

Preface

The linguistics of hate speech

Matsuda (1989) was the first legal scholar to use the term *hate speech*. Although Matsuda's initial examples only referred to *racial hate speech*, she clarified that, in her judgement, racial hate speech is not the only form of hate speech and that *anti-gay* and *anti-lesbian hate speech* are similar phenomena but deserve independent attention.

The meaning of the term hate speech is rather opaque, although, at first sight, it gives the opposite impression. Looking at the semantics of its constituent parts – that is, hate and speech – one may think that the term describes a subcategory of speech associated with the expression of hate or hatred towards people in general. However, we know by experience that its use is neither limited to speech nor to the expression of hatred. We also know that the target is not the general public but, instead, members of groups or classes of people identifiable by *legally-protected characteristics*, such as *race, ethnicity, religion, sex, sexual orientation, disability*,¹ amongst others. Therefore, the meaning of hate speech is not a function of the literal meanings of its constituent parts. On the contrary, it has multiple meanings, as suggested by Brown (2017a; 2017b).

The social phenomenon and the legal concept of hate speech are necessarily intertwined. As a social phenomenon, hate speech fuels broad-scale conflicts that may cause a breach of the peace or even create environments conducive to hate crimes. As a legal concept, hate speech is an abstract endangerment statute because it punishes not actual but hypothetical creation of social risk, and must find a balance between the fundamental rights to freedom of opinion and expression and dignity.

The analysis of hate speech is well documented in law, sociology and media and communication studies. For instance, in the field of law, Brown (2017a; 2017b) analysed the multiple definitions of hate speech; Tsesis (2002) examined how hate speech paves the way for harmful social movements and highlighted the destructive power of hate propaganda; Delgado (1982), Matsuda (1989) and Delgado and Stefancic (2018[2004]; 2018) claimed that the right to freedom of opinion and expression should be constrained by United States constitutional law when individuals abuse this right to shame, cause despair, inflict injury, threaten and harm members of groups or classes of people identifiable by legally-protected characteristics. Actionable hate speech within the legal framework of Spanish

¹ Hall (2019) is one of the few studies on *disability hate speech*.

civil law was analysed by Dolz Lago (2015) and Landa-Gorostiza (2018). Disinformation and hate speech were studied, from an European Union constitutional law perspective, by Pitruzzella and Pollicino (2020). And Brown and Sinclair (2020) analysed the complex relationship between politics and hate speech laws.

Over the last few years, online hate speech has gained increasing interest amongst scholars in the fields of sociology and communication and media studies, as shown in the latest publications on the topic: *Online othering: Exploring digital violence and discrimination on the web* (Lumsden & Harmer 2019); *Digital hate: The global conjuncture of extreme speech* (Udupa, Gagliardone & Hervik 2021); *Social media and hate* (Banaji & Bhat 2022); and *Cyberhate in the context of migrations* (Monnier, Boursier & Seoana 2022).

On reviewing the literature, we find that while most of the research on hate speech principally comes from the social sciences, hate speech has not yet received sufficient attention as a scientific object of study in linguistics.² Only over the last few years can one see a marked turn in linguistics, especially in pragmatics, towards the analysis of language crimes, with special emphasis on (online) hate speech. The following references show that the linguistic analysis of hate speech is currently experiencing a boom: *A corpus linguistic analysis of white supremacist language* (Brindle 2016); *Language and violence: Pragmatic perspectives* (Silva 2017: 141–168); *Online hate speech in the EU: A discourse-analytic perspective* (Assimakopoulos, Baider & Millar 2017). This latter study, by Assimakopoulos, Baider and Millar, presents the findings of C.O.N.T.A.C.T., a project (2015–2017) co-funded by the Rights, Equality and Citizenship Programme of the European Commission Directorate-General for Justice and Consumers (JUST/2014/RRAC/AG) and coordinated by Baider at the University of Cyprus. Their project investigates online hate speech in the context of migration in the European Union from a critical discourse analysis perspective. The special issue of the journal *Lodz Papers in Pragmatics* (2018), edited by Baider and Kopytowska, focused on both hate speech and countering hate speech. In this issue, hate speech was analysed from different linguistic perspectives, such as pragmatics (Technau 2018: 25–43) and discourse analysis (Strani & Szczepaniak-Kozak 2018: 163–179), and approaches, especially corpus linguistics (Ruzaitė 2018: 93–116). Millar (2019: 144–163) explored online hate speech and its social control from institutional and corporate perspectives.

² In previous work, the present author made a similar claim regarding the analysis of other language crimes, such as defamation (Guillén-Nieto 2020), sexual harassment (Guillén-Nieto 2021) and workplace harassment (Guillén-Nieto 2022).

Linguists have also taken a step forward in providing technical services to the security forces and agents of internet platforms to detect and prevent online hate speech. In such cases, the methodology combines a corpus-based approach, deep learning and the qualitative linguistic analysis of lexical and grammatical indicators of hate speech. Some remarkable contributions to this cutting-edge research are represented by Becker (2020; 2021), who investigates antisemitism and the challenges that implicitness raises for automatic online hate speech detection; Bick (2020; 2021), who conducts computer-based research on the annotation of non-direct forms of hate speech, such as emoticons and emojis, in a German-Danish social media corpus;³ and Baider (2020; 2022), who takes a legal-linguistic perspective on covert hate speech, combining The Rabat Plan of Action criteria with pragmatic analysis. Her work also draws attention to strategies for countering hate speech other than censoring it, as an efficient way to combat covert hate speech.

Lastly, *The grammar of hate: Morphosyntactic features of hateful, aggressive, and dehumanising discourse*, edited by Knoblock (2022), departs from traditional lexical and discursive approaches to hate speech detection. It focuses, instead, on the morphosyntactic features that the hate-advocating speaker appropriates, manipulates and exploits to express hate, hostility or violence toward the targets. The collection of chapters in this volume demonstrates how hate speech manifests itself in a wide array of grammatical features, such as morphology (Mattielo 2022: 34–58; Tarasova & Sánchez Fajardo 2022: 59–81), word formation (Beliaeva 2022: 177–196), gender (Lind & Nübling 2022: 118–139; Thál & Elmerot 2022: 97–117), pronouns (Flores Ohlson 2022: 161–176; Peterson 2022: 262–287), imperative verbs (Bianchi 2022: 222–240) and syntactic patterns (Geyer, Bick & Kleene 2022: 241–261), to name a few examples.

Why has hate speech not been studied in linguistics until recently? Linguistic theories have tended to address language as cooperative action (Grice 1975) geared to reciprocally informative polite understanding (Lakoff 1973; Brown & Levinson 1987 [1978]; Leech 1983). As a result of this idealised view of language, negative types of discourse and speech acts, such as defamation (Tiersma 1987; Shuy 2010; Guillén-Nieto 2020), harassment (Guillén-Nieto 2021; 2022; Stein 2022) and hate speech, have been traditionally cast aside as objects of linguistic study. Some linguists, such as Leech (1983), referred to impolite language as

³ The corpus has been developed within the framework of XPEROHS Project (2019–). *Towards a balance and boundaries in public discourse: Expressing and perceiving online hate speech*. Project leader: Klaus Geyer. <https://xperohs.sdu.dk/publications/> (accessed 31 July 2022).

unusual, anomalous or deviant when it is, in effect, widespread. Since the 1990s, Culpeper (1996; 2005; 2008; 2011; 2012), Culpeper and Terkourafi (2016) and Kaul de Marlangeon (1993; 2005; 2008; 2014), amongst other linguists, have paved the way for the analysis of the deliberate use of language to offend. At present, the analysis of offensive communication is still making a place for itself within linguistic research. I concur with Knoblock (2022: 5) on the importance of giving scholarly attention to and discussing real world examples of hate speech openly, however offensive they might be, to improve understanding of how it functions and to seek feasible solutions.

This book is not a volume on hate speech laws nor an introduction to linguistics.⁴ Instead, it is a volume on hate speech from various linguistic perspectives. In this sense, it is applied linguistics. As a linguistic object of study, hate speech is complex and, to a certain extent, elusive because it is not a unitary phenomenon (Brown 2017a; 2017b). Hate speech does not have a unified purpose. Hate speech can take permanent forms – e.g. racial epithets, insults, dehumanising metaphors, group defamation and negative stereotypes – but can also take transient forms. Hate speech can exist in various forms: written words, spoken words and audio-visual materials – e.g. gestures, symbols, images, films and video-games. Hate speech is not ascribed to any specific genre or rhetorical style, as it can range from thoughtful comments in a parliamentary speech to improvised sarcastic comments in an online post. Hate speech can involve many negative illocutionary and perlocutionary acts, such as insulting, degrading, humiliating, harassing, threatening, provoking, inciting hatred, hostility or violence and denying, justifying or glorifying acts of genocide. Hate speech is sometimes overt and unconcealed, but ever-increasingly coded and veiled (Becker 2020: 38). Hate speech can be delivered by identified speakers or can be anonymous, especially in online hate speech.

The book is divided into two different, but at the same time complementary, parts. Each part is devoted to one of the two applied linguistic disciplines in language and law or in law: *Legal linguistics* (Part I) and *Forensic linguistics* (Part II).⁵ Legal linguistics analyses the doctrinal content of the law and its linguistically-based structure, while Forensic linguistics is concerned with helping to establish the facts on which a legal decision is based:

⁴ For an introduction to Linguistic pragmatics, see Alba-Juez and MacKenzie (2016).

⁵ For the differentiation between Legal linguistics and Forensic linguistics, see Guillén-Nieto and Stein (2022: 2–7).

Forensic linguistics is the use of evidence from language use based on records, texts, or traces – not as the live substance, but as vestiges of the use of language, communication or speech acts that took place in the past, however medially constituted, spoken, written, digital, in connection with the resolution of crime (Guillén-Nieto & Stein 2022: 5).

Part I, *Legal linguistics*, consists of three chapters:

In Chapter 1. *Approaches to the meaning of hate speech*, the present author aims to demonstrate how Wittgenstein's concept of *family resemblance* can help our work on the description of the various definitions of hate speech. The discussion is divided into three parts, each corresponding to a different approach to hate speech. The first approach is ordinary language analysis, a philosophical investigation method concerned with how verbal expressions, in our case the term hate speech, is used in non-technical, everyday language (Wittgenstein 2009 [1953]). The discussion then focuses on various legal scholarly attempts to define hate speech: *content-based*, *intent-based* and *harms-based*. Finally, the chapter offers the reader a panoramic view of the existing regulations of hate speech in international law, common law and civil law – European Union law and Member States law.

In Chapter 2. *Hate speech as a legal problem*, the author deals with four significant issues at the core of hate speech as a legal problem: (a) the uneasy balance between the right to freedom of opinion and expression and the prohibition of incitement to hatred, hostility or violence, (b) the lack of an agreed-upon technical legal definition, (c) the difficulty of determining which speech acts are surface linguistic expressions of hate speech and (d) the legal challenges raised by the advent of online hate speech.

Chapter 3. *The legal reasoning in hate speech court proceedings* is devoted to the key foci of legal reasoning about hate speech across different jurisdictions, specifically the United States and the European Union. The chapter begins with a review of several landmark decisions of the United States Supreme Court and the United States Court of Appeals for the Armed Forces. In parallel, some landmark decisions of the European Court of Human Rights are also reviewed. Finally, the chapter analyses what constitutes evidence of hate speech in the United States and the European Union, which have been taken as representative examples of legal practices in common law and civil law jurisdictions.

Part II, *Forensic linguistics*, targets the language cues that various linguistic theories can provide for making hate speech legally actionable. The discussion aims to demonstrate the value of a micro-language approach to hate speech. In order not to build a castle in the sky, the analysis is grounded in several relevant legal cases selected from those presented in Part I (Chapter 3).

Part II consists of five chapters:

Chapter 4. *Critical discourse analysis* reviews some central theories in CDA that are deemed useful for an improved understanding of hate speech: (a) the *theory of social representations* (van Dijk 1997; 2005; 2006b), (b) the *theory of ideology* (van Dijk 1995a; 1995b) and (c) the *theory of power as control* (van Dijk 1996; 2015). The chapter illustrates the benefits of a multilevel analysis (macro level, meso level and micro level) through its application to a case associated with racial hate speech: *Brandenburg v. Ohio* (1969). I argue that a CDA approach may help unveil the social and discourse practices reproducing racism in *Brandenburg v. Ohio* (1969) in the eyes of the law.

In Chapter 5. *Register and genre perspectives on hate speech*, the author elaborates on the surface linguistic forms articulating hateful texts. Specifically, the chapter focuses on the texts – discourse segments of various dimensions – in which hate speech manifests itself and on the genre or genres into which such texts can be classified. My purpose is to demonstrate to what extent the register and genre perspectives can improve our understanding of hate speech.

Chapter 6. *Speech act theory* investigates the types of speech acts that give expression to hate speech. Although the application of Speech act theory to the analysis of hate speech is problematic, due to the inherent complexity of hate speech as an empirical object of study, the chapter points to the useful insights the theory can still provide at both macro and micro levels of linguistic analysis. The author draws the reader's attention to implicitness, indirectness and a loose illocutionary-perlocutionary link as some of the major problems in recognising and identifying the speech acts giving expression to hate speech.

Chapter 7. *(Im)politeness theory* points to the insights this socio-pragmatic theory can provide in the analysis of hate speech. The author argues that to understand the hate-advocating speaker's impolite linguistic behaviour (Culpeper 2011), one needs, first, to analyse their intentional deviation from polite behaviour. The chapter reviews the main approaches to politeness: (a) the *conversational-maxim approach* (Lakoff 1973; Leech 1983) and (b) the *face-saving approach* (Brown & Levinson 1987 [1978]). The discussion then moves forward to *impoliteness* (Culpeper 2011; Kaul de Marlangeon 2005). The chapter also illustrates how the various theoretical views on (im)politeness can be applied to actual legal cases associated with hate speech.

In Chapter 8. *Cognitive pragmatics*, the author deals with the meaning and interpretation of hate speech. Specifically, the discussion concentrates on how the hate-advocating speaker communicates an intention that is not explicitly stated and how this intention is likely to be interpreted differently by subjects not belonging to the same interpretive community. Grice's (1975) *conversational implicature* provides the bridge from what is said to what is meant but not overtly

said. The interpretation of the speaker's intended meaning is considered through Sperber and Wilson's (1995 [1986]) *Relevance theory*, with special reference to *ostensive-inferential communication*, with the aim of demonstrating the extent to which cognitive pragmatics can help unveil the hate-advocating speaker's intended meaning.

As mentioned at the outset of this Preface, hate speech is a complex object of linguistic analysis. Since hate speech is not a unitary phenomenon but, instead, multi-layered, and indeed multi-modal, it cannot be wholly explained from a single linguistic perspective. On the contrary, hate speech demands successive analyses, each focusing on a specific linguistic element. For this reason, the reader will see that in the chapters of Part II, the author sometimes recurs to certain landmark legal cases associated with hate speech, especially *Terminiello v. Chicago* (1949) and *Brandenburg v. Ohio* (1969), to illustrate the insights that each linguistic theory may provide into the same case. In this respect, hate speech resembles the elephant in the well-known poem *The blind men and the elephant*:⁶ A group of six blind men went to see an elephant. They thought that by observation, they could learn the elephant's appearance. Each blind man felt a different part of the elephant's body: the side, the tusk, the trunk, the knee, the ear and the swinging tail. As a result, there were six different descriptions of the elephant. Depending on the part felt by each blind man, elephant was like a wall, a spear, a snake, a tree, a fan or a rope. The blind men's descriptions of the elephant were different from each other because they were based on their own single experiences. All of them were partly right but they were all wrong. From this fable one can learn that if one takes a single perspective to look at hate speech – our elephant – one can only describe one of its linguistic elements, ignoring other elements that may be equally relevant to establish the facts in the judicial narrative. In Part II, the reader will be able to discover, through successive linguistic analyses, several surface linguistic manifestations of hate speech, especially the register, the genre, the speech acts, the strategies and the ostensive stimuli employed by hate-advocating speakers to signal their malicious communicative intents.

⁶ The American poet John Godfrey Saxe (1816–1887) is believed to have introduced a Hindu fable to western readers with the poem *The blind men and the elephant* (cf. Saxe, John Godfrey. 1876. *The Poems of John Godfrey Saxe*, 259–261. Boston). The “blind men” in the poem do not represent real blind people. Nor does “the elephant” in the poem represent a real elephant. On the contrary, “the blind men” and “the elephant” are the fictional characters of a fable whose moral is that humans tend to claim absolute truth based on their limited, subjective experience as they ignore other people's limited, subjective experiences, which may be equally true.

This book also evidences the symbiosis between Linguistics and Forensic linguistics in which both sides benefit from the relationship. Linguistics provides Forensic linguistics with theories, methodologies and tools for analysing hate speech. In contrast, Forensic linguistics prompts adjustments and advances in linguistic pragmatic theories and methodologies because of the new complex input data provided by language crimes, in this case, hate speech.