Corpus of Romanian Sexist and Offensive Tweets

Annotation guideliness

Author: Diana Höfels
diana-constantina.hoefels@student-uni.tuebingen.de

Under the supervision of: Dr. Çagri Cöltekin
ccoltekin@uni-tuebingen.de

Eberhard Karls University of Tübingen

Introduction

This document represents a guide that presents aspects of the sexism annotation task with the purpose of creating a **corpus of sexist and offensive language in Romanian.**

Content disclaimer Please pay special attention to the following information. This guide contains examples that include offensive language, forms of expression that incites, promotes or justifies hatred, violence and discrimination against a person or group of people for various reasons.
Similarly, we warn you that the text data prepared for the annotation task, namely tweets from the social networking platform Twitter, may contain violent, offensive, brutal language, insults, threats, words that incite violence, degrading, humiliating or discriminatory expressions. Please be mindful of the fact that the sole goal of this task is to build a corpus, and do not use the information in the data collected to incite or spread hatred of any type and form towards Twitter users, existing in this data set.

Offensive language is comprised of swearing, cursing, dirty, obscene, insulting words, or slander. Some of these offensive words are directed at a single gender and are mainly used to signal behaviors that are not in line with the ideas of masculinity or femininity of a society (Kremin, 2017).

Sexism is defined as a **prejudice or discrimination based on sex, especially discrimination against women**[1] and it is found in all areas of life.

1 Task Description

The task proposes the annotation of sexist and offensive content in the Romanian language from Twitter. Twitter is a public, online micro-blogging social media platform where users can post up to 280-character of content, known as tweets. The participants of this task are asked to mark the presence of sexist and offensive content by mapping labels to a given tweet.

The following are some of the main aspects of the annotation guidelines:
- the annotators should be as objective as they can.
- the annotators must provide annotation to Romanian Twitter posts by discerning the topic, identifying the intention and the sentiment within it, and finally classifying it based on a predetermined list of labels.
- the global meaning of a tweet must be analysed by considering all the elements of the tweet (@mentions, #hashtags, emojis), as well as irony, sarcasm, slang and colloquialism.
- the task requires annotators to assign labels to each tweet based on whether sexist or offensive content is present, as follows: if a tweet does not contain any sexist elements, its content is then further assessed to determine if the content is offensive. A tweet with sexist elements, however, will be further assessed to identify the type of sexism, specifically if the sexist content directly targets

---

[1]acc.Merriam-Webster

one or more women, or describes a woman, a group of women or all women, or it is a report of a sexual discrimination experience, and label it accordingly. For each of the cases described above, **??** shows the steps and their corresponding labels.
- we do not allow double-labeling of tweets that contain both sexist and offensive language; we encourage the annotator instead to determine which type of language the tweet leans most toward.
- the annotator should use the label 'cannot decide' whenever you are faced with a situation in which you doubt the existence of sexist elements, instead of an annotation with which you are not confident.
- despite the fact that this study focuses on sexism directed at women, we encourage annotators to also mark sexism directed at men;

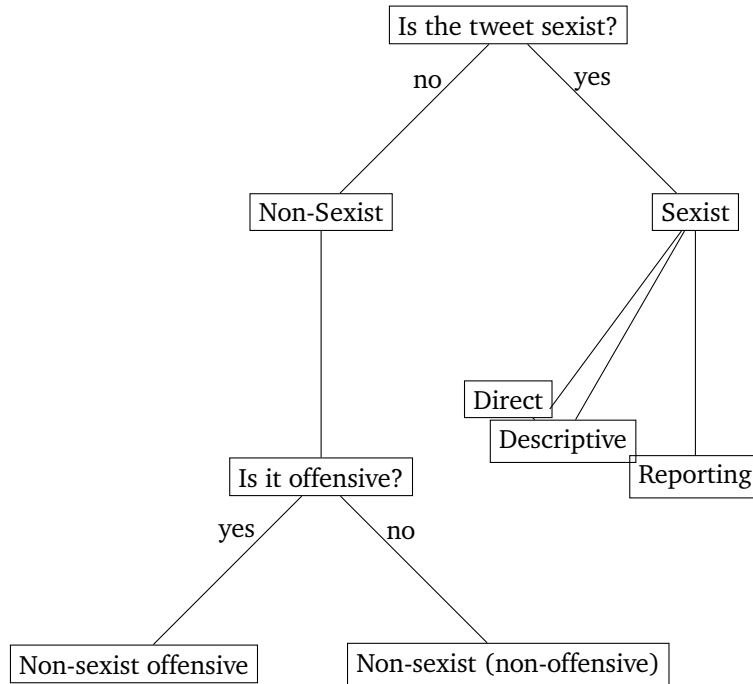View Figure 1 to learn more about the data annotation workflow.

Figure 1: Data annotation workflow

2 DATASET DESCRIPTION

Data is collected from Twitter in real-time for a time period of 6 weeks, from 18.05.2021 until 01.07.2021. It contains a random selection of 40.000 tweets, and each tweet will be annotated three times by three different annotators.

A tweet may contain:
  (a) @Mentions: used to address another Twitter account;
  (b) #Hashtags: categorizing content about a topic;
  (c) Emoji or Emojis.

We encourage that all elements of a tweet to be considered in the annotation process as they may provide information about the topic, context, sentiment, and overall meaning of the tweet.

3 CORPUS LABELLING CONVENTIONS

This section aims at describing the taxonomy of the annotation task.

The annotation taxonomy consists of two levels. The first step is to divide content into two categories: 'sexist' or 'non sexist' and secondly, we establish sub-types of sexist and non sexist content. For sexist

content we defined three categories: 'Sexist Direct', 'Sexist Descriptive', and 'Sexist Reporting' and for non sexist content we defined two categories: 'Non sexist' and 'Non sexist offensive', with all categories being mutually exclusive. The scheme is based on typologies of sexism presented by Chiril et al. (2020) to which we add one extra category to account for offensive language. In the following, we present the annotation categories to which we add some Romanian examples to illustrate each category.

Given a tweet, we categorize it into the following labels:

1. **Non sexist**

   In this case, there are no sexist or offensive elements in the tweet, nor does the tweet carry any sexist or offensive connotations. In some instances, tweets may contain sexist or offensive elements, such as hashtags, but the overall meaning of the tweet is neither sexist nor offensive. Following are some of the most utilized sexist and offensive Twitter hashtags[2] which were included in the annotation guidelines as well: **sexist**: #sexism #feminism #feminist #equality #metoo #sexist #misandry #women #mensrights #misogyny #womenempowerment, #patriarchy #misandrist #gynocentrism#genderequality#womensrights #mensrightsactivist #maleissues #smashthepatriarchy #womensupportingwomen #doublestandards #equalrights #mensrightsactivism, #gynocentric #toxicmasculinity, #feminazi #discrimination #girls, and **offensive**: #hatespeech #hate #freespeech #racism #bullying #freedom #antisemitism #nohate.

   (a) *Campania #MeToo, adică 'și eu', a devenit o mișcare globală.*
      'The #MeToo campaign has become a global movement.'

2. **Non sexist offensive**

   A tweet is classified as 'Non sexist offensive' if it lacks any sexist connotations, but contains offensive language.

   (a) *Ce bine vă stă împreună, doi prosti amândoi.*
      'You two look good together, two fools.'

3. **Sexist direct**

   Tweets that include sexist elements and are addressed directly to women or groups of women are reported with this label. Upon examination of the examples below, we can see the presence of personal pronouns, second person singular and plural, and imperatives that indicate direct addressing:

   (a) *Tu ești femeie, de ce te bagi în discuții despre politică.*
      'You're a woman, why are you talking about politics.'
   (b) *Treci la bucătărie!*
      'Go back to the kitchen!'
   (c) *Voi femeilor, tineți-vă gura.*
      'You women, keep your mouth shut.'

4. **Sexist descriptive**

   A tweet with this label describes a woman or women in general in a sexist manner, without addressing them directly.

   (a) *Locul femeilor este la bucătărie, toată lumea știe asta.*
      'Women's place is in the kitchen, everyone knows that.'
   (b) Despre Viorica Dăncilă: *"Dacă apare o nouă bancnotă de 100 lei, va scrie pe ea una tută lei."*
      'About Viorica Dăncilă:[3] "If a new 100 lei banknote is issued, one stupid lei will be printed on it."'

5. **Sexist reporting**

   An action or an experience of sexism is reported via this label. The person reporting the act witnessed it or heard about it from other sources.

   (a) *Mi-a spus că e normal că femeile însărcinate sunt plătite mai puțin pentru că lucrează mai puțin.*
      'He told me that it is normal for pregnant women to be paid less because they work less.'

---

[2]https://ritetag.com/

[3]Former Prime Minister of Romania. During her mandate as prime minister of Romania, she was mocked for a series of mistakes that some said were proofs of her poor language skills (both English and Romanian).

(b) *Ei s-au retras recent din Convenția prin care se stipula că femeile primesc drepturi...*
'They recently withdrew from the Convention which stipulated that women receive rights...'

6. **Cannot decide**:

A tweet is designated this label when it is unintelligible, lacks context, contains sexist undertones, but is ambiguous in terms of meaning, and cannot be tagged with category 3, 4 or 5. The annotators were not encouraged to use this label if uncertain about a tweet's offensive nature. As this study examines mainly sexist language, the label was designed for tweets, that are possibly sexist, but where there is doubt about it.

4   Sexist and Offensive Language Characterisation

The purpose of this section is to provide more information on sexist and offensive language. If you already understand the concepts, you may skip this section.

4.1   Sexist Language

Sexism is a prejudice that is directed against another person's gender. The fundamental nature of men and women is connected to the roles they should play in society, and sexism is a result of these beliefs. We have outlined some of the most important forms of sexism and provided some examples below.

1. Objectification

    (a) *Om și femeie.*
        'The human being and the woman.'
    (b) *Ce frumos, e copil sau e fetiță?*
        'How beautiful, is it a child or a girl?'

    The examples, 1 (a) and 1 (b) present a mentality that objectifies women and categorizes them as a subspecies, namely women are not perceived as complete human beings nor are girls as children.

2. Inferiority/ Incapacity

    (a) *Bărbatul este capul iar femeia gâtul.*
        'The man is the head and the woman is the neck.'
    (b) *La femei cică pe lângă CIP, le bagă și senzori de parcare.*
        'In addition to a chip, women will also get parking sensors.'

    2 (a) and 2 (b) promote an outdated mindset, in which women are submissive to men, incapable of autonomy or personal power, and are often discredited for their skills.

3. Role stereotyping

    (a) *Locul femeii este la cratiță, iar rolul este să facă copii și să-i crească.*
        'Women's place is in the kitchen, to birth children and raise them.'
    (b) *Când nevasta tace, să n-o întrerupi.*
        'When the wife is silent, do not interrupt her.'
    (c) *Soacră, soacră, poamă acră.*
        'Mother-in-law, mother-in-law, you sour grapes.'
    (d) *Trebuie să te pui și tu la casa ta.*
        'You have got to settle down.'
    (e) *Numai la măritiș ți-e capul.*
        'All you think about is getting married.'
    (f) *Vei muri fată bătrână.*
        'You will end up an old maid.'
    (g) *Aia e femeie? Arată ca un bărbat.*
        'Is that a woman? She looks like a man.'

The belief in 3 (a) eliminates any other purpose for women, and binds their role in society to just motherhood. 3 (b) and 3 (c) insinuate that wives are a nuisance, and mothers-in-law are bad-tempered, unfriendly, and overbearing. Either mothers, wives, or mothers-in-law, their role is stigmatized. In 3 (e) there is an emphasis on a greater desire of women to marry and concomitantly, a greater pressure on them to settle down, as seen in 3 (d).

The marriage represents the ultimate goal, and when that goal is not met, the woman without a partner is viewed with suspicion or pity, as depicted in 3 (f). Finally, in 3 (g) we notice a gender stereotype that suggests women should always display their femininity.

4. DERAILING

   (a) *Femeile care poartă fustă scurtă merită să fie abuzate.*
       'Women who wear mini skirts deserve to be abused.'

To justify women abuse, sexually assaulted victims are partially blamed for the crime. This represents a derailing, and an attempt to absolve perpetrators from full responsibility.

5. SEXUALIZATION AND SEXUAL HARASSMENT

   (a) *Ce bună ești!*
       'You are sexy!'
   (b) *Ești o curvă.*
       'You slut.'
   (c) *Arată-mi sânii ăia frumoși.*
       'Show me those beautiful boobs.'

Women are considered sexual objects and evaluated according to their physical characteristics or sexual character. They are addressed with gendered insults and slurs, most commonly used being *bitch, cunt, slut, or whore*, and subjected to sexual advances and demands for sexual favors, as illustrated in 5 (a), 5 (b) and 5 (c).

6. HATE AND VIOLENCE

   (a) *Urăsc femeile, ar fi mai bine fără ele.*
       'I hate women, it would be better without them.'
   (b) *Petrecerea se încinge cu maneaua și femeia cu cureaua.*
       'The party is heated up by manele[4] and the woman is heated up by the belt.'

We notice an intense dislike for women, aggressive pressure or intimidation, and power assertion over women in 6 (a) and 6 (b).

7. MALE DOMINANCE

   (a) *Asta-i treabă de bărbat.*
       'This is a man's job.'
   (b) *Gigi Becali: "Femeia nu poate fi președinte!"*
       'Gigi Becali[5]: "Women cannot be presidents!"'

Male domination is ingrained in our consciousness to the extend of being imperceptible. 7 (a) and 7 (b) exhibit an androcentric and patriarchal thinking in which women are simply excluded for certain positions or roles in society.

4.1.1 Offensive Language

Generally, language can only be offensive if it is perceived as offensive by the recipient. Obviously, if the addressee considers the language offensive, then it is offensive. Our observations and the relevant literature indicate that offensive language includes profanity, blasphemy, epithets, obscenity, insults, aggression, hatred, racism, xenophobia, homophobia, antisemitism, cyberbullying, and sexism. We will exclude sexism in the classification of offensive language since for its marking we will use more elaborate annotations as described above. In what follows, we present some of the types of offensive language we considered when classifying the offensive content, and they are illustrated with examples:

---

[4]Genre of pop folk music in the Balkans, characterized by sexist lyrics and objectification of women.

[5]Romanian businessman and politician, mostly known for his ownership of a football club and his public sexist comments against women.

1. HATE SPEECH
    (i) *Bucureșteni, vă put gunoaiele? Dar de ce nu purtați masca? Vă sperie mușcătura de căpușa? De vaccin nu ați auzit? Nu mai vreți șobolani? Dar unde vă sunt pisicile?*
    'Bucharestians, do you smell garbage? But why don't you wear a mask? Are you scared of tick bites? Haven't you heard of the vaccine? Don't you want rats anymore? But where are your cats?'
2. PROFANITY
    (i) *Du-te-n mă-ta.*
    'Up yours.'
3. BLASPHEMY
    (i) *Futu-le altarul și cristelnița Sfântului Domn Iisus Hristos ...*
    'Fuck the altar and the baptismal font of the Holy Lord Jesus Christ ...'
4. HOMOPHOBIA
    (i) *Voi ăstia, lgbb lgfg, sau cum se numește comunitatea de gay și lesby, n-am nimic cu voi că sunteți oameni dar faptul că vă sugeți reciproc mă înspăimântă ...*
    'You guys, lgbb lgfg, or whatever the gay and lesbian community is called, it has nothing to do with the fact that you are human, but the fact that you suck each other scares me ...'
5. RACISM
    (i) *Să moară toti țiganii care lucrează la discord.*
    'Let all the gypsies who work at Discord die.'
6. ANTISEMITISM
    (i) *Voi sunteți chiar cea mai a dracu nație de pe Pământ, Voi evreii!! Uite așa ați creat al doilea Holocaust total, Proști sunteți, De ce vreți să ucideți în masă oamenii? .*
    'You are the most damn nation on Earth, You Jews !! This is how you created the second Holocaust, You are stupid, Why do you want to kill people en masse?'

Privacy policy

As far as it concerns the Twitter data sets utilised for this project, the content provided remains subject to the Twitter's Developer Agreement Policy, and must agree to the Twitter Terms of Service, Privacy Policy, Developer Agreement, and Developer Policy.

The purpose of the project

The **Corpus of sexist phrases in Romanian** is a project that has as the sole purpose of quantitative scientific research. To carry out this project, we annotate text, aimed at identifying sexist elements. Using the results of your answers, we build a corpus and conduct a quantitative study regarding the presence of sexism on social media platforms in Romania.

Personal Data Collection

We collect personal information from you, in general, information that could reasonably be used to identify and contact you and may include information such as:

- Contact details such as your name, email address, mobile phone number;

- Demographics, such as gender, professional status, and age.

- Survey data. When you fill out an annotation form, we collect your answers.

We do not collect financial information, data regarding racial or ethnic origin, political opinions, religious or philosophical beliefs. We may use your name for the purpose of assigning contributions to this project (mentioning your name in the *Acknowledgements* section of the scientific paper) if we have your prior consent.

How is your data collected?

We use two methods to collect data from and about you, by:

- Direct interactions

- You provide us with your email address when you fill out the annotation form.

References

P. Chiril, V. Moriceau, F. Benamara, A. Mari, G. Origgi, and M. Coulomb-Gully. An annotated corpus for sexism detection in French tweets. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 1397–1403, Marseille, France, May 2020. European Language Resources Association. ISBN 979-10-95546-34-4. URL https://www.aclweb.org/anthology/2020.lrec-1.175.

L. V. Kremin. Sexist swearing and slurs: Responses to gender-directed insults. *LingUU*, 1:18–25, 2017.

Merriam-Webster. sexism. In *Merriam-Webster.com dictionary*. URL https://www.merriam-webster.com/dictionary/sexism.