

## Linguistic Cyberbullying, a Legislative Approach

**Authors :** Simona Maria Ignat

**Abstract :** Bullying online has been an increasing studied topic during the last years. Different approaches, psychological, linguistic, or computational, have been applied. To our best knowledge, a definition and a set of characteristics of phenomenon agreed internationally as a common framework are still waiting for answers. Thus, the objectives of this paper are the identification of bullying utterances on Twitter and their algorithms. This research paper is focused on the identification of words or groups of words, categorized as "utterances", with bullying effect, from Twitter platform, extracted on a set of legislative criteria. This set is the result of analysis followed by synthesis of law documents on bullying(online) from United States of America, European Union, and Ireland. The outcome is a linguistic corpus with approximatively 10,000 entries. The methods applied to the first objective have been the following. The discourse analysis has been applied in identification of keywords with bullying effect in texts from Google search engine, Images link. Transcription and anonymization have been applied on texts grouped in CL1 (Corpus linguistics 1). The keywords search method and the legislative criteria have been used for identifying bullying utterances from Twitter. The texts with at least 30 representations on Twitter have been grouped. They form the second corpus linguistics, Bullying utterances from Twitter (CL2). The entries have been identified by using the legislative criteria on the the BoW method principle. The BoW is a method of extracting words or group of words with same meaning in any context. The methods applied for reaching the second objective is the conversion of parts of speech to alphabetical and numerical symbols and writing the bullying utterances as algorithms. The converted form of parts of speech has been chosen on the criterion of relevance within bullying message. The inductive reasoning approach has been applied in sampling and identifying the algorithms. The results are groups with interchangeable elements. The outcomes convey two aspects of bullying: the form and the content or meaning. The form conveys the intentional intimidation against somebody, expressed at the level of texts by grammatical and lexical marks. This outcome has applicability in the forensic linguistics for establishing the intentionality of an action. Another outcome of form is a complex of graphemic variations essential in detecting harmful texts online. This research enriches the lexicon already known on the topic. The second aspect, the content, revealed the topics like threat, harassment, assault, or suicide. They are subcategories of a broader harmful content which is a constant concern for task forces and legislators at national and international levels. These topic - outcomes of the dataset are a valuable source of detection. The analysis of content revealed algorithms and lexicons which could be applied to other harmful contents. A third outcome of content are the conveyances of Stylistics, which is a rich source of discourse analysis of social media platforms. In conclusion, this corpus linguistics is structured on legislative criteria and could be used in various fields.

**Keywords :** corpus linguistics, cyberbullying, legislation, natural language processing, twitter

**Conference Title :** ICAL 2023 : International Conference on Applied Linguistics

**Conference Location :** Paris, France

**Conference Dates :** November 27-28, 2023