Bachelorscriptie 2018/2019 - Abusive Language

DATA ANNOTATION

ABUSIVE LANGUAGE: impolite, harsh, or hurtful language (that may contain profanities or vulgar language) that result in a debasement, harassment, threat, or aggression of an individual or a (social) group, but not necessarily of an entity, an institution, an organisation, or a concept.

Our annotation will be conducted using two attribute sets:
- EXPLICITNESS: EXPLICIT | IMPLICIT | NOT
- TARGET: INDIVIDUAL | GROUP | OTHER

Explicitness: the focus is on the content of the message. It takes into account how the message is realised. The level of annotation focuses on the assumed intentions of the users (is the message debasing someone?) as wells the effect of the receivers (can the message be perceived as debasing by a targeted individual or a community?). Explicitness is measures by referring to the presence of profanities, slurs, offensive terms.

Target: the focus of this attribute is on the (potential) receiver of the message. It makes explicit to whom the message is "addressed to".
We distinguish between two major types of targets:
- INDIVIDUAL : this value is used when the target of a message is a specific individual
- GROUP: this value is used when the target of the message is a (social) group of people; and
- OTHER: this value is used when the target of the message are concepts, institutions and organisations, or non-living entities.

ANNOTATION GUIDELINES

Tweets to be considered for the annotation:
- The message is not a retweet:
- The whole message is not a quotation of someone else
- The message is not a meme or simply a link to an external website

The first attribute to annotate concern EXPLICITNESS:

- A message is marked as EXPLICIT if it is interpreted as potentially abusive (intension of the speaker to debase/offend; effect on the message on the receiver) if it contains a profanity or a slur.

A. *Liberalen zijn idioten. :* EXPLICIT
B. *Islam is onzin.* EXPLICIT

- A message is marked as IMPLICIT if it is interpreted as potentially abusive (intension of the speaker to debase/offend; effect on the message on the receiver) if it DOES NOT contain a profanity or a slur.

C. *Minder minder Marokkanen* : IMPLICIT
D. *Europa heeft een plan om blanke mensen te vervangen door mensen uit Afrika.* IMPLICIT

- A message is marked as NOT if it is interpreted as being potentially NOT abusive (no intension of the speaker to debase/offend; no effect on the message on the receiver). We do not consider as ABUSIVE messages that debase or offend the author of the message (e.g. messages at the first singular or plural person)

E. *Ik ben een idioot* : NOT - the intension to offend is from the author to him/herself
F. W*ij experts zijn soms zo dom*: NOT - the author of the message puts himself as part of a group and s/he is claiming something about the group of people including her/himself.

NOTE: a message can contain a profanity or a slur but NOT being abusive:

G. *Godverdomme:* NOT - contains a profanity, but there is not intension to debase or offend someone, nor the effect of the message can be perceived as debasing or being offensive.

Similarly for the following cases:

H. *Wat een kutstreek! :* NOT
I. *Wat een shitdag!:* NOT
J. *Wat een klotezooi!:* NOT
K. *Wat een zeikerige voetbalwedstrijd.:* NOT
L. *Dat vind ik verdomd vervelend!:* NOT
M. *Ik ben het spuugzat!:* NOT

IF IN DOUBT, CHECK IF THERE IS A TARGET, OR DISCARD THE MESSAGE.

The second attribute to annotate concern the TARGET.

NOTE: a message CANNOT be ABUSIVE and NOT HAVE a target.
N. *Wat een shitdag!:* NOT

Annotate this attribute AFTER you have annotated the EXPLICITNESS attribute. In case of doubts, you can check if there is a target of in the message.

- A target is marked as INDIVIDUAL if the message is ABUSIVE and the target of the abuse is an individual (e.g. a specific person)

O. *hij is een idioot :* INDIVIDUAL

- A target is marked as GROUP if the message is ABUSIVE and the target of the abuse is a (social) group (e.g. an ethnic or religious minority)

P. *Liberalen zijn idioten. :* GROUP
Q. *Conservatieven zijn corrupt. :* GROUP
R. *Minder minder Marokkanen* : GROUP

- A target is marked as OTHER if the message is ABUSIVE and the target of the abuse is an organisation, an institution, or a concept

S. *Europa is een deken van fascisten. :* NOT
T. *Europa heeft een plan om blanke mensen te vervangen door mensen uit Afrika.* OTHER
U. *Religie is onzin. :* NOT
V. *Islam is onzin.* OTHER
W. *Het katholicisme is onzin* OTHER
X. *De VVD is corrupt.*: NOT