

## **Context and content of online hate speech**

Valentina Bianchi & Silvio Cruschina

*University of Siena, University of Helsinki*

Research on hate speech has highlighted the impact of specific aspects of the online medium (Brown 2018 a.o.). In this paper we examine the phenomenon from the perspective of formal pragmatics. Starting from (Grice 1975), the conversational exchange is conceived of as a cooperative activity involving rational agents, who add information to their common ground with the aim of distinguishing among possible ways that the world might be (Stalnaker 1978). A discourse goal is a particular way of distinguishing among these, and it is pursued through an ordered set of questions under discussion (Roberts 1998 a.o.). This model of the conversational dynamics presupposes a relatively stable conversational community, which guarantees a relatively stable common ground and a shared discourse goal. We argue that the online contexts where hate speech proliferates lack this stability (e.g. comments to a public post on a social). The conversational group is volatile: anyone can join at any moment and leave without signaling it. The original post sets a topic, but there is no discourse goal to be cooperatively pursued by all participants: the exchange is not driven by questions under discussion; participants do not aim at increasing the shared information; a discourse move is not guaranteed to get feedback. These properties characterize such a context as inherently non-cooperative. Without a proper discourse goal, discourse moves are not information-seeking, and they only retain an expressive function: the author aims at expressing their evaluation of the topic, so as to be included in (or to intrude into) a generic ‘community’ of like-minded individuals. Such moves are justified to the extent that they convey some nontrivial evaluative content. But since all participants are expected to be like-minded in their attitude towards the general topic, the evaluative content is nontrivial only if it is extreme on the relevant evaluative scale. This gives rise to excessive language which, when the attitude is negative, degenerates to hate speech. In order to test our hypothesis, we analyse interrogative clauses in a corpus (c.a. 10.000 words) of comments to posts published on the FB page of an Italian party, as compared to a reference corpus of similar size. We show that in the former, information-seeking questions are sporadic, and purely rhetorical questions outnumber those of the reference corpus to a very significant extent.