

CyberAggressionAdo-v1: a Dataset of Annotated Online Aggressions

in French Collected through a Role-playing Game

Anaïs Ollagnier¹, Elena Cabrio¹, Serena Villata¹, Catherine Blaya²

¹ Université Côte d'Azur, Inria, CNRS, I3S

² Université Côte d'Azur, CNRS, Unité de Recherche Migrations et Société (Urmis)



Introduction

Annotated scenarios mimicking cyber aggression situations that may occur among teens on private instant messaging platforms

Vision

- Proposing a new source to **explore diversity** in abusive language content and cyber aggression practices;
- Using role-playing games as a mean to collect **naturalistic interactions** relying on authentic language.

Challenges

- Creating **scenarios** addressing common **cyber aggression topics** (i.e. ethnic origin, religion, obesity and homophobia);
- Establishing a **clear taxonomy** with meaningful and theoretically sound categories (i.e. role, target, verbal abuse, hate speech and humour).

Objectives

Promote

Role-playing games carried out in French high schools to collect realistic data.

Explore

New **annotation layers** to better capture cyber aggression practices on private instant messaging platforms.

Develop

Fully-annotated dataset aiming at developing effective abusive language detection tools.

Data collection

Role-playing game setting :

- Three French high schools;
- 142 students aged (16-18 years-old, 90 girls, 52 boys);
- 45 minutes per scenario.

Topic	Scenario
Obesity	Zoe is overweight. After the gym class, Marjorie and Lucie, who are jealous of her good academic results, take a picture of her in a posture that highlights her extra pounds. They share it to the whole class with harmful comments. Natacha, a friend of Zoe, tries to defend her. She is helped by Pauline who also has a few extra pounds and is a friend of Marjorie. Julien, who was obese when he was younger, tries to intervene with Marjorie and Lucie as well as Zoe to stop the situation.

Figure: A sample of scenarios adopted in our experimentation

Data description

- 19 conversations collected;
- 172.7 messages exchanged on average.

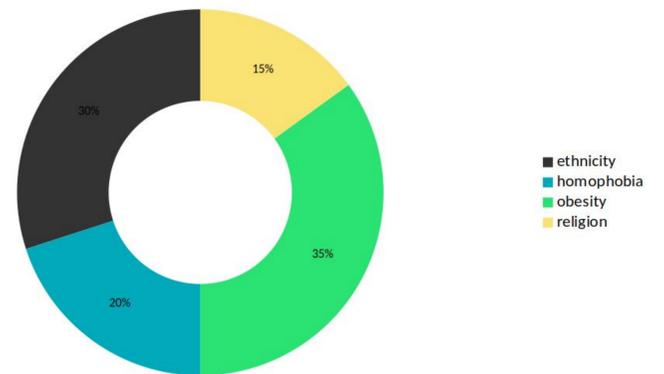


Figure: Proportion of the different topics in the dataset

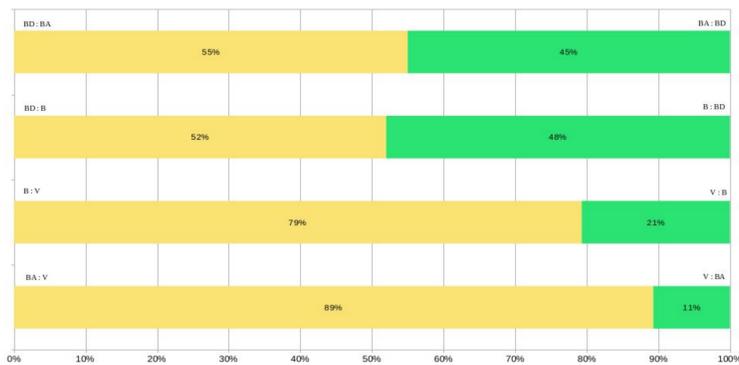


Figure: Proportion of exchanged messages between participants.

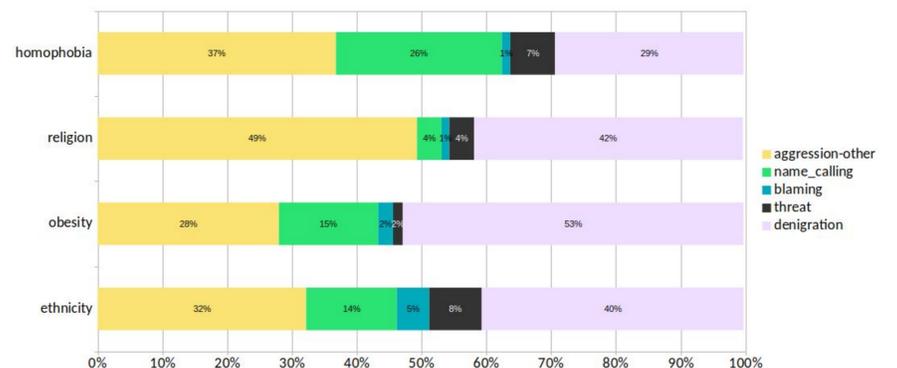


Figure: Distribution of verbal abuse targeting victim and bystander-defender across topics.

Ethical issues

Prior informed parental consent - Students' consent to participate - Students were informed on the objectives of the study and data management. They could withdraw any time with no consequence. The school administration board provided their consent prior study. In case the participants would feel unsettled, a referent in the school could provide support. Students were also provided information about resource agencies or resource adults in case of trouble.

Acknowledgements

This work is funded under the IDEX UCA OTESIA "L'intelligence artificielle au service de la prévention de la cyberviolence, du cyberharcèlement et de la haine en ligne", and by the UCA Academy 1 project with the reference number C870A021 - D103 - ACAD1_FIN_17_20Y. It has also been supported by the French government, through the 3IA Côte d'Azur Investments in the Future project managed by the National Research Agency (ANR) with the reference number ANR-19-P3IA-0002.

