

Annotation Guidelines of  
“Hate Towards the Political Opponent:  
A Twitter Corpus Study of the 2020 US Elections  
on the Basis of  
Offensive Speech and Stance Detection”

Lara Grimmering

The data set stemming from this annotation study is described  
in detail in the following paper:

Lara Grimmering, Roman Klinger: Hate Towards the Political Opponent:  
A Twitter Corpus Study of the 2020 US Elections  
on the Basis of Offensive Speech and Stance Detection.  
WASSA at EACL 2021.

# 1 Annotation Tasks

**Warning: This file contains offensive and hateful content.**

We have collected tweets about the 2020 US presidential campaigns and election. Given the text of a tweet, we want to annotate the stance the tweet holds towards our pre-determined targets and the presence or lack of hateful and offensive speech.

The file has 5 columns:

1. text
2. Trump
3. Biden
4. West
5. HOF

## 1.1 Stance Detection

Given the text of a tweet, we want to find out which position the text holds towards our pre-determined targets, Donald Trump, Joe Biden and Kanye West. The detected stance can be one of the following annotation labels:

- *Favor*:  
Text argues in favor of the target
- *Against*:  
Text argues against the target
- *Neither*:  
Target is not mentioned; neither implicitly nor explicitly
- *Mixed*:  
Text mentions positive as well as negative aspects about the target
- *Neutral mentions*:  
Text states facts or recites quotes; can not be said for sure, whether text holds any position towards the target; hence, there is no judgment of the target

## 1.2 Guidelines

Each text of each tweet needs to be rated with one of the named labels, the default value is *Neither*.

The annotation of tweets for whether the text is *favorable*, *against*, *neither*, *mixed* or *neutral* towards the target is also executed on party level, which means that “voteBlue” represents a tweet not only being in favor of the Democratic Party but also of Joe Biden.

Likewise, mentions of the Democratic or Republican Party, such as “(@)TheDemocrats” or “(@)GOP” can represent a tweet being *favorable*, *against*, *mixed* or *neutral* towards Biden and/ or Trump if the content of the tweet is about the electoral process, a discussion of the candidate’s vision, the election manifesto etc.

To label the text of the tweets with *Favor*, *Against*, *Mixed*, *Neutral mentions*, the targets need to be named in the tweet. Targets can be named explicitly like “Trump” or “Biden” or implicitly like “president”, “POTUS” and “voteRed” for Trump or “Democratic presidential candidate” and “voteBlue” for Biden.

The slogans of the respective candidates such as “MAGA”, “KAG” or “Build-BackBetter” should come together with an explicit or implicit reference to the respective candidate. Then, the tweet can be labeled *Favor*, *Against*, *Mixed* or *Neutral mentions* towards the target.

Exception: The official campaign slogans of the presidential candidates’ websites, “MAGA2020”, “BattleForTheSoulOfTheNation” and “2020Vision”, respectively, already represent implicit references to the targets.

The mention of the vice presidential candidates in the tweets can also indicate the position of the text towards our targets.

@\_mentions of the targets or the vice presidential candidates also need to be taken in consideration when labeling the text of a tweet.

Hashtags like “#BackTheBlue” and “#BlueLivesMatter” do not refer to the Democrats but to a countermovement to “BlackLivesMatter”.

**Examples:**

- “Vote for Harris”:  
**Neither Trump, Favor Biden, Neither West**
- “Vote for Biden, vote Trump out”:  
**Against Trump, Favor Biden, Neither West**
- “FBI agents back Christopher Wray in letters to Trump, Biden and warn that firing him could ‘damage’ bureau via @Yahoo”:  
**Neutral mentions Trump, Neutral mentions Biden, Neither West**
- “@realDonaldTrump Cuz we are smart here. Now give Ivanka your phone and go back to bed. #BidenHarris2020 #ByeDon2020 #VoteThemOut #PennsylvaniaForBiden #TyphoidDonny #VoteBlueToEndThisNightmare #TrumpIsANationalDisgrace #Resist”:  
**Against Trump, Favor Biden, Neither West**
- “Kamala Harris said she believed Biden’s sexual assault accuser. Both parties are riddled with contradictions, but the Democrats seem to be more violent to those they disagree with”:  
**Mixed Trump, Mixed Biden, Neither West**
- “Why send new ballots? Just send corrected return envelopes!!! It is as if @TheDemocrats WANT millions of extra ballots out their to aid in their voter fraud scheme!! #MAGA #TRUMP2020 #LiberalTears”:  
**Favor Trump, Against Biden, Neither West**
- “That selfish idiot just put the Secret Service at risk for a photo op. Vote him out! #BidenHarris2020”:  
**Against Trump, Favor Biden, Neither West**

### 1.3 Hate Speech Detection

At the same time, we want to annotate tweets for hateful and offensive speech (HOF). The column HOF can be annotated as *Hateful* or *Non-Hateful*. The default value is *Non-Hateful*.

We follow the definition of Gao and Huang [2017] who define hate speech

“to be the language which explicitly or implicitly threatens or de-means a person or a group based upon a facet of their identity such as gender, ethnicity, or sexual orientation.”

We do not distinguish between offensive and hateful speech. Further, we adapted this definition slightly to our setting (tweets about the presidential election of 2020) and annotate name-calling and down talking of the political opponent as hateful and offensive.

### 1.4 Guidelines

Hateful and offensive speech includes

- Abusive speech
- Degrading speech
- Violent threats
- Insults
- Wishing harm on a person and/or group of people
- Threatening with violent acts
- Racial and sexist slur

Calling names, such as “Slow/ Sleepy Joe”, “Phony/ Monster Kamala”, “TrumpCrimeFamily”, “PutinsPuppet”, “DontheCon” is also considered offensive.

Think of offensive and hateful abbreviations, such as SOB (= son of a b\*tch), BS (= bullsh\*t), POS (= piece of sh\*t).

Words can become offensive/ hateful when used in combination with offensive/ hateful adjectives.

Be aware of slang! N\*gga and b\*tch are not always meant in a hateful way.

When in doubt ask yourself: “Would I feel offended by the tweet?”

### Examples:

- “Sleepy/ Slow Joe”:  
**Hateful**
- “Phony Kamala”:  
**Hateful**
- “Joe Biden is a pedophile and Kamala Harris is a whore”:  
**Hateful**
- “Go say a pray for PRESIDENT-ELECT JOE BIDEN AND VICE PRESIDENT ELECT KAMALA HARRIS OH YEAH. F###k god”:  
**Non-Hateful**
- “But God willing, Trump won’t be! Maybe the Taliban can take him and his crime family in when they are evicted from the White House”:  
**Hateful**
- “Karens”:  
**Hateful**
- “Trump is a national disgrace”:  
**Hateful**
- “I do not believe a liar”:  
**Non-Hateful**
- “I do not believe a pathological liar”:  
**Hateful**
- “Jim Crow Joe is an old senile softcore racist.”:  
**Hateful**

### References

Lei Gao and Ruihong Huang. Detecting online hate speech using context aware models. In *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017*, pages 260–266, Varna, Bulgaria, September 2017. INCOMA Ltd. doi: 10.26615/978-954-452-049-6\_036. URL [https://doi.org/10.26615/978-954-452-049-6\\_036](https://doi.org/10.26615/978-954-452-049-6_036).