epic fantastic superb so so proud 🖕⚽🖕⚽

● sarcasm — a popular category; the label is an umbrella term for irony and sarcasm

    ○ example: #TheresaMay today we publish our #LGBTActionPlan where we firmly establish the Tory

*cough
..party as the defender of
*cough
#LGBT rights and
*cough
What, what do you want...?
Sorry PM but do you remember those Irish people you gave £1bill to?
Yes....Oh shit

METAPHORS & MULTIWORD EXPRESSIONS

    A. Metaphors are annotated in accordance with the conceptual metaphor idea. For a good introduction consult: Lakoff, G. and Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

        a. Example: *Love is a journey*. Here, *journey* is the **source domain** and *love* is the **target domain**.

    B. Multiword expressions is perhaps a more controversial category and includes (but is not limited to) idioms and proverbs.

        a. To annotate idioms, use the canonical (default) form of the idiom. Consult the following dictionary to cite the canonical form: Sinclair, J. (1995). The COBUILD Dictionary of Idioms. Glasgow: Harper Collins. If a given idiom is not present in the aforementioned dictionary (new expressions emerging due to a natural linguistic process), consult the following dictionaries (in this order) and cite the form given therein: https://dictionary.cambridge.org and https://idioms.thefreedictionary.com.

        b. To annotate proverbs, use the canonical (default) form of the proverb. Consult the following dictionary to cite the canonical form: Simpson, J. and Speake, J. (2008). The Oxford Dictionary of Proverbs (5 ed.). Oxford: Oxford University Press. Available online:

            i. http://www.oxfordreference.com/view/10.1093/acref/9780199539536.001.0001/acref-9780199539536

KEYWORDS & OTHER

    A. The 'Keywords' field is supposed to contain words / phrases that helped you make the polarity / mood decision. Sometimes it can be very difficult to pinpoint the exact words. In that case, just leave the field blank, or better yet, put N/A there. There are also two (self-explanatory) additional categories, namely: `<emoticon>' and `<punctuation>'. Mind you, <punctuation> serves here as an umbrella term for both punctuation and capitalisation. Perhaps not a neat solution for pedantic linguists, but I assume it'll be better for computational purposes.

        a. Cite the keywords verbatim. However, if there is a spelling mistake, correct the word(s) by using square brackets. For instance, the string `two terrible mens' would be `two terrible m[en]'.

    B. Finally, the `Other' category is precisely that — a vague field for anything that doesn't fit the other fields. Remarks related to figurative language shall also be noted there (for instance, proverb or idiom modifications).

GENDER & BIAS

A. If possible, try to annotate the author's gender, i.e., by looking at their profile picture, profile description, and name. It usually works fine. If in doubt, put N/A.

B. `Bias' refers to political bias. We use two labels for simplicity: left- and right-wing. In the vast majority of cases, it is really easy to decide what to put there. Consult the Appendix below for a list of the most popular politicians and their political orientation.

THEME & POLITICIAN

A. `Theme` refers to a political theme (based on the hashtags and content). Please, use the following labels (unless a new political event is significant enough to merit its own label, e.g., `Brexit'):

    a. US politics: `Republicans vs Democrats'

    b. UK politics: `Tories vs Labour' or `Brexit'

    c. Australian politics: `Australian politics'

    d. New Zealand politics: `New Zealand politics'

    e. Canadian politics: `Canadian politics'

B. `Politician': It's meant to filter out tweets by famous / important politicians. I plan to get it done once in a while, simply by searching the author category with a list of usernames of the leading politicians.

SWEAR WORDS / OFFENSIVE LANGUAGE

A. I assumed a rather `conservative' approach to this category to ensure more varied results, that is to say: I even put there such words as `stupid' or `idiot', even though I wouldn't really consider them offensive. This approach seems to be more in line with the message conveyed by most tweets — when people use those words in tweets, they are (usually) really angry and want to insult the addressee.

    a. An important note: these words / phrases are put in this category as a lemma. For example: instead of `fucks', type `fuck'.

    b. If the swear word is misspelled, abbreviated, or even semantically weakened, put the full form anyway. For instance: `What the actual f***' -> `fuck'. Hashtags also count as potential swear words / offensive language.

    c. The `poo Emoji' (💩) should be interpreted as the word `shit'.

        i. Example: It is highly unlikely that Calgarian's will be throwing roses at #JustinTrudeau should he go. Other things perhaps! ((💩)) #Trudeau may need more "panic attack time off" I'm sure. Best he just #Resign. 🤣👍🟧 #CalgaryStampede #TheGreatAwakening #MCGA https://t.co/FlmNJyBrgy

SPEECH ACT

A. The following labels are used:

- assertive — for example: suggesting, describing a situation, swearing, boasting, etc. (extremely common!)
    - example: It's Shocking, & disgusting this government wants people to work longer then their parents did (work till you die) Depriving young people of jobs & old people of retirement & #pensions #shame #TheresaMay

- directive — trying to make sb perform an action
    - example: @JustinTrudeau RESIGN!!! Do the honourable thing, RESIGN!!! #downwithtrudeau #trudeaumustgo #groping

- commissive — the speaker commits to doing sth in the future (e.g., `I'm going to vote on X')
    - example: @JustinTrudeau #TrudeauMustGo we will be voting him out in 2019!! He's a joke

- expressive — how the speaker feels about a particular situation
    - example: Even in Central London, I'm still scared to hold my soulmate's hand. #oneday #Pride2018 #TheresaMay https://t.co/wbiZ42gSLT

- declarations — change the state of the world in an immediate way (extremely rare in political tweets; examples: `I declare war on country X', `I pronounce you husband/wife and husband/wife', 'I baptise you in the name of X', etc.)
    - example: so far no tweets with this label

- metalocutionary — (somewhat rare) refers to metalinguistic discourse; example: `I think the word "preposterous" will soon disappear from the English language.'
    - example: @TrueCanadaFirst We'll replace it with the more appropriate expression and less offensive one: #Trudeau. Instead of saying: you're so crazy, #peoplekind should sayn instead: You are a real Trudeau 😂😂😂🤣😹😹#trudeaumustgo #trudeaugrope #trudeaufortreason

APPENDIX

the USA:
1. Donald Trump: right-wing
2. Hillary Clinton: left-wing
3. Barack Obama: left-wing
4. Bernie Sanders: left-wing
5. Gary Johnson [libertarian]: more or less right-wing
6. Maxine Waters: left-wing
7. Nancy Pelosi: left-wing
8. Republicans: right-wing
9. Democrats: left-wing

Canada:
1. Justin Trudeau: left-wing
2. Pierre Trudeau (PET): left-wing
3. Andrew Scheer: right-wing
4. John Tory: right-wing

the UK:
1. Theresa May: right-wing
2. Jacob Rees-Mogg: right-wing
3. Jeremy Corbyn: left-wing
4. Boris Johnson: right-wing
5. David Cameron: right-wing
6. Gordon Brown: left-wing
7. Tony Blair: left-wing
8. Iain Duncan Smith (IDS): right-wing
9. Philip Hammond: right-wing
10. George Osborne: right-wing
11. Labour Party: left-wing
12. Conservative Party: right-wing
13. Green Party: left-wing

Australia:
1. Pauline Hanson: right-wing
2. Malcolm Turnbull: right-wing
3. Tony Abbott: right-wing
4. Julie Bishop: right-wing

5. Kevin Rudd: left-wing
6. Bill Shorten: left-wing
7. David Leyonhjelm [libertarian]: more or less right-wing

New Zealand:
1. Phil Goff: left-wing
2. Jonathan Coleman: right-wing
3. Helen Clark: left-wing
4. Don Brash: right-wing
5. Winston Peters: right-wing