

CHAPTER 1

Words that matter

The use of language with hate purposes

Silvio Cruschina
University of Helsinki

Chiara Gianollo
University of Bologna

Abstract

This chapter introduces the rationale behind the volume and the main topics discussed in the various chapters. It surveys the difficulties surrounding the definition of hate speech and singles out the main issues that are relevant for its linguistic investigation: besides the lexical elements (slurs, insults, derogatory epithets), more hidden pragmatic and grammatical strategies are also argued to characterise hate speech and aggressive language. In this respect, a rigorous evaluation of the contextual conditions by means of the tools provided by linguistics helps towards establishing a more precise identification of types of hate speech in

How to cite this book chapter:

Cruschina, Silvio, and Chiara Gianollo. 2024. 'Words that matter: The use of language with hate purposes'. In *An Investigation of Hate Speech in Italian: Use, Identification and Perception*, edited by Silvio Cruschina and Chiara Gianollo, 1–25. Helsinki: Helsinki University Press. DOI: <https://doi.org/10.33134/HUP-27-1>.

conversational dynamics (explicit and implicit hate speech, intensity and degree of offensiveness, intentions, and effects).

Keywords: hate speech, aggressive language, context-dependency, implicit meaning, speaker intentions

1.1 Introduction

Language is a key element in the construction and reinforcement of social identities, and, as a consequence, also in the creation and diffusion of stereotypes, discrimination, and social injustices. The use of language to attack an individual or group based on attributes such as race, ethnicity, religion, gender, nationality, political ideology, disability, or sexual orientation constitutes the basis of what is now known as *hate speech*. Hate speech is certainly not a new phenomenon. However, it is only since the 2000s that continuously evolving and fast-paced digital communication, combined with the amplifying function of social media, has made it a major topic of research in a variety of areas, including discourse analysis, psychology, sociology, philosophy of language, media, political, legal, and gender studies (see, among others, Leets 2002; Sternberg 2004; Van Blarcum 2005; Parekh 2006; Lillian 2007; Daniels 2008; Maitra and McGowan 2009; Bleich 2011; Herz and Molnar 2012; Waldron 2012; Foxman and Wolf 2013; Gagliardone et al. 2015; Gelber and McNamara 2016; Brown 2017; Richardson 2018; Knoblock 2022; Ermida 2023a; Guillén-Nieto 2023).

Social media channels have come to play an increasingly large role in our everyday lives and communication, creating novel discursive practices and technological affordances (see the recent overview in Esposito and KhosraviNik 2024). They provide a context in which people across the world can communicate, share knowledge, exchange messages, and interact with each other, irrespective of the distance and the social differences between them, thus allowing greater freedom of expression and empowering individual voices. At the same time, however, social media channels also enable anti-social behaviour, cyberbullying, online aggression, and hate speech. These manifestations are intensified

by the virtual nature of the interaction, which tends to remove socially imposed inhibitions, and can be particularly harmful because of the more persistent nature of the message, which is in written form and has the potential to reach a wide audience.

In this new culture of communication, which has been adopted and adapted in other contexts such as politics and other forms of public speech, appeals to emotions and personal beliefs are important persuasive devices. Verbal aggressions, offensive propaganda, and the construction of authority and subordination both in speech and in writing have become frequent features and instruments of communication in multiple spheres of society.

In the research on hate speech, attention has been focused on two main aspects: i) the individuals or groups who have overtly or covertly been victims of aggression or discrimination through hate speech, and ii) the legal and ethical controversy around the boundaries between the right to freedom of speech and the use of hate speech. This volume contributes to the investigation of hate speech by adopting the methodological and theoretical tools of linguistics.¹

The focus of this volume is on the use and perception of hate speech, which can be produced either by lexical means (e.g. insults, derogatory terms or epithets) or via more subtle grammatical and pragmatic strategies related to implicit meanings or atypical conversational dynamics. The aim of this investigation is to identify the common linguistic characteristics and features of hate speech in different domains of communication and to establish a set of

1 The contributions in this volume stem from the work and collaboration of a research network sponsored and funded by Una Europa, an alliance of European universities (UNA Europa seed-funding scheme, funding number: SF2019003). This network brought together academic researchers to investigate the topic of hate speech in Italian within a project entitled 'A Linguistic Investigation of Hate Speech (ALIHAS): How to Identify It and How to Avoid It'. The preliminary results were presented in a digital workshop organised by the University of Helsinki on 17 May 2021. The papers were later developed and presented in a second workshop that took place at the University of Bologna on 12 November 2021.

criteria that can help distinguish between hate speech and freedom of expression. The studies collected in this volume all focus on Italian, with the aim of collecting data and drawing generalisations starting from relatively homogeneous conditions and allowing for immediate comparability. However, the analyses, methodologies, and findings of the individual chapters can easily be extended to other languages for comparative and contrastive purposes.

The emphasis is on the linguistic strategies and tools that are typically employed for hate purposes, and on the context and the communication situation that foster hate speech. The investigation is not limited to the more obvious and easier to recognise lexical elements of aggressive language and hate speech (e.g. slurs, derogatory epithets, metaphorical offences and insults; see Faloppa 2004, 2012; Croom 2013; Bianchi 2014, 2015; Capone 2014; Bolinger 2017; Cepollaro 2015, 2017; Retta 2023); the use of more hidden pragmatic and grammatical strategies, and the concomitant properties of the contexts in which hate speech proliferates will also be explored. Most of the studies contained in this collection address hate speech on social media, exploiting the potential of these communication channels as an invaluable source of linguistic data that would otherwise be difficult to collect and analyse in a systematic way. The focus on social media, however, should not obscure the fact that online discourse is inextricably connected to the offline settings in which hate speech and aggressive language emerge as a product of the same social context.

1.2 Definition and identification

The first problem to be addressed in an investigation of hate speech is how to define it. Hate speech is a concept that is intuitively easy to grasp but difficult to rigorously define. Indeed, its definition both as a scientific and as a legal notion raises many complex questions ranging across several disciplines: what is hate speech? Who is the target of hate speech? What are the boundaries between hate speech and freedom of speech? (See, among many others, Assimakopoulos 2020; Baider, Millar, and Assimakopoulos

2020; Faloppa 2020: Ch. 1; Anderson and Barnes 2022; Ermida 2023a, Määttä 2023).

The need for a precise and binding definition of hate speech is felt particularly strongly by legislative institutions and international organisations looking to recommend specific measures or legislations against hate speech. In 1997 the European Council provided the governments of its member states with a set of principles and recommendations to combat hate speech, arriving at the following definition:

the term ‘hate speech’ shall be understood as covering all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance, including: intolerance expressed by aggressive nationalism and ethnocentrism, discrimination and hostility against minorities, migrants and people of immigrant origin. (Council of Europe 1997)

This definition makes reference to the purposes and targets of hate speech but remains quite *open* in its scope of application, using the verb ‘including’ before the list of potential cases of hate speech. Moreover, it puts a strong emphasis on hate speech motivated by racism and xenophobia. In this respect, the concise definition provided by the United Nations is at the same time more precise and more comprehensive:

the term hate speech is understood as any kind of communication in speech, writing or behaviour, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor. This is often rooted in, and generates intolerance and hatred and, in certain contexts, can be demeaning and divisive. (United Nations 2019)

This definition identifies both the operational means of hate speech and its targets—that is, it clarifies the type of speech (‘any kind of communication in speech, writing or behaviour, that

attacks or uses pejorative or discriminatory language’) and the type of target (‘a person or a group on the basis of their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor’), expanding the scope to include all factors that contribute to personal or group identity, which can be targeted in a discriminatory way. According to this definition, hate speech is therefore not limited to extreme verbal aggressions characterised by abusive and offensive language (see [Section 1.3](#) of this chapter), but includes all forms of communication that lead to discrimination or contempt.

Bianchi (2021) reaches a similar definition from the perspective of philosophy of language:

quelle espressioni e quelle frasi che comunicano derisione, disprezzo e ostilità verso gruppi sociali, e verso individui in virtù della loro mera appartenenza a un certo gruppo; le categorie che sono bersaglio o target dei discorsi d’odio vengono anche in questo caso identificate sulla base di caratteristiche sociali (reali o percepite) come etnia, nazionalità, religione, genere, orientamento sessuale (dis)abilità, e così via. (Bianchi 2021: 5)

(those expressions and sentences that communicate ridicule, contempt and hostility towards social groups, and towards individuals simply because they belong to a certain group; in this case too, the categories that are the target of hatred discourse are identified on the basis of social characteristics (real or perceived) like ethnicity, nationality, religion, gender, sexual orientation, (dis)ability, and so on.)²

Two important aspects emerge from this definition. First, there can be different degrees of intensity of hate speech, building a spectrum from ridicule to outright hostility, with different consequences for the communication dynamics and for the legal sanctioning of discourses of hatred (see [Section 1.3](#) below for examples of implicit and covert forms of hate speech). Secondly, the

2 Our translation.

characteristics upon which discrimination is built can have an objective counterpart in reality, or can be represented by perceived social constructs. Moreover, it is clear from the typical social characteristics involved in producing hate speech that discriminatory discourses often capitalise on intersectionality—that is, the intertwining and overlapping of dimensions of identity.

From the definitions above, we clearly see that offensive communication can only be considered as hate speech if it is directed at groups or single individuals (usually as representatives of a group) and does not include verbal attacks on political institutions, public offices, or leaders. In fact, the target of hate speech is typically understood as a minority group—or an individual taken as representative of a minority group—which has been subject to persecution and discrimination. For instance, if a politician verbally attacks another politician with excessive and offensive language, this can hardly be viewed as hate speech because it does not give rise to the creation and diffusion of stereotypes, discrimination, and social prejudices against a (minority) group. From the legislative viewpoint, possible offences in political debate are regulated by specific laws against insults and defamation, which relate solely to the individual. By contrast, if a politician directs violent or offensive language towards a person or a group, for instance on the basis of their nationality (as immigrants) or religion, this *does* count as hate speech, because it reinforces stereotypes and incites or justifies discrimination.

On the basis of these considerations, we distinguish hate speech from more general aggressive language—that is, offensive, violent, and excessive language which is not directed at the groups or single individuals who can often be the targets of hate speech. The two notions are nevertheless not unrelated: since aggressive language can easily turn to hate speech as soon as the target of the abusive communication changes, understanding the linguistic properties of aggressive language may be pivotal to the study of hate speech (on this distinction, see also Ermida 2023b).

It is important to note, moreover, that the definitions provided by the European Council and the United Nations do not have

legal validity, insofar as they do not include strict guidelines for distinguishing hate speech from free speech. This issue is particularly relevant to legislators who can only promulgate and apply laws against hate speech if it is defined as a coherent and clearly demarcated concept. In fact, since the perception and the characteristics of the groups listed in definitions such as that provided by the United Nations can be historically and culturally conditioned, it is not surprising that the laws regarding hate speech, when they exist, vary significantly across countries (Fish 1994; Butler 1997; Perry 2001; Brown 2017; Määttä 2020).

Another argument that is often brought up as an obstacle to the definition of hate speech and to its study as a coherent concept is the role of context. We know that the interpretation of linguistic expressions depends on the context; for some scholars, therefore, hate speech should also be examined in the light of the contextual conditions that characterise the situation of communication, such as the (offensive vs non-offensive) intentions of the speaker, the relationship between the interlocutors, and the use of sarcasm. According to these scholars, in the presence of certain conditions the context would *not* make an utterance hate speech. Because of these complexities, several scholars have even rejected the idea that hate speech constitutes a coherent concept that can be investigated as such (see Boromisza-Habashi 2021).

In approaching the aims of the research project presented in this volume, we acknowledge the difficulties in establishing an unambiguous and universal definition of hate speech, both in theoretical and in applied terms; at the same time, we would like to emphasise two undeniable aspects of hate speech that call for an improved understanding of its linguistic determinants.

First of all, although hate speech is also difficult to define *a priori* from a linguistic viewpoint, the identification and perception of hate speech is much easier. As speakers of a language, we have clear intuitions regarding the presence of hate speech in verbal—or even non-verbal—communication, and we are also able to perceive different levels of intensity. From a scientific viewpoint, and independently of reaching a precise definition, the investigation

of hate speech is therefore crucial in understanding where such speech comes from, how it is generated, what its triggers are, and what contributes to its intensity.

Secondly, hate speech can be investigated independently of the context, on the basis of its content and its intrinsic properties. However, even if we admit that hate speech is strongly context-dependent, this should not be used as a reason to reject the systematic study of hate speech. On the contrary, linguists have the theoretical means to analyse and examine the contextual domain in which hate speech is produced, distinguishing between triggering and concomitant features, and between descriptive and performative functions of language (see, e.g., Bianchi 2018, 2021), and identifying the environments that legitimate the use of certain slurs, as in the case of metalinguistic uses or reappropriation (see Galinsky et al. 2003 and Brontsema 2004 on the reappropriation of terms such as ‘bitch’ and ‘queer’; see also chapters [2](#) and [6](#) in this volume). In sum, the contextual variability of hate speech can also be subject to a principled account, allowing for a more nuanced understanding, with important empirical consequences.

The goal of the chapters in this volume is to contribute to the establishment of reliable criteria for the identification of the common linguistic characteristics of hate speech in different contexts of communication. In turn, this could be seen as a first essential step towards the far more complex issue of a scientific definition of hate speech and its practical application at the level of regulation and legislation.

1.3 Explicit and implicit hate speech

From a linguistic perspective it is generally recognised that there are two forms of hate speech (Gao, Kuppersmith, and Huang 2017; Caselli et al. 2020; Faloppa 2020; Brambilla and Crestani 2021): explicit and implicit. The explicit manifestations of hate speech are easier to identify, in that they are typically represented by lexical expressions that contain insults, derogative terms or epithets, threats, or overt references to stereotypes. With these words we

can offend, insult, discriminate, and impose our purported superiority or authority. The intensity of the offensiveness associated with these lexical items or expressions is commonly recognised and can also be measured (see chapters 4 and 5 in this volume).

However, stereotypes, prejudices, discrimination, intolerance, and hatred more generally can also be reinforced implicitly, in more subtle and hidden ways (see Ben-David and Fernández 2016; Bhat and Klein 2020; also Baider 2019, 2023, where the term ‘covert hate speech’ is used). These implicit manifestations of hate speech are more complex and difficult to detect and to investigate, in particular for natural language processing, which is an important regulatory resource used by social media (see Schmidt and Wiegand 2017; Fortuna and Nunes 2018; [Chapter 5](#) in this volume).

If I utter ‘You fucking faggot!’, I am using a lexical device to attack, explicitly and directly, a specific target. I might argue that I was just joking or that my intention was just mockery or some kind of irony. This argument built on non-offensive intentions, however, cannot be used as an excuse or as an attempt to *normalise* hate speech. Words have their effect independently of what the speakers claim about their intentions: words can abuse others by ridiculing, hurting, and humiliating them, and are therefore powerful tools of oppression and aggression. Indeed, by uttering the sentence above, I am not simply expressing my opinion or an evaluation, I am producing a speech act that conveys offence, aggression, and denigration, not only towards an individual but also towards a group of people who can be identified on the basis of a specific sexual identity or orientation.³

3 On the context of the utterance as an essential element to understand the illocutionary dimension of the speech act, see Meza, Vincze, and Mogoş (2018), Baider (2020) and Culpeper (2021). On sarcasm and humour as forms of implicit or covert hate speech, see Bansal et al. (2020), Frenda et al. (2022). Note that, in this introductory chapter, we roughly use the term ‘implicit hate speech’ as a synonym of ‘covert hate speech’, but implicit hate speech could in fact be understood as limited to the sub-type of covert hate speech that is associated with implicit meanings such as presuppositions and implicatures (see [Chapter 7](#) in this volume).

But what happens if I use demeaning and divisive communication while carefully avoiding slurs or derogatory terms (for instance, in order not to incur sanctions)? An example could be political slogans that refer to nations or nationality such as ‘Britain first!’ or ‘Prima gli italiani!’ (Italians first). In fact, when uttering such expressions, superficially I am not attacking or insulting anybody. In reality, however, these are still messages of hate, which incite intolerance and prejudice towards other social groups by distinguishing and contrasting groups on the basis of nationality. Even though the social groups who are discriminated against are not explicitly mentioned, their identity can be gathered from the context, in the same way as it can be inferred that these *other* groups are to be placed in a position of subordination or inferiority with respect to the group that is prominently singled out in the slogan.

Implicit hate speech can be associated with various speech acts, beyond insults, and with a range of implicit meanings, such as presuppositions and implicatures, which may contribute to the construction or reinforcement of stereotypes (see Lombardi Vallauri 2019a, 2019b; see also [Chapter 7](#) in this volume). The discussion of political slogans like ‘Britain first!’ has already provided an example of the use of items that trigger often unwarranted presuppositions, such as the existence of a ranking: the use of ‘first’ presupposes that there is going to be a second, third, and so on—that is, a hierarchy of social groups is necessarily superimposed in the interpretation.

As for implicatures, consider the examples in (1):

- (1) a. Some Italians pay taxes.
- b. Finns are honest, but Italians cook well.

Both examples are cases in which, in normal communication, the interpretation of the utterance is crucially enriched by implicit meaning, which is triggered in a very systematic way. In (1a) the use of the quantifier ‘some’ triggers an implicit comparison with a stronger quantifier, ‘all’, which the speaker could have used but

chose not to. From this scale of strength among possible alternatives of expression, the hearer will extract the non-literal meaning that not all Italians pay taxes. Note that, strictly logically, the utterance in (1a) is compatible with a scenario in which all Italians pay taxes, and does not deny it explicitly. However, well-studied mechanisms of human communication lead the hearer to interpret (1a) by enriching its literal meaning with a so-called scalar implicature ('if some, then not all'), with the effect of contributing to the stereotype that many Italians are tax evaders. This mechanism thus has the potential to generate implicit offensive speech.

Similarly, the conventional implicature evoked by the contrastive connective 'but' in (1b) conveys the implicit meaning that the opposite of what is predicated of one nationality holds for the other, although this is not stated literally. Finns and Italians are compared and a different quality is attributed to each nationality; the fact that a contrast is established by means of 'but' generates the implicit meaning that it cannot be said that Italian are honest, hence they are not, and it cannot be said that Finns cook well, hence they do not.

While the implicit meaning generated by conversational dynamics does not go beyond reinforcing and propagating stereotypes in the examples in (1), the effects of these strategies, in the appropriate context, can lead to hate speech proper. Consider:

- (2) Calano fatturato e Pil ma aumentano i gay.

'Sales volume and GDP go down but the number of gays increases.'

The sentence in (2) is the headline of a front-page article published on 23 January 2019 in the right-wing Italian newspaper *Libero*. The headline brings together Italian economic trends in sales and gross domestic product, and the number of people who identify as gay. At first sight, the two facts strike the reader as completely unrelated. However, similarly to (1b), the contrastive connective *ma* (but), by contrasting the two facts, establishes a connection: it suggests that there is a relationship between the two trends, but

leaves it to the reader to figure out what that relationship is. The only way to make sense of the contrast in (2) is to extract from the two conjuncts a value judgement: the decrease in the economic trends is bad; ‘but’ leads to the expectation that the contrasted fact is, instead, good. This, in turn, forces the reader to recognise irony in the message: clearly, the author does not see the increase in the numbers of gay people as a positive thing that can counterbalance the negative trend. In fact, the opposite is true: what the author wants to imply is that it is a further sign of a decaying society. This means that not only is the result a message of hate, but the author of that message can hide behind its implicit nature, which emerges only because of the reader’s pragmatic inferences.

The construction of stereotypes or other forms of generalised prejudices can also be expressed through various grammatical means used for referential purposes: for instance, the contrastive use of personal pronouns (‘*we* Italians’, ‘*you* immigrants’), or the use of generic pronouns to refer to an unspecified group (or subgroup) of individuals (e.g. ‘*they* are lazy and never work’), or the derogative use of demonstratives (‘*this* immigrant’, ‘*these* people’) (see Fumagalli 2019; [Chapter 8](#) in this volume). The following is an example from a Facebook post, discussed in Fumagalli (2019, 66), where repeated reference is made to *loro* (they) in opposition to *noi* (we) in the context of racist comments:

- (3) Ma quale odio razziale ... l’odio razziale è da parte loro verso noi. Sono loro che rifiutano i nostri usi e costumi. Sono loro che rifiutano di integrarsi. Sono ancora loro che pretendono, pretendono, pretendono, senza un minimo di riconoscenza.

‘What racial hate ... racial hate is by them towards us. It is they who reject our customs and traditions. It is they who reject integration. It is yet they who demand, demand, demand, without any gratitude whatsoever.’

The construction of authority and subordination (see Bianchi 2017; Langton 2018) may also arise from linguistic devices such as

the use of prohibition and obligation modals (e.g. ‘foreigners must/cannot...’), the exclusive adverb ‘only’ (e.g. ‘only Italians can...’), inclusive pronouns to refer to unspecified groups or parties (e.g. ‘we all know this’, ‘we must act now’), high-degree/-intensity adverbs and adjectives (e.g. ‘entire(ly)’, ‘total(ly)’, ‘absolute(ly)’), and evidential and epistemic adverbs and adjectives (e.g. ‘evident(ly)’, ‘obvious(ly)’, ‘undoubted(ly)’, ‘undisputed(ly)’). All these authoritative strategies have the power to legitimise beliefs and behaviours and to influence perception of reality.⁴

The use of tools that express an opinion or an evaluation in a parenthetical form are also relevant here. These ‘asides’ are apparently not at the centre of the communicative exchange, but this is precisely why they are particularly insidious since they tend to be more passively accepted by the hearer together with the main core of the message (see Lombardi Vallauri 2000). This is true, for instance, of evaluative adverbs and adverbials (e.g. ‘fortunately’, ‘unfortunately’, ‘unluckily’, ‘regrettably’; for instance, ‘Regrettably, many [of a particular minority group] live in this neighbourhood’) and of special evaluative constructions with comparatives (e.g. *Lei è più bella che intelligente*, ‘She’s more beautiful than intelligent’: apparently a compliment, which becomes an offence when addressed to someone who is not characterised by outstanding beauty).

In terms of intensity and degree of offensiveness, implicit hate speech is certainly weaker than explicit hate speech. Consider the examples in (1), for instance, particularly compared to ethnic slurs or offensive terms related to gender or sexual orientation. Many of us would agree that the statements in (1), especially (1b), can be considered only slightly offensive. However, the role of context is fundamental: if sentences similar to (1a) were uttered in a context (say, a country) in which Italians constituted an immigrant minority that has historically been attacked and accused of

4 See also the contributions collected in Knoblock (2022) for the exploitation of morphosyntactic features, such as word formation strategies, to convey offensive messages.

criminal practices and behaviour, they could have stronger repercussions on the audience, even more so if uttered by a public person with some sort of authority.

Because of its intrinsic nature as a hidden message, implicit hate speech can also be subject to different interpretations and perceptions. It would therefore be difficult, if not totally impossible, to subject it to national or international regulations—this might not even be a desideratum from a legislative viewpoint. However, it is precisely because of its invisible and disguised character that implicit hate speech is employed and exploited in propaganda and with persuasive purposes, both in mass media and in political debate. Since the addressees of these messages very often process and accept the implicit meanings automatically and unconsciously, these communicative expedients have the power to justify and legitimise beliefs and behaviours and to influence our perception of reality. A linguistic analysis of both the explicit and implicit strategies involved in hatred discourse can therefore help us to identify hate speech in all its manifestations, including its most subtle and hidden forms and expressions.

The concepts of explicit and implicit hate speech come very close to, but are not interchangeable with, the notion of hard and soft hate speech, as applied for instance in Baider, Assimakopoulos, and Millar (2017) (see also Assimakopoulos 2020). Under hard hate speech we find all those manifestations that are prosecutable by law because they explicitly incite discriminatory hatred. Soft hate speech, instead, is not prosecutable in legal terms, because it does not explicitly manifest incitement to hatred, although it expresses prejudice and intolerance, and, as such, is capable of considerable harmful impact. Here, implicitness and explicitness refer to the intention to incite hatred, and are therefore tied to a differential legal treatment. In our linguistically minded investigation, instead, implicitness and explicitness refer to the linguistic means that are adopted in conveying discriminatory messages. This linguistic notion of explicitness and implicitness can also more generally be applied to aggressive language, and is not univocally linked to a differential legal status. For instance, there can

be cases of implicit hate speech that qualify as hard hate speech under the above definition: imagine a situation in which a term like ‘daisy’ assumes a derogatory meaning in the discourse of an in-group (see further [Section 1.4](#)); in such a situation, an utterance like ‘We must eradicate the daisies’ becomes an explicit incitement to violence against a whole group, although the target of hatred will only be recognisable within the in-group. The linguistic means are implicit, but within the in-group the intention to incite hatred is explicit.

1.4 Haters and hated, intentions and effects

Since Paul Grice’s ([1975] 1989) distinction between two levels of meaning—*what is said* and *what is meant*—speaker intentions have been assumed to play an essential role in the interpretation of a linguistic utterance. Indeed, the pragmatic level of what is meant involves aspects of meaning that are drawn from the context in which the utterance occurred, so the process of interpretation that leads from *what is said* to *what is meant* can be viewed as an inferential process that is based on principled, pragmatic mechanisms, and that also relies on reasoning about speaker intentions.

In the debate on the boundaries between hate speech and freedom of speech, context-dependency—and especially, speaker intentions—are often put forward as a reason to deny the existence of hate speech in particular situations. Comedians may justify potentially offensive expressions as irony, in the same way as politicians may sanction the use of historically and culturally *loaded* words as a way to demonstrate their aversion towards the silencing that is allegedly imposed by the censorship of the *politically correct* or their rights to freely express their opinion and judgements. They may simply claim that they have no intention of verbally attacking or offending any group or individual, so there is no hate speech. According to this view, hate speech is a slippery concept that depends on the context, so potentially offensive hate expressions change their meaning and their impact depending on who uses them and how they are used.

On the one hand, it is true that insults and other derogatory terms are context-dependent. One example of this is the reappropriation of slurs through which a group reclaims words that have been used in an offensive and discriminatory way against that group (see chapters 2 and 6 in this volume and references therein). Another interesting case is represented by so-called *dog-whistles*—that is, apparently neutral terms (e.g. ‘inner-city’) that assume an ideologically loaded interpretation for an in-group (in the case of ‘inner-city’, the stereotypical, negatively connotated reference to African American neighbourhoods; see Henderson and McCready 2019). On the other hand, the effects of a word or a linguistic expression are not entirely tied to the speaker’s intentions. Independently of the claimed intentions and of the appeal to free speech, in choosing to use a slur instead of a neutral equivalent, the speakers signal—even unconsciously—that they endorse the term, its connotations, and its associations. It might be a joke for the speaker, but it will still vilify the target. Indeed, it has long been acknowledged that language has a performative function that may be independent from speaker intentions: it does not simply describe the world but is also able to effect changes in the world.

The distinction between constative language, which describes the world and can be evaluated in terms of true or false statements, and performative language, which *does things with words*, was first described by the philosopher John L. Austin (1962). For Austin, the performative function of language includes speech acts such as swearing, promising, betting, and officiating a wedding ceremony. Performative language is not completely independent of the context and the situation of the utterance; indeed, Austin identifies a number of felicity conditions that must be met in order for an utterance to be used performatively. However, the effects and consequences that a word or an utterance can have on the interlocutor or on a listener—the so-called perlocutive effects— may or may not coincide with the intentions of the speaker: *‘Saying something will often, or even normally, produce certain consequential effects upon the feelings, thoughts, or actions of the audience, or of the speaker, or of other persons: and it may be done with the design,*

intention, or purpose of producing them' (Austin 1962: 101). But, as implied in this quotation ('often', 'normally'), it may also be done without that design, intention, or purpose.

Performativity of language and its perlocutive effects have clearly played a prominent role in the scholarly debates on the pragmatic and social functions of languages (see Culpeper 2021 and references therein). Language is not only the mirror of society in its descriptive function, reflecting hierarchies, social injustices, conflicts, discriminations, classifications, and divisions; in its performative power, it is also able to generate and transform social identities, connections, and practices by creating, reinforcing, or removing classifications, hierarchies, and conflicts. In this sense, language performativity plays a key role in the construction of human and social identities, such as gender (Butler 1990, 1993, 1997; Bianchi 2021; [Chapter 2](#) in this volume).

1.5 The volume and its structure

This volume undertakes an investigation of hate speech that tackles the linguistic strategies adopted by speakers when they employ language as a means of aggression. The authors apply established methods of data collection and analysis to a novel body of evidence, specifically collected for their studies, primarily from social media and other forms of public speech, with the aim of identifying the common linguistic characteristics of aggressive language and hate speech in different domains of communication. Most of the chapters deal directly with hate speech, while some address issues related to aggressive language, either with respect to their differing offensive potential (e.g. [Chapter 3](#), which analyses different types of insults) or in comparison with hate speech (e.g. [Chapter 9](#), which examines the contextual conditions that favour the emergence of excessive language and hate speech).⁵

5 On the distinction between aggressive language and hate speech, see [Section 1.2](#).

The focus on Italian as an empirical testbed by our international group of authors offers both practical and methodological advantages in terms of comparability of data from different domains of communication and of cohesion in the research questions addressed by the various chapters. The resulting picture will pave the way for future research on other languages, with the aim of identifying cross-linguistic and language-specific strategies and constructions. Although our investigation has a primarily linguistic perspective, the authors contribute their expertise and strengths in different linguistic areas to arrive at a multifaceted and interdisciplinary approach, with insights from language variation, dialectology, language and migration, multilingualism, semiotics, language education, bilingualism, language development and cognition, and computational linguistics. Observations from or about other disciplines are also often present, especially philosophy of language, political studies, and sociology.

The volume is divided into two parts, preceded by this introductory chapter written by the editors and by an invited position paper in [Chapter 2](#), ‘Call me by my name: hate speech and identity’, written by Claudia Bianchi, a leading figure in the international debate on hate speech at the crossroads between linguistics and philosophy of language. The first of the two main parts of the book concerns the interaction of lexical strategies and context in hate speech, while the second part enlarges the scope by including grammatical and pragmatic analyses. Because of the fluidity and interconnectedness of the themes discussed, however, this categorisation and subdivision of the chapters should not be understood in a rigid way. Indeed, some chapters within each part often overlap in their examination of lexical, contextual, and pragmatic strategies, blurring the delineation between sections.

Three contributions in the first part discuss data from social media in two particularly polarising domains: political debate and homotransphobic discourse. [Chapter 3](#) by Borreguero Zuloaga focuses on contextual insults, which require more than lexical items to be performed and rely on the relevance of cultural knowledge and stereotypes. Meanwhile, [Chapter 4](#) by Safina

and [Chapter 5](#) by De Pascale, Cavarani, and Marzo present data from different social media that show the mechanisms by which homotransphobia emerges from lexical and discursive tools, identifying clusters of meanings that, although not necessarily connected to sexual aspects, contribute to the propagation of linguistic aggressions in this domain. A fourth contribution, [Chapter 6](#) by Zingaretti, Garraffa, and Sorace, addresses the dimension of bilingualism in connection with the perception of hate speech, in particular with respect to slur appropriation, by employing a specifically designed questionnaire.

The second part begins with [Chapter 7](#) by Retta, which takes advantage of the new discourse context provided by the COVID-19 pandemic to analyse the emergence of xenophobic and racist discourse in social media, identifying specific argumentative strategies. The study by Paleta and Dyda ([Chapter 8](#)) focuses on linguistic aggressions performed by lexical and pragmatic means, especially pronouns, in a Facebook group comprising expatriates. The last chapter in this part, [Chapter 9](#) by Bianchi and Cruschina, investigates the use of questions in online political debate, showing how they can help to distinguish between cooperative and non-cooperative contexts.

References

- Anderson, Luvell, and Michael Barnes. 2022. 'Hate speech'. *The Stanford Encyclopedia of Philosophy* (Spring 2022 edition), edited by Edward N. Zalta. <https://plato.stanford.edu/archives/spr2022/entries/hate-speech/>.
- Assimakopoulos, Stavros. 2020. 'Incitement to discriminatory hatred, illocution and perlocution'. *Pragmatics and Society* 11 (2): 177–195. <https://doi.org/10.1075/ps.18071.ass>.
- Austin, John Langshaw. 1962. *How to Do Things with Words*. Cambridge, MA: Harvard University Press.
- Baider, Fabienne. 2023. 'Accountability issues, online covert hate speech, and the efficacy of counter-speech'. *Politics and Governance* 11 (2): 249–260. <https://doi.org/10.17645/pag.v11i2.6465>.
- 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>.

- Baider, Fabienne. 2019. 'Le Discours de haine dissimulée: le mépris pour humilier'. *Déviance et Société* 43 (3): 359–387. <https://doi.org/10.3917/ds.433.0359>.
- Baider, Fabienne, Stavros Assimakopoulos, and Sharon Millar. 2017. 'Hate speech in the EU and the C.O.N.T.A.C.T. project'. In *Online Hate Speech in the European Union: A Discourse-Analytic Perspective*, edited by Stavros Assimakopoulos, Fabienne Baider, and Sharon Millar, 1–6. Cham: Springer.
- Baider, Fabienne, Sharon Millar, and Stavros Assimakopoulos, eds. 2020. 'Hate speech: definitions, interpretations and practices'. Special issue, *Pragmatics and Society* 11 (2). <https://doi.org/10.1075/ps.11.2>.
- Bansal, Srijan, Vishal Garimella, Ayush Suhane, Jasabanta Patro, and Animesh Mukherjee. 2020. 'Code-switching patterns can be an effective route to improve performance of downstream NLP applications: a case study of humour, sarcasm and hate speech detection'. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, edited by Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault, 1018–1023. Kerrville, TX: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-main.96>.
- Ben-David, Anat, and Ariadna Matamoros Fernández. 2016. 'Hate speech and covert discrimination on social media: monitoring the Facebook pages of extreme-right political parties in Spain'. *International Journal of Communication* 10: 1167–1193.
- Bhat, Prashanth, and Ofra Klein. 2020. 'Covert hate speech: white nationalists and dog whistle communication on Twitter'. In *Twitter, the Public Sphere, and the Chaos of Online Deliberation*, edited by Gwen Bouvier and Judith E. Rosenbaum, 151–172. Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-030-41421-4_7.
- Bianchi, Claudia. 2021. *Hate speech: il lato oscuro del linguaggio*. Bari and Rome: Laterza.
- Bianchi, Claudia. 2018. 'Perspectives and slurs'. In *Beyond Semantics and Pragmatics*, edited by Gerhard Preyer, 187–198. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780198791492.003.0011>.
- Bianchi, Claudia. 2017. 'Linguaggio d'odio, autorità e ingiustizia discorsiva'. *Rivista di Estetica* 64: 18–34. <https://doi.org/10.4000/estetica.2059>.
- Bianchi, Claudia. 2015. 'Parole come pietre: atti linguistici e subordinazione'. *Esercizi Filosofici* 10 (2): 115–135.
- Bianchi, Claudia. 2014. 'Slurs and appropriation: an echoic account'. *Journal of Pragmatics* 66 (May): 35–44. <https://doi.org/10.1016/j.pragma.2014.02.009>.
- Bleich, Erik. 2011. *The Freedom to be Racist? How the United States and Europe Struggle to Preserve Freedom and Combat Racism*. Oxford: Oxford University Press.

- Bolinger, Renée Jorgensen. 2017. 'The pragmatics of slurs'. *Nous* 51 (3): 439–462. <https://doi.org/10.1111/nous.12090>.
- Boromisza-Habashi, David. 2021. 'There's no such thing as hate speech and it's a good thing, too'. In *Digital Hate: The Global Conjecture of Extreme Speech*, edited by Sahana Udupa, Iginio Gagliardone, and Peter Hervik, 23–33. Bloomington, IN: Indiana University Press.
- Brambilla, Marina, and Valentina Crestani. 2021. "'Der kriminelle ...", "I soliti ...": parole dell'odio in tedesco e in italiano'. In *Il linguaggio dell'odio: fra memoria e attualità*, edited by Marilisa D'Amico, Marilisa Brambilla, Valentina Crestani, and Nannerel Fiano, 85–105. Milan: Franco Angeli.
- Brontsema, Robin. 2004. 'A queer revolution: reconceptualizing the debate over linguistic reclamation'. *Colorado Research in Linguistics*, 17. <https://doi.org/10.25810/dky3-zq57>.
- Brown, Alexander. 2017. 'What is hate speech? Part 1: The myth of hate'. *Law and Philosophy* 36 (4): 419–468. <https://doi.org/10.1007/s10982-017-9297-1>.
- Butler, Judith. 1997. *Excitable Speech: A Politics of the Performative*. London and New York: Routledge.
- Butler, Judith. 1993. *Bodies That Matter: On the Discursive Limits of "Sex"*. London and New York: Routledge.
- Butler, Judith. 1990. *Gender Trouble: Feminism and the Subversion of Identity*. London and New York: Routledge.
- Capone, Alessandro. 2014. 'Indirect reports, slurs, and the polyphonic speaker'. *Reti, Saperi, Linguaggi* 2: 301–318.
- Caselli, Tommaso, Valerio Basile, Jelena Mitrović, Inga Kartoziya, and Michael Granitzer. 2020. 'I feel offended, don't be abusive! Implicit/explicit messages in offensive and abusive language'. Paper presented at *LREC 2020 – 12th International Conference on Language Resources and Evaluation*, Conference Proceedings, 6193–6202. Marseille: European Language Resources Association.
- Cepollaro, Bianca. 2017. 'The Semantics and Pragmatics of Slurs and Thick Terms'. PhD thesis, Paris Sciences et Lettres Research University.
- Cepollaro, Bianca. 2015. 'In defense of a presuppositional account of slurs'. *Language Sciences* 52: 36–45. <https://doi.org/10.1016/j.langsci.2014.11.004>.
- Council of Europe. 1997. 'Recommendation No. R (97) 20 of the Committee of Ministers to Member States on "Hate Speech"', 30 October 1997. Accessed 19 July 2024. <https://rm.coe.int/1680505d5b>.
- Croom, Adam. 2013. 'How to do things with slurs: studies in the way of derogatory words'. *Language and Communication* 33: 177–204. <https://doi.org/10.1016/j.langcom.2013.03.008>.
- Culpeper, Jonathan. 2021. 'Impoliteness and hate speech: compare and contrast'. *Journal of Pragmatics* 179: 4–11. <https://doi.org/10.1016/j.pragma.2021.04.019>.

- Daniels, Jessie. 2008. 'Race, civil rights, and hate speech in the digital era'. In *Learning Race and Ethnicity: Youth and Digital Media*, edited by Anna Everett, 129–154. Cambridge, MA: MIT Press.
- Ermida, Isabel, ed. 2023a. *Hate Speech in Social Media: Linguistic Approaches*. Cham: Palgrave Macmillan.
- Ermida, Isabel. 2023b. 'Distinguishing online hate speech from aggressive speech: a five-factor annotation model'. In Ermida, *Hate Speech in Social Media*, 35–75. https://doi.org/10.1007/978-3-031-38248-2_2.
- Esposito, Eleonora and Majid KhosraviNik, eds. 2024. *Discourse in the Digital Age: Social Media, Power, and Society*. London: Routledge.
- Faloppa, Federico. 2020. *#ODIO: manuale di resistenza alla violenza delle parole*. Turin: UTET.
- Faloppa, Federico. 2012. *Razzisti a parole (per tacer dei fatti)*. Bari: Laterza.
- Faloppa, Federico. 2004. *Parole contro: la rappresentazione del 'diverso' nella lingua italiana e nei dialetti*. Milan: Garzanti.
- Fish, Stanley. 1994. *There's No Such Thing as Free Speech and It's a Good Thing, Too*. New York: Oxford University Press.
- Fortuna, Paula, and Sérgio Nunes. 2018. 'A survey on automatic detection of hate speech in text'. *ACM Computing Surveys* 51 (4): Article no. 85. <https://doi.org/10.1145/3232676>.
- Foxman, Abraham H., and Christopher Wolf. 2013. *Viral Hate: Containing Its Spread on the Internet*. New York: St. Martin's Press.
- Frenda, Simona, Alessandra Teresa Cignarella, Valerio Basile, Cristina Bosco, Viviana Patti, and Paolo Rosso. 2022. 'The unbearable hurtfulness of sarcasm'. *Expert Systems with Applications* 193: 116398. <https://doi.org/10.1016/j.eswa.2021.116398>.
- Fumagalli, Corrado. 2019. 'Discorsi d'odio come pratiche ordinarie'. *Biblioteca della libertà* 224: 55–75. https://doi.org/10.23827/BDL_2019_1_3.
- Gagliardone, Iginio, Danit Gal, Thiago Alves, and Gabriela Martinez. 2015. *Countering Online Hate Speech*. Paris: UNESCO Publishing.
- Galinsky, Adam D., Kurt Hugenberg, Carla Groom, and Galen V. Bodenhausen. 2003. 'The reappropriation of stigmatizing labels: implications for social identity'. In *Identity Issues in Groups*, edited by Jeffrey Polzer, 221–256. Bingley: Emerald Group Publishing Limited. [https://doi.org/10.1016/S1534-0856\(02\)05009-0](https://doi.org/10.1016/S1534-0856(02)05009-0).
- Gao, Lei, Alexis Kuppersmith, and Ruihong Huang. 2017. 'Recognizing explicit and implicit hate speech using a weakly supervised two-path bootstrapping approach'. *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, retrieved from *Computation and Language* (cs.CL). <https://doi.org/10.48550/arXiv.1710.07394>.
- Gelber, Katharine, and Luke McNamara. 2016. 'Evidencing the harms of hate speech'. *Social Identities* 22 (3): 324–341. <https://doi.org/10.1080/13504630.2015.1128810>.

- Grice, H. Paul. 1989. 'Logic and conversation.' In *Studies in the Way of Words*, 22–40. Cambridge, MA, and London: Harvard University Press. Originally published in 1975 in *Syntax and Semantics*, vol. 3: *Speech Acts*, edited by Peter Cole and Jerry L. Morgan, 41–48. New York: Academic Press.
- Guillén-Nieto, Victoria. 2023. *Hate Speech: Linguistic Perspectives*. Berlin: De Gruyter.
- Henderson, Robert, and Elin McCready. 2019. 'Dogwhistles and the at-issue/non-at-issue distinction.' In *Secondary Content: The Semantics and Pragmatics of Side Issues*, edited by Daniel Gutzmann and Katharina Turgay, 222–245. Leiden: Brill. https://doi.org/10.1163/9789004393127_010.
- Herz, Michael, and Peter Molnar, eds. 2012. *The Content and Context of Hate Speech: Regulation and Responses*. Cambridge: Cambridge University Press.
- Knoblock, Natalia, ed. 2022. *The Grammar of Hate: Morphosyntactic Features of Hateful, Aggressive, and Dehumanizing Discourse*. Cambridge and New York: Cambridge University Press.
- Langton, Rae. 2018. 'The authority of hate speech.' In *Oxford Studies in Philosophy of Law*, vol. 3, edited by John Gardner, Leslie Green, and Brian Leiter, 123–152. Oxford: Oxford University Press.
- Leets, Laura. 2002. 'Experiencing hate speech: perceptions and responses to anti-semitism and antigay speech.' *Journal of Social Issues* 58 (2): 341–361. <https://doi.org/10.1111/1540-4560.00264>.
- Lillian, Donna L. 2007. 'A thorn by any other name: sexist discourse as hate speech.' *Discourse and Society* 18 (6): 719–740.
- Lombardi Vallauri, Edoardo. 2019a. 'Implication and power of choice.' *Rivista italiana di filosofia del linguaggio* 13 (2): 73–93.
- Lombardi Vallauri, Edoardo. 2019b. *La lingua disonesta: contenuti impliciti e strategie di persuasione*. Bologna: Il Mulino.
- Lombardi Vallauri, Edoardo. 2000. *Grammatica funzionale delle avverbiali italiane*. Rome: Carocci.
- Määttä, Simo. 2023. 'Linguistic and discursive properties of hate speech and speech facilitating the expression of hatred: evidence from Finnish and French online discussion boards.' *Internet Pragmatics*, online-first article. <https://doi.org/10.1075/ip.00094.maa>.
- Määttä, Simo. 2020. 'Syrjivän kielenkäytön rajoilla.' *Niin&Näin* 3: 71–77.
- Maitra, Ishani, and Mary Kate McGowan. 2009. 'On racist hate speech and the scope of a free speech principle.' *Canadian Journal of Law and Jurisprudence* 23 (2): 343–372. <https://doi.org/10.1017/S0841820900004963>.
- Meza, Radu, Hanna-Orsolya Vincze, and Andreea Mogoş. 2018. 'Targets of online hate speech in context: a comparative digital social science analysis of comments on public Facebook pages from Romania and Hungary.' *Intersections. East European Journal of Society and Politics* 4: 26–50. <https://doi.org/10.17356/ieejsp.v4i4.503>.

- Parekh, Bhikhu. 2006. 'Hate speech: is there a case for banning?' *Public Policy Research* 12 (4): 213–223. <https://doi.org/10.1111/j.1070-3535.2005.00405.x>.
- Perry, Barbara. 2001. *In the Name of Hate: Understanding Hate Crimes*. London and New York: Routledge.
- Retta, Mattia. 2023. 'A pragmatic and discourse analysis of hate words on social media.' *Internet Pragmatics* 6 (2): 197–218. <https://doi.org/10.1075/ip.00096.ret>.
- Richardson, Louise. 2018. 'Woman-hating: on misogyny, sexism, and hate speech.' *Hypatia: A Journal of Feminist Philosophy* 33 (2): 256–272. <https://doi.org/10.1111/hypa.12398>.
- Schmidt, Anna, and Michael Wiegand. 2017. 'A survey on hate speech detection using natural language processing.' *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*, 1–10. Valencia: Association for Computational Linguistics. <https://doi.org/10.18653/v1/W17-1101>.
- Sternberg, Robert J. 2004. *The Psychology of Hate*. Washington, DC: American Psychological Association.
- United Nations. 2019. 'United Nations strategy and plan of action on hate speech'. Accessed 19 July 2024. <https://www.un.org/en/genocideprevention/documents/UN%20Strategy%20and%20Plan%20of%20Action%20on%20Hate%20Speech%2018%20June%20SYNOPSIS.pdf>.
- Van Blarcum, Christopher D. 2005. 'Internet hate speech: the European framework and the emerging American haven.' *Washington and Lee Law Review* 62 (2): 781–830.
- Waldron, Jeremy. 2012. *The Harm in Hate Speech*. Cambridge, MA: Harvard University Press.