Annotating linguistic features of extreme narratives: a literature review and a proposal

Fabienne Baider and Alexandros Gregoriou

Abstract: The research questions addressed in this paper are twofold: 1. How can we define an extremist narrative? 2. Can a narrative's toxicity be determined according to certain linguistic features? To this end, we first define what is currently considered an extreme narrative, with a focus on identifying the differences between hate speech and extreme speech. We next review annotations of hate speech and extreme speech, with a specific emphasis on linguistic features; then, drawing on our results of that review we propose an annotation schema to label extreme speech. Finally, we share the results of a small sample we annotated and reveal the most frequently and most consistently annotated features that could be representative markers of extreme speech.