CHAPTER 9

Hate speech in non-cooperative contexts

Question types as a measure of cooperativity

Valentina Bianchi
University of Siena

Silvio Cruschina
University of Helsinki

Abstract

Research on hate speech has identified various aspects of social media that affect the speaker's attitude in this specific type of communication. In this chapter we discuss some structural aspects of the context of utterance as analysed in dynamic pragmatics, and we show that with respect to these, certain online contexts qualify as inherently non-cooperative; we hypothesise that non-cooperativity favours the emergence of excessive language and, in particular, of hate speech. To test our hypothesis, we analyse three small corpora of discussion threads from two different social platforms. We propose that different types of canonical and non-canonical

How to cite this book chapter:
Bianchi, Valentina, and Silvio Cruschina. 2024. 'Hate speech in non-cooperative contexts: Question types as a measure of cooperativity'. In *An Investigation of Hate Speech in Italian: Use, Identification and Perception*, edited by Silvio Cruschina and Chiara Gianollo, 315–346. Helsinki: Helsinki University Press. DOI: https://doi.org/10.33134/HUP-27-9.

questions are indices of (non-)cooperativity, and we analyse their distribution in each discussion thread.

Keywords: online context, cooperativity, dynamic pragmatics, question types, canonical questions, non-canonical questions

9.1 Introduction

The proliferation of hate speech and excessive language on social media has become a central issue from various disciplinary perspectives. Research since the late 2010s converges on the view that alongside social, psychological, and ideological factors, the format and affordances of the medium itself plays a crucial role (for general discussion, see Brown 2018; Baider 2020; Biri 2023; Esposito and KhosraviNik 2024). The reciprocal invisibility between the conversational participants, due to the lack of visual or auditory contact, reduces empathy on the part of the author, as well as their sense of accountability and their moral engagement; the latter may be further reduced by the author's perceived anonymity in the online environment. The speed and instantaneousness of the medium may lead to a less reflective attitude, leaving considerable space for unfiltered emotive content. Moreover, in addition to the immediate addressee, the author is also aware of a potential remote 'audience', since the medium makes the written exchange available to any reader for an undefined period: thus, hate speech is often used to mark the author's affiliation to a generic community of like-minded users.

Besides these general factors, however, something specific to the discourse context must be at play, since not all discussion threads are conducive to hate speech. In this chapter we propose that one crucial aspect is the degree of cooperativity among participants.

Following Grice (1975), the conversation is a cooperative activity involving rational agents who jointly pursue a common discourse goal. This conception has been adopted and implemented in the framework of dynamic pragmatics that evolved from the seminal work of Stalnaker ([1978] 1999), which defines the *context of the conversation* by characterising different components

and the way they are involved in cooperative speech acts (see Section 9.2).[1] We will argue that the analysis of conversation that follows from this approach presupposes the stability of the conversational group for cooperative speech acts to be successfully performed (Section 9.3). We then extend this approach to online written exchanges on social platforms, assuming that a conversation context corresponds to *a single continuous discussion thread*. We characterise a certain type of online context as lacking to a significant extent the required stability, which makes it inherently non-cooperative (Section 9.4). We then hypothesise that the lack of cooperativity favours the emergence of hate speech and that, conversely, cooperative online conversation contexts show a comparatively low incidence of hate speech (Section 9.5).

To operationalise our hypothesis, we adopt the characterisation of speech acts in formal pragmatics, and we propose that certain types of questions—information-seeking questions and deliberative questions—are evidence for context cooperativity, as opposed to rhetorical questions. On these grounds, we expect that discussion threads with a higher proportion of questions of these types show a lower incidence of hate speech as compared to discussion threads where such questions are scarce (Section 9.6). As a proof of concept, we analyse three small corpora composed of discussion threads from two social networks, Facebook and Reddit (sections 9.7 and 9.8). We compare one corpus from Facebook and one from Reddit where the discussion thread concerns gender identity, and we observe that the two corpora differ significantly, in line with our expectation; this shows that it is not the topic *per se*, however potentially divisive, that triggers hate speech. We then compare the first Facebook corpus to another Facebook corpus whose discussion threads concern potentially less divisive topics (the life of Italian immigrants in Finland): again, we observe a stark difference, which leads us to conclude that the social media platform *per se* is not crucial either: what is relevant is the specific conditions under which the online conversation develops.

1 On the connection with Grice's approach, see Stalnaker (2002: 702–705).

9.2 The dynamic pragmatic approach

In the approach starting from Stalnaker ([1978] 1999) and Lewis (1979), the conversational context is represented in the terms of possible world semantics. Each conversation is based on a set of propositions that constitute the participants' common ground: the presupposed propositions that each participant accepts as true for the purposes of the conversation, and assumes to be accepted by all participants.[2] In a non-defective context, all the participants implicitly agree on which propositions are in the common ground (Stalnaker [1978] 1999: 84–85).[3] The common ground circumscribes a region of logical space: the subset of possible worlds in which all the propositions are true, dubbed the 'context set'. The multiplicity of worlds in the context set represents the fact that the common ground information is partial and leaves various possibilities undecided.

In a typical conversation, the essential goal is to increase the information jointly accepted by all the participants. A proposition is *informative* relative to a common ground if and only if it is not true in all the worlds of the context set, but it is true in some of them (i.e. it is neither entailed by the context set, nor inconsistent with it).[4] When an informative proposition is asserted by one participant and is jointly accepted by the conversational group, it is added to the common ground and it eliminates from the context set those worlds in which it is not true. Thus, the increase in shared information reduces the region of logical space to be taken into account (Stalnaker [1978] 1999: 86).

2 The propositional attitude of acceptance, or 'common belief', is a public attitude, whereby each participant is committed to act as if the common ground propositions were true in the actual world. For an in-depth discussion of the formal properties of the relevant attitude relation, see Stalnaker (2002: 706–708).

3 'The information state will include two different kinds of information: first, information about the participants in the conversation—about what they know about each other and their common environment; second, information about the subject matter of their discourse' (Stalnaker 2018: 384).

4 See Stalnaker's ([1978] 1999: 88–89) principle I.

The increase in shared information does not proceed randomly, but is guided by discourse goals, modelled as *questions under discussion* (QUDs; Ginzburg 1996; Roberts 1996, 2018). When a question is asked, its effect is to *partition* the context set into disjoint subsets of worlds, each corresponding to a possible answer. The discourse goal, then, is to select one of these subsets.

By way of illustration, consider a toy context with just two restaurants A and B. The question in (1) partitions the context set into the four cells schematically represented in Figure 9.1.[5]

(1) Which of the two restaurants has a vegan menu?

Each cell corresponds to a complete (exhaustive) answer. It is possible, however, to provide a partial answer, by breaking down the QUD into two sub-questions:

(2) Does restaurant A have a vegan menu?

(3) Does restaurant B have a vegan menu?

A yes-answer to (2) retains only the upper half of the context set, while a no-answer retains only the lower half. Symmetrically, a yes-answer to (3) retains only the left-hand half of the context set, and a no-answer retains the right-hand one. Each of these answers is relevant to the super-question (1)—since it discards some cells from the context set—but it is partial in that it does not single out a unique cell (Roberts 2012: 11–12).[6] Notably, different QUDs induce different partitions on the context set, which predetermine some specific possible evolutions of the common ground. Thus, for the conversation to proceed, the addressee must proffer an answer that is as informative as possible, or else indicate that no informative answer can be provided.

5 We adopt here the partition semantics for questions (Groenendijk and Stokhof 1984) because it allows for a neat and concise definition of a discourse goal.

6 The addressee can assert a reply from which an answer can be obtained via inference. We leave aside this phenomenon, limiting ourselves to an overview of the general approach.

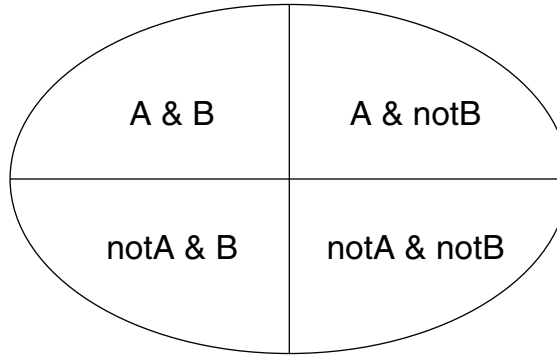

Figure 9.1: Partition of the context set.

9.3 Context stability

The view of the cooperative conversation outlined above rests on an implicit assumption—namely, that the context is (relatively) stable. Firstly, the group of participants is stable, in that when a participant joins or leaves the group, this is explicitly acknowledged by all the others. For instance, if a newcomer is assumed to lack some common ground information, the other participants will inform them, so that they will be able to contribute to the conversation in a relevant and informative way.[7]

Crucially, group stability guarantees that the common ground can be *monotonically updated* by new information—that is, an assertion can be straightforwardly assessed as compatible with the current common ground and informative with respect to it, and if it is, it can update the common ground without requiring any revision. In addition, the common ground of the conversation can be *non-generic*—that is, it may contain information that is only shared by the specific group in that specific moment. In turn, a stable common ground allows the group to pursue a common discourse goal by asking and answering QUDs.

7 See note 3 above.

Group stability and mutual acknowledgement also guarantee that every discourse move by a speaker is taken into account by all the others (even when the move is directed to a specific addressee in the group). In a situation of disagreement, where not all the participants accept as true an asserted proposition, the latter is not added to the common ground, but all the participants are aware of the incompatible commitments that have arisen; on the other hand, if no participant explicitly rejects an assertion, this typically counts as tacit acceptance by the conversational group, since acceptance can be considered the default reaction to an assertion (Farkas and Bruce 2010: 86, 99). When disagreement arises, it is often possible to open a conversational negotiation through which a shared commitment is eventually reached. If this is (assumed to be) impossible, the participants will agree to disagree. However, such a context of settled disagreement can still be cooperative: for instance, the participants may decide to pursue another relevant QUD.

It is important to stress that a given conversation context will display these properties to different degrees, and these properties may change in the course of the conversation. In a context where the participants' group is not fully stable, there may be a stable subgroup. Moreover, there are different proportions of common ground information specific to the conversational group; this in turn affects the range of possible QUDs. But crucially, there must be a stable core, however minimal, for cooperativity to be possible at all.

9.4 Unstable contexts

We defined above an online conversation context in a very narrow way, as a *single continuous discussion thread* starting from a post on a specific topic. Of course, the environment in which a post is published will already define a theme or orientation for the communication, so as to condition the participants' interests and beliefs: for instance, a post on a politician's Facebook page (see Corpus A in Section 9.7.3) will mostly attract comments from

users with an interest in politics, raising a number of expectations about the development of the discussion. The online environment thus constitutes a 'context' in a very broad sense, and some general information about it will be taken for granted by any participant. From the present perspective, however, what is relevant is the conversation dynamics, and for this reason we focus on the narrower notion of conversation context.

The type of conversation context that we are interested in is the discussion thread that follows a post on an open page, such that participation may be occasional. Here, the conversational group is unstable in a specific way: any participant can join in at any moment, participate for an undefined stretch of time, and leave at any moment, without prior notice and without the other participants being aware of it; note that this holds independently of how long and how often a participant intervenes. Because of this fundamental instability, replies or reactions to a discourse move are not guaranteed: even the smallest or most minimal responses may not be forthcoming. However, lack of reaction on the part of a given participant cannot be interpreted as default tacit acceptance, contrary to what may happen in face-to-face conversations. Whenever a reaction is manifested, this typically gives rise to interaction within a subgroup of participants – which is also unstable – and it is not possible to perceive its effect on any other participants in the discussion thread who do not directly intervene in it. The conversation context thus lacks an essential ingredient, namely mutual acknowledgement: since the discourse moves are not mutually acknowledged by a stable group of participants, it is virtually impossible—beyond the occasional interactions just mentioned—to maintain a common representation of how the conversation is evolving, and how the common ground is being updated.

Moreover, the very instability of the conversational group also implies that the participants can only presuppose a minimal and *generic* common ground; consequently, when a question is asked, it creates a partition on an exceedingly wide context set, and the pursuit of a specific discourse goal would preliminarily require an

unusually long chain of ordered QUDs. Under these conditions, it is extremely hard for the participants to engage in an exchange aiming at a common discourse goal. This implies that they will not be consistently committed to expanding the common ground by sharing new information among themselves.[8] For these reasons, we characterise this type of unstable context as *structurally non-cooperative*.

9.5 Our hypothesis

We have argued that in non-cooperative online contexts, speech acts do not—actually, *cannot*—aim to pursue an information-oriented discourse goal. The question, then, is what such speech acts aim at. We propose that they have an essentially *expressive* function: the author expresses their evaluation and their sentiment regarding the topic of the discussion thread (or some subtopic), and expects other participants (not *the* other participants, since the group is unstable) to express their own similar or opposite sentiment. Indeed, any (pseudo-)factual information is reported to justify their expressive stance.[9]

These speech acts have two alternative essential goals. One is that of maintaining the author's inclusion in an undefined 'in-group' of like-minded users, by manifesting and encouraging agreement, so as to yield the so-called 'chorus effect'.[10] The

8 Again, we may find fragments of cooperative exchanges among subgroups of participants, but crucially, the discussion thread as a whole does not pursue a general discourse goal.

9 In an information-oriented conversation, expressive content is present but is not at issue. See Potts (2007) for an approach to expressive content in possible world semantics.

10 This can be viewed as the common goal of a group of individuals. Note, however, that this type of cooperativity does not correspond to our definition: such a common goal is not a *discourse* goal—that is, it is not aimed at sharing information that restricts the context set. It is also important to emphasise that the actual communication expectations may depend on the affordances of the specific virtual environment—that is, the technical features of a social media platform that enable and

in-group does not coincide with the conversational group, which is unstable, nor with a recognisable collective entity; it is an imagined entity, whose membership condition is declaring oneself a member of it.[11] The alternative goal is to achieve intrusion into a discussion thread characterised by some 'out-group' stance that is perceived as opposite to one's own.

Note that when participants in such a discussion thread are expected to be largely like-minded, the expressive and emotive content conveyed by the author may well go unnoticed within the general chorus effect. We hypothesise that this is what leads the participants to intensify the content that they express on a relevant emotive and/or evaluative scale. This mechanism gives rise to excessive language, which exacerbates the polarisation between the perspectives of the in-group and the target out-group(s). Excessive language involves emotive, offensive, and aggressive communication that is not necessarily directed at specific groups or single individuals, as is the case with hate speech. However, the lack of information exchange easily shifts the target of excessive language from the topic itself to public persons who are perceived as representative of a target out-group, and to the participants who are perceived as opponents. Thus, excessive language easily degenerates into hate speech.

We thus propose that the purely expressive function of speech acts is linked to the structural non-cooperativity of the context. This leads us to expect that the less cooperative a context is, the more likely it is for excessive language to be found. In order to avoid a circular argument, it is necessary to identify some independent linguistic features that characterise (non-)cooperativity.[12]

---

constrain the actions and interactions of the network community (see Biri 2023 and references therein).

11 On the linguistic underpinnings of generic in-groups and out-groups, see the corpora analysis in Olmastroni, Bianchi, and Duguid (2021: 203–207).

12 As discussed above, disagreement does not *per se* imply lack of cooperativity: this is why we do not consider the presence of overt expressions of disagreement as a reliable indicator.

To this aim, in the next section we return to the notion of 'question under discussion'.

9.6 A typology of questions

In the framework described in Section 9.2, questions are normally asked in order to enhance the informative content of the participants' common ground, by partitioning the context set into disjoint alternatives (see the discussion around (1) above). Farkas (2020) proposes that at the pragmatic level, canonical questions are associated with the following default assumptions about the participants' epistemic states:

(i) *Open issue:* the speaker assumes that all the alternatives introduced by the question are neither positively nor negatively decided with respect to the current context set; in other terms, the context set neither entails not excludes any of the alternatives.
(ii) *Speaker ignorance*: the speaker presents themselves as having an epistemic state that does not support their commitment to any of the alternatives.
(iii) *Addressee competence*: the speaker presents themselves as assuming that the addressee's epistemic state supports the commitment to the 'true' alternative—that is, the cell of the partition that contains the actual world.
(iv) *Addressee compliance:* the speaker presents themselves as assuming that the addressee will resolve the issue by publicly committing to the true alternative. (Farkas 2020: 21)

Canonical information-seeking questions then require a cooperative addressee who fulfils (iv) by providing the most informative answer (complete or partial) that is supported by their epistemic state. Therefore, we can consider them as marking a cooperative dynamics. According to Farkas, non-canonical questions deviate from one or more of these assumptions.[13]

13 On non-canonical questions, see also Obenauer (2004), Garzonio (2004), Cruschina (2012), Giorgi (2016), Giorgi and Dal Farra (2019),

Deliberative self-addressed questions are non-canonical questions that introduce an open issue but do not require the addressee to provide the true answer as per (iii) and (iv):

(4) Should we go there by car or by train?

The goal of a deliberative question is thus to start the negotiation of a possible choice: the question is asked with the aim of partitioning the context set into disjoint cells that correspond to different action choices (Cariani, Kaufmann, and Kaufmann 2013), each of which is then cooperatively evaluated. Since there is no informational asymmetry between the participants, the default assumptions of speaker ignorance and addressee competence are suspended. However, deliberation-oriented questions introduce an open issue (i) and elicit a cooperative conversational dynamics; for this reason, we assume that they too characterise cooperative contexts.

Another type of non-canonical questions is *exclusively self-addressed questions* like (5):

(5) Teacher: Why do you have to show your work? Because I want to know how you reached the solution. (Farkas 2020: 24)

These questions introduce an open issue, as per (i), and are intended to increase the information publicly available to all participants; however, the speaker asks the question to point out an issue on which the addressee is assumed to be ignorant, and they immediately provide the answer. Here assumptions (ii)–(iv) are suspended; however, the question highlights an issue that is open in the current context set (i). Although commonly dubbed 'rhetorical', this question type introduces a sequence of speech acts that are cooperative, in that they are aimed at enhancing the participants' common ground.

Bianchi and Cruschina (2022), Cruschina and Bianchi (2022a,b), among others.

Real rhetorical questions, instead, do not comply with the open issue assumption:

(6) Should we leave them to drown in the sea?

The speaker does not assume that the alternatives in the question denotation are undecided, as per (i): on the contrary, they ask the question to emphasise that the resolution of the issue is obvious to everyone (this implies that addressee competence is assumed). Thus, the question does not ask for an answer on the part of the addressee, nor does it elicit a cooperative discourse dynamics (see Biezma and Rawlins 2017). Indeed, rhetorical questions have the effect of marking one of the alternatives as *non-negotiable* (Farkas 2020). We wish to stress that rhetorical questions are not *per se* an indication of context non-cooperativity: they can be used by an author to emphasise a point, or to mark expected similarity with the other participants' stances. It is rather the scarcity of the other question types (on the total number of questions asked) that characterises a non-cooperative conversation context.

To summarise, within this approach, canonical, deliberative, and self-addressed questions signal an information-oriented attitude on the part of the speaker, who assumes that they are participating in a cooperative exchange. For the purposes of our analysis, these three types are grouped together under the label 'cooperative questions'.[14] On the other hand, rhetorical questions are not assumed to introduce an open issue and do not require the addressee to answer; therefore, they are compatible with a non-cooperative context.

Based on this typology, our hypothesis can be operationalised in the following way: the more cooperative questions are found in a discussion thread (on the total number of questions asked), the less hate speech is expected to emerge. Conversely, in a context with a high incidence of hate speech, we expect a low proportion of cooperative questions. In the next section we present a proof-of-concept study conducted through corpus analysis.

14 Short for 'questions marking a cooperative conversation context'.

9.7 Corpora analysis

9.7.1 Independent factors

In designing our proof-of-concept study, two factors independent of our hypothesis were taken into account that might influence the incidence of hate speech in a discussion thread. The first one is the nature of the topic under discussion. It is likely that a highly controversial issue would arouse heightened emotions and would engender among the participants a less reflective attitude that could be conducive to hate speech. In this case, the very nature of the topic could lead the conversational community to a radical polarisation independently of the structural aspects of the context that we discussed above.

We believe that there is indeed a topic effect. Note, however, that if this were the main factor, it should condition the incidence of hate speech independently of the type of context in which it is discussed. In contrast, from our perspective, a polarised conversation context can still remain cooperative in the sense that we define above; in this case, a controversial issue can be discussed without giving rise to excessive language and hate speech.

The second independent factor is the social platform on which a discussion thread develops. It is probably the case that different social platforms differ with respect to their prevalent use in the online community, as well as in the ethical guidelines stated in the terms and policies with which users must agree when signing up. We claim, however, that the social platform should not be taken as the relevant notion of 'context': it is at the level of a single discussion thread that a conversational community arises and its dynamics is deployed. We will indeed show that one and the same social network can host quite different types of contexts.

In order to check for these independent factors we carried out two pairwise comparisons. In the first comparison, we built two corpora of discussion threads concerning two issues related to the LGBTQIA+ community, which were *equally controversial* in the Italian public debate at the time of observation. Corpus A is a corpus of Facebook comments about equality between homosexual

and heterosexual couples. Corpus B is a long discussion thread on Reddit about the nature of transgender identities, on which different participants express quite different views. The rationale is that if the controversy of the topic is the main conditioning factor, then we should expect a similar incidence of hate speech in the two corpora.[15]

In the second pairwise comparison, we compared Corpus A to another Facebook corpus, Corpus C, containing five discussion threads about the life of Italian immigrants in Finland, where the issues raised are unrelated to those of corpora A and B and are not particularly controversial. In this case, if the social platform is the main conditioning factor, we should expect a similar incidence of hate speech in A and C.

9.7.2 Criteria

In our analysis, we manually counted the number of posts/comments containing hate speech and the occurrence of different types of questions in each corpus, following the typology outlined in Section 9.6. For the purposes of this study, we adopted a fairly broad definition of hate speech, which subsumes the more recent definitions with an emphasis on the aggressive side of hate speech (see, e.g., Bianchi 2021), but which also includes excessive language not directly aimed at a specific individual or group (e.g. swear words). We thus classified as hate speech the following expressions:[16]

- insults, swear words, irony with explicit reference to sex;
- slurs, including innovative slurs referring to a political stance, e.g. *sinistroidi* 'leftoids', *pidioti* 'PD idiots';[17]

15 On the differences between the two social media platforms in terms of technological affordances, see Biri (2023) and references therein.

16 Note that this classification is functional to the specific goals and data of this study and may therefore differ from other definitions of hate speech (see Chapter 1).

17 PD (*Partito Democratico*, the Democratic Party) is a social-democratic political party in Italy.

- evaluative epithets such as *questa incapace* 'this incompetent [woman]'. Crucially, in this use the negative evaluation is not part of the at-issue content, but it is presupposed and hence not subject to explicit negotiation (Potts 2005; Murray 2014);
- verbs in an irrealis mood denoting violent actions, e.g. *frustarla sulle chiappe e poi mandarla ai lavori forzati* '[we should] whip her butt and then give her hard labour'.

All of these expressions were classified in the general category of hate speech. We did not classify as hate speech other forms of irony, GIFs, and emoji, except for one case: when an emoji was used to substitute a word used as a derogatory term—for example, the emoticon for 'shit' referring to an individual.

We manually counted the occurrences of questions classified as follows:

- canonical information-seeking questions, which manifest the speaker's intention to gain information from the interlocutor(s) by introducing a QUD;
- canonical questions directly answered by another participant, which manifest the occurrence of a cooperative exchange;
- deliberative questions, which address an open issue and invite the participants to express their evaluation of alternative action choices;
- rhetorical questions, which are characterised by the lack of an open issue and thus do not elicit an answer;
- self-addressed questions, through which the speaker introduces an open issue and provides the answer themselves. As discussed above, these differ from rhetorical questions proper in that they are intended to structure an argumentation providing information to the interlocutor, and in this respect can be considered cooperative speech acts.

We collected the results from each corpus in tables, and then compared the tables (see sections 9.7.2.–9.7.4 and Section 9.8).

9.7.3 Corpus A

Corpus A consists of two discussion threads from the official Facebook page in Italian of Simone Pillon (@SenatorePillon), a Member of Parliament who famously opposed the 'DDL Zan', a bill that concerned, most prominently (at least in public opinion), the legitimacy of equal rights between homosexual and heterosexual couples.[18] The first discussion thread is dated 27 October 2021—the end of the first parliamentary debate on the DDL Zan; the second is dated 13 January 2022—the date of a second discussion round for a modified version of the DDL. The two discussion threads consist of 246 and 166 comments, respectively, for a total of 412 comments, amounting to *c.*10,000 words. The initiating posts were not included in the corpus, since they were not introduced by a participant in the discussion but rather by the page owner or manager.

Most hate speech in the discussion threads was produced by users who expressed hate or disgust toward the Member of Parliament himself and his stance on the relevant issue. Table 9.1 shows the number of comments containing hate speech.

Table 9.1: Hate speech in Corpus A.

	Comments with hate speech	Total comments	%
Post 1	60	246	24.4%
Post 2	27	166	16.2%
Corpus A total	87	412	21.2%

18 The *disegno di legge* (DDL) Zan is a bill proposed by Alessandro Zan, a Partito Democratico MP. In Italy, a DDL is the initial phase of the process in which proposed new legislation is introduced by one or more members of parliament. The bill contains a set of articles that need to be discussed and (eventually) approved, one at a time, by the different branches of the parliament before becoming an effective law.

In Corpus A, we find a significant proportion of comments with hate speech. As we can see, the number of posts with hate speech is different in the two discussion threads, amounting to 60 out of 246 in the first thread and 27 out of 166 in the second thread. Overall, 21.2 per cent of comments in Corpus A contain hate speech.

In (8)–(10) we reproduce some examples of comments that were classified as containing hate speech. (For privacy reasons, we do not provide the users' names.)

(8) Devi morire! Ammazzato!!!!!

'You must die! Killed!!!!!'

(9) Figa se mi fai vomitare

'Fuck, you make me vomit'

(10) Sei una merda senza fine

'You're an endless [piece of] shit'

The manually counted totals of the different question types are summarised in Table 9.2.

Table 9.2: Question types in Corpus A.

	Canonical Qs not answered	Canonical Qs answered	Deliberative Qs	Exclusively self-addressed Qs	Rhetorical Qs	Total
Post 1	12	3	2	1	25	43
Post 2	13	1	1	0	73	88
Corpus A total	25	4	3	1	98	131
%	19.1%	3.1%	2.3%	0.8%	74.8%	100%

We observe a very low proportion of cooperative questions, which are significantly outnumbered by rhetorical questions (see examples (11)–(13)), the latter amounting to 78 per cent of the total. Among the former, the low number of answered questions shows the low incidence of cooperative exchanges addressing a QUD.

(11) E tu quando ti togli dal cazzo?

'When are you getting the fuck out of here?'

(12) Perché tua mamma non ti ha ingoiato?

'Why didn't your mum swallow you?'

(13) Che squallore certi contenuti fb dovrebbe bloccarli, cosa si prova nel sentirsi così inutile?

'What squalor! Certain fb contents should be blocked; what does it feel like to feel so useless?'

Finally, the virtual lack of self-addressed questions in this corpus is compatible with the absence of structured argumentation.

9.7.4 Corpus B

Corpus B consists of a single discussion thread published on Reddit on 3 May 2021.[19] The author who initiated the discussion thread published a long and detailed post entitled 'Persone transgender e identità di genere' (Transgender persons and gender identity), which was followed by a discussion between a limited number of participants. The discussion thread contains 216 posts for a total of *c*.20,000 words. The ratio shows that the posts in this thread were, on average, significantly longer than those in Corpus A.

19 CYP4502D6, 'Persone transgender e identità di genere', Reddit (r/italy), 3 May 2021, https://www.reddit.com/r/italy/comments/n3rftm/persone_transgender_e_identit%C3%A0_di_genere/.

The same classification criteria were adopted as for Corpus A. Crucially, the topic of transgender identities was controversial in the discussion, where different positions were expressed. Despite this, the ratio of hate speech is negligible, as shown in Table 9.3: virtually all instances of hate speech consisted in the expression of a heightened tone through swear words.

Table 9.3: Hate speech in Corpus B

	Comments with hate speech	Total comments	%
Corpus B	7	216	3.2%

On the other hand, the manual count of different question types yielded very different results from Corpus A, as shown in Table 9.4.

Table 9.4: Question types in Corpus B.

	Canonical Qs not answered	Canonical Qs answered	Deliberative Qs	Exclusively self-addressed Qs	Rhetorical Qs	Total
Corpus B	16	32	2	20	38	108
%	14.8%	29.6%	1.8%	18.5%	35.2%	100%

We can observe that 18.5 per cent of the posts in Corpus B contain self-addressed questions, which suggests the presence of structured argumentation in the discussion thread. The percentage of canonical questions (14.8 per cent) and especially of answered canonical questions (29.6 per cent) shows that there was a significant incidence of cooperative exchanges driven by QUDs or attempts at initiating them. The combined percentage of canonical questions, answered or otherwise, amounts to 44.4 per cent

and exceeds that of rhetorical questions (35 per cent). In (14) an example is reported of an answered canonical question, while in (15) we can see examples of canonical questions consecutively produced by the same speaker:[20]

(14) A: Il problema di fatto è che le definizioni di genere, identità di genere, uomo e donna sono o vuote anche nell'astratto o con conseguenze pratiche non ideali. Tu che definizioni daresti a quelle 4 parole?

'The actual problem is that the definitions of gender, gender identity, man and woman are either empty also at an abstract level, or they have undesirable practical consequences. How would you define those 4 words?'

B: Io darei le definizioni che ho utilizzato nel post.

'I would give the definitions that I used in the post.'

C: Provo a rispondere io, anche se, tieni presente, al momento ciascuno usa i vari termini un po' come preferisce perché non c'è un consenso netto

'I'll try to answer on my part, even though, you should realise, as of now everyone uses the terms as they prefer, because there is no clear consensus.'

(15) Hai fonti da linkare? Ci sono affermazioni e conclusioni da parte di studiosi? O più semplicemente potresti motivare la tua affermazione?

'Do you have sources to link? Are there statements and conclusions by scholars? Or could you simply justify your statement?'

20 In the following, capital letters are used to list examples from different speakers and to list their conversational turns (see (14), (18), and (19)), while the use of lower-case letters, as in (16), indicates that the examples were consecutively produced by the same speaker.

The examples in (16) illustrate two consecutive self-addressed questions, answered by the speaker themselves, while in the comment in (17) we find a rhetorical question:

(16) a. Allora perché non siamo tutti transgender? La differenza sta nella cosiddetta DISFORIA DI GENERE.

'So why aren't we all transgender? The difference lies in so-called gender dysphoria.'

b. Cos'è la disforia? È un sentimento spiacevole che si prova nei confronti di se stessi e di come si viene percepiti dagli altri in merito al proprio GENERE.

'What is dysphoria? It is an unpleasant feeling that one experiences about oneself and about how one is perceived relative to one's gender.'

(17) Per le persone transgender non capisco questo accanimento sulla questione, dopo un accurato controllo medico e psicologico, se i medici lo ritengono necessario che problema c'è?

'For transgender people I do not understand such doggedness on the issue, after an accurate medical and psychological examination, if the physicians think it [transition] necessary what's the problem?'

The results summarised in Table 9.4 show that the overall discussion thread qualifies as a (mostly) cooperative context in the sense defined above.

9.7.5 Corpus C

Corpus C was taken from the Facebook private group *Gruppo degli italiani in Finlandia* (Italians in Finland), at a time when the members numbered around 4000. The corpus consists of eight discussion threads that appeared between 14 and 29 September

2021. The total number of posts and comments is 290, for a total of *c.*10,000 words. In this case, as for Corpus B, the initiating posts were counted together with the comments because, unlike in Corpus A, they were introduced by an actual participant.

The discussion threads were about various aspects of life in a foreign country, with an emphasis on bureaucratic procedures and the issue of learning the official language; the posts mostly focused on sharing useful information. Two threads were initiated by a participant who was planning to move to Finland, and another participant who was planning a holiday there. In both cases, the long-term Italian immigrants responded by providing information and at the same time commenting on living conditions in Finland.

The manual count followed the same classification criteria as in corpora A and B. Table 9.5 shows that the percentage of comments with hate speech is 0.

Table 9.5: Hate speech in Corpus C.

	Total comments	Comments with hate speech	%
Corpus B	290	0	0%

The occurrences of question types are reported in Table 9.6.

Table 9.6: Question types in Corpus C.

	Canonical Qs not answered	Canonical Qs answered	Deliberative Qs	Exclusively self-addressed Qs	Rhetorical Qs	Total
Corpus C	20	35	0	1	10	66
%	30.3%	53.0%	0%	1.5%	15.1%	100%

The distribution shows a very high percentage of canonical questions, most of which (53 per cent) were answered. There are virtually no self-addressed questions, which is compatible with the lack of structured argumentation in the posts. The percentage of rhetorical questions is 15.1 per cent; all in all, the ratio of rhetorical to non-rhetorical questions is almost symmetrical to that of Corpus A. In (18) and (19) we present examples of cooperative exchanges in Corpus C:

(18) A: Buongiorno, adesso vivo in Finlandia posso rimanere solo 3 mesi. Ma la mia ragazza è finlandese posso rimanere qui oltre questo tempo visto che siamo una coppia? Cosa potrei fare? Grazie

'Good morning, I currently live in Finland, I can only stay for 3 months. But my girlfriend is Finnish, can I stay here longer since we are a couple? What could I do? Thanks'

B: Buonasera. Sapresti lavorare come cameriere?

'Good evening. Would you be able to work as a waiter?'

C: scusa ma che lavoro stai cercando? Qui la situazione non è delle più rosee.

'Sorry, but what kind of job are you looking for? The situation is not so good here.'

(19) A: Ciao a tutti, mi consigliate qualcosa di bello da fare per trascorrere il mio compleanno in famiglia? Abbiamo 2 bimbi di 6 e 8 anni. Non conosco granché essendo qui da poco più di un anno, magari qualcuno di voi ha qualche bella idea da propormi.

'Hello, everyone, could you suggest something nice to do for my birthday with my family? We have 2 children aged 6 and 8. I don't know much around here

since I have been here for little over a year, perhaps someone has a nice idea to suggest.'

B: Escape room, ce ne sono di semplici da fare coi bambini. Parco acquatico Flamingo. Se il tempo è bello, ci sono i parchi di arrampicata sugli alberi. A costo zero, passeggiata a Nuuksio con salsiccia grigliata. Cinema e pizza.

'Escape room, there are easy ones that you can do with children. Aqua Park Flamingo. If the weather is good, there are parks where you can climb trees. For free, a walk in Nuuksio with grilled sausages. Cinema and pizza.'

A: Abbiamo scelto Flamingo, grazie mille! Tu ci sei stata? Hai qualche consiglio da darmi? Basta portarsi solo costume e accappatoio o serve altro?

'We chose Flamingo, thanks a lot! Have you been there? Any advice? Do we need to take just swimming costumes and bathrobe or anything else?'

Most of the rhetorical questions in Corpus C, examples of which are presented in (20) and (21), came from a single discussion thread—that is, a post on integration and language policies in Helsinki:

(20) Abbassare le tasse? Ahah ma se sono più basse che in Italia

'Lowering the taxes? Haha, but if they are lower than in Italy'

(21) perché mai uno dovrebbe sbattersi ad imparare il finlandese, quando con le stesse qualifiche può andare altrove, con meno problemi, meno tasse, ed uno stile di vita anche migliore?

'why on earth should one make the effort to learn Finnish, when with the same qualifications one can go elsewhere, with less problems, less taxes, and an even better quality of life?'

9.7.6 Summary

The results from the three corpora are summarised in Table 9.7.

Table 9.7: Summary of results for all three corpora.

	Canonical Qs not answered	Canonical Qs answered	Deliberative Qs	Exclusively self-addressed Qs	Rhetorical Qs	Comments with hate speech
Corpus A	25	4	3	1	98	87/412
131 Qs	19.1%	3.1%	2.3%	0.8%	74.8%	21.1%
Corpus B	16	32	2	20	38	7/216
108 Qs	14.8%	29.6%	1.8%	18.5%	35.2%	3.2%
Corpus C	20	35	0	1	10	0/290
66 Qs	30.3%	53.0%	0%	1.5%	15.1%	0%

In Table 9.8, we group together all cooperative question types.

Table 9.8: Comparison of the three corpora: cooperative vs rhetorical questions.

	Cooperative questions	Rhetorical questions	Comments with hate speech
Corpus A	25.2%	74.8%	21.1%
Corpus B	64.7%	35.2%	3.2%
Corpus C	84.8%	15.1%	0%

9.8 Discussion

The trends that we expected are confirmed. There is, in particular, a dramatic split between Corpora B and C, where cooperative questions are significantly above 50 per cent, and Corpus A, where cooperative questions make up around 25 per cent of all questions; the incidence of hate speech in B and C is at least seven times less than in Corpus A.

On the other hand, the incidence of rhetorical questions is more than double in Corpus A with respect to Corpus B, and it is almost five times higher in Corpus A than in Corpus C. There still is a non-negligible proportion of rhetorical questions in B and C; however, as noted above, rhetorical questions *per se* are not incompatible with cooperativity. What is relevant, instead, is the incidence of cooperative questions that introduce, at least potentially, a QUD.

The asymmetries between the three corpora are even sharper if we compare the number of *answered* canonical questions, as shown in Table 9.9.

Table 9.9: Comparison of the three corpora: answered canonical questions.

	Canonical Qs answered	Comments with hate speech
Corpus A	3.1%	21.1%
Corpus B	29.6%	3.2%
Corpus C	53.0%	0%

We conclude that typology of questions seems to be a reliable indicator for distinguishing different types of conversation contexts. Following this criterion, the discussion threads in Corpus B and those in Corpus C qualify as significantly more cooperative than the two discussion threads in Corpus A. The comparison of A and C shows that there is indeed a topic effect, such that a highly controversial topic, unsurprisingly, leads to polarisation.

But crucially, the topic *per se* is not a *sufficient* condition for the emergence of hate speech. In other terms, it is not only the topic that counts but also the conversation context in which it is discussed.

As a side note, we observe a difference between Corpora B and C when it comes to self-addressed questions. We take this question type to mark the presence of a structured argumentation in the discussion thread. The asymmetry corresponds to the different nature and goals of the contexts of the two corpora: while in B the discussion revolves around a general topic at an abstract ideological level, in Corpus C it mostly revolves around practical questions.

9.9 Summary and conclusions

In this chapter we have proposed an approach to hate speech that focuses on the intrinsic properties of the online conversation contexts in which it might emerge. Through the tools of formal pragmatics, we identified some properties that distinguish cooperative contexts from less (or non-)cooperative ones. In less cooperative contexts, the conversational group is unstable and, for this reason, they do not share a specific common ground; reactions to any speech act are not guaranteed, and it is virtually impossible to establish a common discourse goal to be pursued through QUDs. We argued that under these conditions, the participants' speech acts have a merely expressive function, conveying the author's sentiment, and are aimed at an ideologically homogeneous group, for the author to be recognised as a member of that group (or to intrude into it). To produce an expressive speech act that will have a significant impact, the author is led to express extreme evaluations or sentiments, which may lead to hate speech against a specific target: a person or an out-group. This led us to hypothesise that unstable, non-cooperative conversation contexts are more likely to host hate speech.

To operationalise our hypothesis, we adopted a typology of questions that distinguishes those that potentially introduce a

QUD from those that do not; we assumed that the former are an indicator of context cooperativity. We conducted a proof-of-concept study of three corpora, which suggests that the rate of cooperative questions is a reliable indicator of a significant property of online conversation contexts.

We are aware that our results are far from conclusive: the number of corpora analysed and their dimensions are very limited. The procedure of manual counting is time-consuming, and to the best of our knowledge, there is no way to automatically classify questions according to the typology that we adopted, because there are no systematic lexical or syntactic cues that distinguish the various types. However, we have developed a methodology that we hope has been shown to be promising. Note that the index that we have proposed here—question types—should in principle be valid across languages: thus, we envisage a possible cross-linguistic comparison of online exchanges in different linguistic communities. In future work we plan to investigate other possible indicators of context cooperativity.

We conclude with a bold and hopeful suggestion. At the beginning of the era of social media, one could hope that public debate would be enhanced and supported by easy access to information; but things turned out differently, with a proliferation of 'fake news' and of hate speech. We believe that by shedding light on the relevant *structural* properties of online contexts, an approach might emerge that goes beyond mere censorship of pathological online behaviour, and instead aims to raise the users' awareness of the conversational dynamics that characterise online contexts.

References

Baider, Fabienne. 2020. 'Pragmatics lost? Overview, synthesis and proposition in defining online hate speech'. *Pragmatics and Society* 11 (2): 196–218. https://doi.org/10.1075/ps.20004.bai.

Bianchi, Claudia. 2021. *Hate speech: il lato oscuro del linguaggio*. Bari and Rome: Laterza.

Bianchi, Valentina, and Silvio Cruschina. 2022. 'Ignorance and competence implicatures in central Sicilian polar questions'. In 'Romance Languages and Linguistic Theory 17', edited by Ora Matushansky, Laurent Roussarie, Michela Russo, Elena Soare, and Sophie Wauquier, special issue, *Isogloss: Open Journal of Romance Linguistics* 8 (2): 1–20. https://doi.org/10.5565/rev/isogloss.117.

Biezma, María, and Kyle Rawlins. 2017. 'Rhetorical questions: severing asking from questioning'. *Proceedings of Semantics and Linguistic Theory* (SALT) 27: 302–322. https://doi.org/10.3765/salt.v27i0.4155.

Biri, Ylva. 2023. 'Stancetaking in Interest-Based Online Communities: A Corpus Pragmatic Comparative Analysis'. PhD thesis, University of Helsinki. https://helda.helsinki.fi/items/49596ae2-12c9-490f-b978-aef57bc36177.

Brown, Alexander. 2018. 'What is so special about online (as compared to offline) hate speech?' *Ethnicities* 18 (3): 297–326. https://doi.org/10.1177/1468796817709846.

Cariani, Fabrizio, Magdalena Kaufmann, and Stefan Kaufmann. 2013. 'Deliberative modality under epistemic uncertainty'. *Linguistics and Philosophy* 36: 225–259. https://doi.org/10.1007/s10988-013-9134-4.

Cruschina, Silvio. 2012. *Discourse-Related Features and Functional Projections*. Oxford and New York: Oxford University Press.

Cruschina, Silvio, and Valentina Bianchi. 2022a. 'The syntactic encoding of conventional implicatures in Sicilian polar questions'. *Probus* 34 (1): 77–109. https://doi.org/10.1515/probus-2022-0006.

Cruschina, Silvio, and Valentina Bianchi. 2022b. 'Variation at the syntax–pragmatics interface: discourse particles in questions'. *Languages* 7 (3): 183. https://doi.org/10.3390/languages7030183.

Esposito, Eleonora, and Majid KhosraviNik, eds. 2024. *Discourse in the Digital Age: Social Media, Power, and Society*. London: Routledge.

Farkas, Donka. 2020. 'Canonical and non-canonical questions'. Unpublished manuscript, University of California at Santa Cruz. Accessed 22 August 2024. https://semanticsarchive.net/Archive/WU2ZjIwM/questions.pdf.

Farkas, Donka, and Kim B. Bruce. 2010. 'On reacting to assertions and polar questions'. *Journal of Semantics* 27 (1): 81–118. https://doi.org/10.1093/jos/ffp010.

Fogal, Daniel, Daniel Harris, and Matt Moss, eds. 2018. *New Work on Speech Acts*. Oxford: Oxford University Press.

Garzonio, Jacopo. 2004. 'Interrogative types and left periphery: some data from the Fiorentino dialect'. In *Quaderni di Lavoro ASIS* 4, edited by Barbara Patruno and Chiara Polo, 1–19. Padua: Unipress.

Ginzburg, Jonathan. 1996. 'Dynamics and the semantics of dialogue'. In *Logic, Language, and Computation*, vol. 1, edited by Jerry Seligman and Dag Westerstahl, 221–237. Stanford, CA: CSLI Publications.

Giorgi, Alessandra, and Chiara Dal Farra. 2019. 'On the syntax/pragmatics interface: expressing surprise and disapproval'. *Intercultural Pragmatics* 16 (3): 335–361. https://doi.org/10.1515/ip-2019-0017.

Giorgi, Alessandra. 2016. 'On the temporal interpretation of certain surprise questions'. *SpringerPlus* 5 (1): 1390. https://doi.org/10.1186/s40064-016-2951-5.

Grice, H. Paul. 1975. 'Logic and conversation'. In *Syntax and Semantics*, vol. 3: *Speech Acts*, edited by Peter Cole and Jerry L. Morgan: 41–58. New York: Academic Press.

Groenendijk, Jeroen, and Martin Stokhof. 1984. 'Studies on the Semantics of Questions and the Pragmatics of Answers'. Unpublished PhD thesis, University of Amsterdam.

Lewis, David. 1979. 'Scorekeeping in a language game'. *Journal of Philosophical Logic* 8: 339–359. https://doi.org/10.1007/BF00258436.

Murray, Sarah E. 2014. 'Varieties of update'. *Semantics and Pragmatics* 7 (2): 1–53. https://doi.org/10.3765/sp.7.2.

Obenauer, Hans-Georg. 2004. 'Nonstandard *wh*-questions and alternative checkers in Pagotto'. In *Syntax and Semantics of the Left Periphery: Interface Explorations 9*, edited by Horst Lohnstein and Susanne Trissler, 343–384. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110912111.343.

Olmastroni, Francesco, Valentina Bianchi, and Alison Duguid. 2021. 'A deliberative bridge over the mass-elite rift: Effects of online deliberation on support for European integration'. In *The EU through multiple crises*, edited by Maurizio Cotta and Pierangelo Isernia, 186–220. New York: Routledge.

Potts, Christopher. 2007. 'The expressive dimension'. *Theoretical Linguistics* 33 (2): 165–198. http://dx.doi.org/10.1515/tl.2007.011.

Potts, Christopher. 2005. *The Logic of Conventional Implicatures*. Oxford: Oxford University Press.

Roberts, Craige. 2018. 'Speech acts in discourse context'. In Fogal, Harris, and Moss, *New Work on Speech Acts*, 317–359. https://doi.org/10.1093/oso/9780198738831.003.0012.

Roberts, Craige. 2012. 'Information structure: towards an integrated formal theory of pragmatics'. *Semantics and Pragmatics* 5 (6): 1–69. https://doi.org/10.3765/sp.5.6. Originally published in 1996 in 'Papers in

Semantics', edited by Jae-Hak Yoon and Andreas Kathol, *Working Papers in Linguistics* 49: 91–136. Columbus, OH: Ohio State University, Department of Linguistics.

Stalnaker, Robert. 2018. 'Dynamic pragmatics, static semantics'. In Fogal, Harris, and Moss, *New Work on Speech Acts*, 394–399. https://doi.org/10.1093/oso/9780198738831.003.0014.

Stalnaker, Robert. 2002. 'Common ground'. *Linguistics and Philosophy* 25: 701–721. https://doi.org/10.1023/A:1020867916902.

Stalnaker, Robert. 1999. 'Assertion'. In *Context and Content: Essays on Intentionality in Speech and Thought*, Chapter 4, 78–95. Oxford: Oxford University Press. Originally published in 1978 in *Syntax and Semantics 9: Pragmatics*, edited by Peter Cole, 315–332. New York: Academic Press.