What is Hate Speech? The Case for a Corpus Approach

Maxime Lepoutre[1] · Sara Vilar-Lluch[2] · Emma Borg[3] · Nat Hansen[3]

Abstract

Contemporary public discourse is saturated with speech that vilifies and incites hatred or violence against vulnerable groups. The term "hate speech" has emerged in legal circles and in ordinary language to refer to these communicative acts. But legal theorists and philosophers disagree over how to define this term. This paper makes the case for, and subsequently develops, the first corpus-based analysis of the ordinary meaning of "hate speech." We begin by demonstrating that key interpretive and moral disputes surrounding hate speech laws—in particular, surrounding their compatibility with the rule of law, democracy, and free speech—depend crucially on the ordinary meaning of "hate speech." Next, we argue, drawing on recent developments in legal philosophy, that corpus linguistics constitutes a distinctively promising tool for ascertaining the ordinary meaning of "hate speech." Finally, we offer a proof of concept, by outlining, and analyzing the interpretive and moral implications of, the first such study.

Keywords Hate speech · Free speech · Democracy · Legal interpretation · Legal corpus linguistics · Ordinary language philosophy

✉ Maxime Lepoutre
   m.c.lepoutre@reading.ac.uk

   Sara Vilar-Lluch
   s.vilar-lluch@reading.ac.uk

   Emma Borg
   e.g.n.borg@reading.ac.uk

   Nat Hansen
   n.d.hansen@reading.ac.uk

1   Politics and International Relations, University of Reading, Reading, Berkshire, UK

2   Reading Centre for Cognition Research, University of Reading, Reading, Berkshire, UK

3   Department of Philosophy, University of Reading, Reading, Berkshire, UK

⌂ Springer

1 Introduction

Contemporary public discourse is saturated with speech that vilifies and incites hatred or violence against vulnerable social groups (Futtner and Brusco 2021). The proliferation of this kind of speech has prompted various responses from liberal democracies. Some of these responses take the form of counterspeech, which aims to counter this bad speech with "more speech" (Howard 2021; Lepoutre 2021). But the vast majority of liberal democracies have gone further, and adopted legal responses as well.[1] For example, the UK's Public Order Act of 1986 prohibits the use of "threatening," "abusive," or "insulting" words when these words are either intended to "stir up racial hatred," or are likely to do so.

To pick out these problematic communicative acts—and, by extension, to characterize attempted responses to them—the term "hate speech" has gained widespread usage. The term originally emerged in the 1980s among legal theorists who investigated the deployment of legal measures to counter harmful racist utterances. But, as the legal philosopher Alexander Brown (2017a: 424) has noted, the term "hate speech" is no longer confined to legal circles. Instead, it has gained currency in everyday talk, and features prominently, for instance, in news articles ("French presidential candidate Zemmour convicted of hate speech") (Abboud 2022), on social media (e.g., Twitter's #hatespeech), and in popular culture.[2] Thus, Brown suggests, the term "hate speech" does not simply have a legal meaning: it also has an *ordinary meaning* which, though it initially grew out of legal scholarship, has since taken on "a life of its own" (2017a: 424).

What *is* the ordinary meaning of "hate speech"? Brown (2017b: 574–81) mentions, and critically scrutinizes, a number of common claims (or "folk platitudes") about this term. The term "hate speech" is often thought (a) to imply a negative judgment about the speech in question; (b) to pick out speech that is, or is liable to be, legally regulated; (c) that targets particular social groups; and (d) that is liable to harm these groups. Finally—though Brown ultimately rejects this claim—"hate speech" is commonly said (e) to be connected, in some necessary way, to feelings or attitudes of hatred. To date, however, the status of these folk platitudes, and thus the ordinary meaning of "hate speech" itself, remains underexplored. For one thing, Brown's analysis (2017a; 2017b) remains the sole investigation of the term "hate speech" that is specifically dedicated to examining its ordinary meaning. Moreover, Brown himself does not purport to have given the final word on the topic. On the contrary, he explicitly "advocat[es] a new research agenda, to be pursued from a variety of academic disciplines across the arts, humanities, and social sciences" aimed at "excavating the ordinary concept *hate speech*" (2017a: 430).

The present paper aims to contribute to this interdisciplinary research agenda. Specifically, it aims to do so by making the case for, and developing, the first

---

[1] For an influential overview of this body of law, see Brown (2015).

[2] See, for example, the common occurrence of the term "hate speech" in pop songs: https://www.lyrics.com/lyrics/hate%20speech.

sustained corpus analysis of "hate speech."[3] In recent years, corpus linguistics has emerged as an increasingly popular tool for ascertaining the ordinary meaning of legal terms (e.g., Lee and Mouritsen 2017, 2021; Gries 2020). We aim to bring this fruitful interpretive methodology to "hate speech"—and, in so doing, to further our understanding of this term's ordinary meaning.

The rest of the paper will proceed as follows. Section 2 demonstrates why it matters, from a legal, political and ethical standpoint, that we grasp the ordinary meaning of "hate speech." Sect. 3 then introduces corpus linguistics and argues, drawing on recent debates in legal philosophy, that it is a promising tool for ascertaining the ordinary meaning of "hate speech." Having thus made the case for a corpus approach to "hate speech," Sects. 4 and 5 offer a proof of concept: they outline (Sect. 4), and analyse the results of (Sect. 5), the first such study. Section 6 briefly concludes.

2 Why the Ordinary Meaning of "Hate Speech" Matters

As noted above, Brown argues for an interdisciplinary research programme aimed at determining the ordinary meaning of "hate speech." But why should we care what "hate speech" means ordinarily, when it is used outside of legal circles?

2.1 Interpretation

The first reason has to do with legal *interpretation*. Some influential theories of legal interpretation—e.g., textualism or "public meaning" originalism—stipulate that the meaning of legal terms is, to a significant degree, a function of their ordinary meaning. For those who embrace these interpretive theories, ordinary meaning is therefore directly relevant to interpreting legal terms (e.g., Gries 2020: 628–29).

But the interpretive significance of ordinary meaning doesn't require embracing these specific legal theories. Even those who deny that legal terms should be interpreted according to their ordinary meaning may nonetheless agree that ordinary meaning is epistemically relevant to interpretation. For example, purposivists and intentionalists (who give greater emphasis to lawmakers' intentions) can take the ordinary meaning of a term as defeasible evidence of what lawmakers intended—and so, of legal meaning (Tobia 2021: 741).

In our context, the upshot is that the ordinary meaning of "hate speech" is relevant, directly or indirectly, to interpreting "hate speech" in legal texts that contain this term, as in South Africa's Promotion of Equality and Prevention of Unfair Discrimination Act 2000 (especially s.10), and parts of the European Court of Human Rights' case law.[4]

---

[3] Note that Culpeper (2021) has launched a preliminary investigation into this topic. But the term he actually investigates is "hateful," not "hate speech." This is a significant limitation because the relationship between "hate speech" and "hate" or "hatefulness" is hotly debated. See Sect. 5e below.

[4] See, especially, *Gündüz v. Turkey* (in Brown 2017a: 435).

There is a complication here. Although some bodies of law do contain the term "hate speech," many of the laws that are conventionally referred to as "hate speech laws" do not (Brown 2015). For instance, the UK's Public Order Act of 1986 makes reference to language that is "insulting," "threatening," "abusive," and "stir[s] up hatred," but does not include the term "hate speech." The ordinary meaning of "hate speech" is therefore not directly relevant to legal interpretation in *these* cases. In acknowledgement of this fact, however, we will complement our corpus analysis of the ordinary meaning of "hate speech" with an analysis of several other terms commonly contained in hate speech laws (see Sect. 4 below).

2.2 Implementation

There is a second core reason why the ordinary meaning of "hate speech" (and other terms contained in hate speech law) matters. Even if the ordinary meaning of these terms were not relevant to the interpretation of hate speech laws, it would still be relevant to their *implementation*.

This is because, when implementing hate speech laws, it is important to know whether there is a mismatch between the legal meaning of "hate speech" (and other terms contained in hate speech law), and the way these terms are ordinarily understood. As Jeffrey Howard (2019: 211–213) has argued, the moral justification of laws aimed at prohibiting speech depends partly on the expected consequences of implementing those laws. The problem, in this context, is that a mismatch between the legal meaning of "hate speech" (and other terms contained in hate speech law), and the ordinary meaning of these terms, risks leading to adverse consequences when implementing hate speech laws.

One such potential consequence concerns *fair notice*. A key principle of the rule of law is that those subjected to the law must have "fair notice." They must be able to organize their lives and adjust their behaviours around existing laws, so as to avoid violating them—and for that, they need to be able to understand what the law says (e.g., Gries 2020: 629; Tobia 2020: 737). Accordingly, if the legal meaning of terms associated with hate speech laws is systematically disconnected from the way these terms are ordinarily understood, the public may lack the "fair notice" required adequately to adjust their behaviours.

A further possible consequence relates to *trust in democratic norms*. If the actual meaning of hate speech laws is disconnected from the way the public ordinarily understand "hate speech," this may lead to the public perception—justified or unjustified—that those who are being sanctioned by these laws have unfairly been silenced. This feeling of silencing, in turn, may fuel a loss of trust in democratic norms of inclusion and public debate. This risk is especially high amidst the present "crisis of representation," where some sections of the electorate—rightly or wrongly—already feel ignored or silenced.

The final risk we will mention here concerns the so-called *chilling effect*. A common critique of hate speech laws is that their implementation could lead to widespread self-censorship among the electorate—including to self-censorship of legitimate speech (Howard 2019: 245–46). Whether this worry is well founded depends

crucially on how hate speech laws are ordinarily understood. If, for example, "hate speech" (and other terms contained within hate speech laws) are ordinarily understood in a way that is more expansive than their actual legal meaning, deploying such laws may result in a chilling effect.

Thus, identifying the ordinary meaning of "hate speech" matters for reasons of implementation as well as interpretation. Notice, moreover, that the considerations relating to implementation apply whether or not the term "hate speech" is contained in hate speech laws. If it is, then its ordinary meaning is clearly relevant to how those laws are publicly understood. But even laws that do not contain the term "hate speech" (such as the Public Order Act of 1986) are conventionally labelled, and regarded as, "hate speech laws."[5] Consequently, the ordinary meaning of "hate speech" provides insight into how these laws are interpreted by the public.

Having said that, it is important to remember that our analysis in Sect. 4 will not merely examine the term "hate speech" itself. As explained in 2.1, we will complement our analysis of "hate speech" by examining several other terms that are commonly contained in hate speech laws. This will help provide further insight into how hate speech laws are ordinarily understood, independently of whether they contain the term "hate speech." And, by extension, it will help further clarify whether, and to what extent, implementing hate speech laws is likely to violate the principle of fair notice, erode trust in democratic norms, and produce a free speech-impairing chilling effect.

3 Why Use Corpus Linguistics?

We have argued that there are strong reasons—relating to legal interpretation and to the ethical and political costs of implementation—to ascertain the ordinary meaning of "hate speech." But how should we go about doing so? There are various ways of assessing the ordinary meaning of a legal term. For example, one can proceed by examining one's own intuitions about meaning; by looking up a dictionary definition of the term(s) in question; or, by surveying ordinary citizens or experts in experimental settings (e.g., Gries 2020: 629; Mouritsen 2011: 180–90). Brown, for his part, predominantly proceeds by appealing to intuitions about permissible uses of "hate speech."[6]

In recent years, however—in part due to dissatisfaction with these existing methods—corpus linguistics has emerged as a distinct—and distinctly promising—way of ascertaining the ordinary meaning of legal terms. Corpus linguistics is a branch of linguistics that studies language by examining large bodies of text (corpora).

---

[5] For instance, the Public Order Act of 1986 is taxonomized as a "hate speech law" in crowdsourced sites such as Wikipedia, in academia (Brown 2015), in think tanks (Alkiviadou et al. 2020), etc. The corpus data we report in Sect. 4 appears to support this point. Uses of "hate speech" in UK newspapers are prototypically associated with the idea of legal regulation. This suggests that, despite not containing the term "hate speech," laws banning hate speech in the UK are ordinarily understood in these terms.

[6] In particular, he often appeals to his intuitions regarding whether certain uses of the term "hate speech" appear meaningful or appropriate. See, e.g., Brown (2017a: 445, 448, 451, 452, 453, 457, 461).

These texts are typically drawn from real-world (or "natural") communicative settings. And the examination of these texts is usually facilitated by software-assisted statistical analyses (Gries 2020: 631–32).

More specifically, corpus analysis of a term's meaning can take at least three forms. *Collocation* analysis examines which words or phrases most frequently appear near a term under consideration. Closely related to collocation analysis is *keywords* analysis. Keywords are terms that are particularly salient within a corpus (partly because they appear more often in it than we would normally expect them to) and which are commonly used to shed light on what a given corpus is about. When a corpus is assembled around a target term (e.g., "hate speech"), keywords analysis can therefore provide insight into what texts containing this term are usually about. *Concordance line* analysis, finally, involves extracting target terms with excerpts from the contexts in which they occur ("concordance lines"). For example, one might examine a random sample of hundreds of concordance lines (the size of a sample will vary depending on the number of times the target term appears in the corpus), to evaluate how the term is used (Gries 2020).Concordance line analysis combines quantitative and qualitative approaches to understanding a text, since it aggregates individual judgments about how an expression is being used in various contexts.

The jurisprudential movement known as "legal corpus linguistics," which was pioneered by Judge Thomas Lee and Stephen Mouritsen (2017), proposes to apply corpus linguistics to the ordinary meaning of legal terms.[7] There are at least two benefits to this approach.

The first relates to the size of its sample. Legal corpus linguistics examines the way a legal term is used across very many cases: often, it considers hundreds or thousands of uses. Why does this large number of cases matter? The idea is that, insofar as the corpus is balanced—i.e., insofar as it contains a range of different speakers and communicative contexts—this will yield a representative picture of the way the term is ordinarily used (Gries 2020: 637).

This arguably constitutes an improvement over two alternative ways of assessing ordinary meaning: the use of individual intuitions; and the appeal to dictionary definitions.[8] Intuitions about the ordinary meaning of a term vary across different individuals (Gries 2020: 629–30). And a similar observation applies to dictionaries: definitions of a given term vary across different dictionaries—so much so, that different sides in legal disputes often appeal to competing dictionary definitions (Solan and Gales 2016: 257). Accordingly, having a single person introspect their intuitions, or appealing to a particular dictionary definition, is unlikely to yield a representative picture of the way a term is ordinarily used or understood.

The second core benefit of using corpus linguistics to gauge the ordinary meaning of legal terms is its focus on language in "natural" communicative settings. Corpora are typically not composed of speech produced simply for the

---

[7] See also Mouritsen (2011), Gries (2020), and Lee and Mouritsen (2021).

[8] It is more debatable whether this constitutes an advantage relative to experimental survey methods. Tobia (2020) argues that it does not; but Sytsma et al. (2019: 231–32) disagree.

purposes of a study. Rather, they contain speech that was generated naturally, as part of real-world communication (e.g., newspapers, social media, transcribed speeches) (Lee and Mouritsen 2021: 320–32; Sytsma et al. 2019: 232–33). This stands in contrast to "experimental jurisprudence" which instead involves "artificial" speech—speech that was deliberately elicited, for the purposes of study, in experimental settings.

Why does the preference for natural speech matter? One reason is that speech that is artificially elicited may not accurately reflect ordinary usage. According to Lee and Mouritsen (2021: 320), subjects may deviate from ordinary usage quite simply because they "know that they are being observed and subjected to analysis." Put differently, the mere awareness of being observed, and the linguistic introspection prompted by this awareness, may lead subjects to use language in ways they would not in ordinary communicative settings. Another concern with artificially elicited speech concerns *the specific way* that it is elicited. Speech that is elicited in experimental surveys may be distorted by pragmatic effects of the survey design. As Sytsma et al. observe, the wording or structuring of experimental survey questions often inadvertently invites particular interpretations of the question, in a way that biases the answers that follow (Sytsma et al. 2019: 231–32).

To be clear, these concerns may not be insurmountable. It may be possible to devise a survey experiment in a way that mitigates the "observer effect" and simulates a natural communicative setting. Moreover, well-designed surveys may be able to avoid unwanted pragmatic effects (Sytsma et al. 2019: 232–33). But the point remains that doing so is difficult. Because corpus linguistics uses language that was produced independently of the study, in real-world communicative settings, it largely sidesteps these difficulties.

Of course, legal corpus linguistics faces challenges of its own. To begin, conducting an adequate corpus analysis is no easy task. It requires, notably, assembling a time-relevant and balanced corpus; carefully designing search terms for that corpus; and systematically interpreting keywords, collocates, and concordance lines. Early defences of legal corpus linguistics have sometimes downplayed the expertise needed to carry out these tasks (e.g., Mouritsen 2011: 203). This is not an argument against legal corpus linguistics per se. But it does suggest, contrary to what some corpus advocates have argued, that it may often be impractical for judges to deploy this tool, at least without expert assistance.

But some critics have raised a more fundamental challenge. According to Tobia (2021), even when properly conducted, a corpus search only provides partial insight into the ordinary meaning of a legal term. Specifically, Tobia (2021) argues that legal corpus linguistics tends to reveal the *prototypical* meaning of a term—i.e., its most salient meaning. This is, in part, because collocation analysis highlights which words are most frequently used alongside the term under investigation. The problem, for Tobia (2021: 759), is that terms can also have *permissible* uses that are less common and less salient. For example, although the prototype of "vehicle" may be a machine with wheels and an engine (e.g., a car), "vehicle" may nonetheless permissibly be used to refer to non-prototypical modes of conveyance, such as a canoe. The concern, then, is that focusing on prototypical meaning will tend to obscure such permissible meanings.

There are three things to say in response to this objection. The first is that the prototypical meaning of "hate speech"—the meaning that is most salient to people—is arguably the most important one for our purposes (Lee and Mouritsen 2017: 788). Consider, for example, fair notice. If hate speech laws are interpreted according to an extremely uncommon—but permissible—meaning of hate speech, this may not give the public fair notice of what to expect from the law. Accordingly, if lawmakers wish to give the public fair notice, it seems crucial to consider whether the content of hate speech laws is aligned with the most salient public understanding of "hate speech."

The second point is that, on closer inspection, corpus linguistics may also provide insight into permissible but non-prototypical meanings of "hate speech." It is true that collocation highlights the terms or phrases that are most commonly associated with "hate speech." But corpus linguistics does not reduce to collocation analysis.[9] Examining a large enough sample of concordance lines containing "hate speech" would reveal obscure uses of "hate speech" as well as more common ones.

The final—and most important—point is that corpus linguistics needn't be the only tool we use to ascertain the ordinary meaning of "hate speech." As Lee and Mouritsen (2021: 358) acknowledge, the best way to assess ordinary meaning may be "a sort of triangulation," which combines corpus evidence with other approaches, such as survey experiments and dictionary definitions (see also Sytsma forthcoming). Thus, even if it were true that corpus linguistics provides limited insight into permissible meanings of "hate speech," we can use survey methods, dictionaries, or even individual intuitions to complement it.

More generally, our point is not that corpus linguistics is perfect, or even that it should replace alternative modes of analysis (such as Brown's). The point is simply that, given its distinct strengths—in particular, its large sample size, and its emphasis on natural language—we have good reason to enrich existing examinations of the ordinary meaning of "hate speech" (and other terms contained in hate speech law) using a corpus approach. This is what we do in the rest of this article.

4 A Corpus Study of "Hate Speech"

The corpus study we carried out sought answers to two questions. First and foremost: how is the term "hate speech" ordinarily understood? Answering this first question can shed light on how the public perceive hate speech laws—and by implication, on how we should interpret these laws, and the ethical costs associated with their implementation. However, as discussed in Sect. 2, not all hate speech laws contain the term "hate speech." So, to provide further insight into public perceptions and interpretations of hate speech laws, we complemented our first question with another: namely, how are key terms contained in hate speech laws ordinarily understood? To make this second question tractable, we focused on a number of key

---

[9] Moreover, not all frequent collocates are signs of prototypical meaning. A word may be frequently associated with another precisely to modify its prototypical meaning.

terms used in UK law (specifically, the Public Order Act of 1986) to characterize hate speech.

To investigate these questions, we assembled relevant corpora. From a corpus linguistics perspective, the fact that "hate speech" is a legal term as well as an ordinary term poses a potential challenge. Searches of "hate speech" in a general corpus are likely to retrieve some legal uses (e.g., in institutional regulations or other legal documents) alongside non-legal ones. This can make it more difficult to examine how people outside the legal community use "hate speech." To mitigate this difficulty, we built a small specialized "journalistic hate speech" corpus for the purpose of the study to conduct a pilot analysis that would inform a second analysis on a "general" corpus.

The "journalistic hate speech" corpus we assembled comprises 255 news reports about hate-speech related events (e.g., incidents and offences, political debates, hate speech-related laws) consisting of 164,183 words. News articles were retrieved from the Nexis database Lexis Library News, and date from 1990 to 2021. All news reports were published by British media, though they in principle cover both national and international news. In order for this journalistic corpus to be as representative as possible, it included a diverse range of media outlets, including both tabloid and non-tabloid newspapers, and both large national newspapers (e.g., *The Guardian, The Times,* the *Daily Mail*) and smaller local newspapers (e.g., the *Yorkshire Post,* the *Belfast Telegraph*, and the *Birmingham Post*).

The pilot analysis of this journalistic hate speech corpus was intended to provide an initial look into the public understanding of "hate speech." This first analysis was then followed by a second study of "hate speech" in the "general" corpus, *English Web 2020 (enTenTen20).*[10] The *English Web 2020 (enTenTen20)* contains 38 billion words and is comprised of texts retrieved from internet domains of states whose official language is English. In addition to being much larger than the journalistic corpus, the general corpus is also much more varied. It contains texts drawn from a wide range of sources, including not just newspapers, but also blogs, discussion sites, and (to a much smaller extent)[11] legal sources. Moreover, its texts examine an extremely broad range of topics, including arts, society, business, science, sports, technology, and so on.

The first stage of our analysis (the pilot analysis), which focused on the journalistic hate speech corpus, took the following form. We began by studying the keywords for this corpus. As Sect. 3 discussed, keywords are words that are particularly salient within a corpus, and which therefore provide insight into what this corpus is about (Baker 2006: 125). We then examined the top collocates, within this corpus, for the term *"*hate speech."[12] Collocates, recall, are words that

---

[10] This "general corpus" is available in *Sketch Engine* (app.sketchengine.eu)*,* the software employed to perform the corpus analysis.

[11] As explained in *Sketch Engine* (see previous note), the legal genre represents by far the smallest of the listed *enTenTen20* genres (compared to "blogs," "news," and "discussion").

[12] We also conducted follow-up searches on some related terms, such as "hatred," for reasons that will become clear in 5.5.

frequently co-occur in a particular corpus, thereby providing insight into how these terms tend to be used (Xiao 2015). Collocations were retrieved using *Sketch Engine's* Word Sketch tool, which provides fine-grained information about the grammatical patterns in which collocates appear. This grammatical information yields greater understanding of how a given term relates to its top collocates (and so, greater understanding of how this term is used).[13] Finally, the collocation analysis was supported by an examination of concordance lines (lines of text that show uses of the term considered in context).

Broadly speaking, this pilot analysis tended to offer support for—as well as important precisifications of—the folk platitudes about "hate speech" mentioned in the Introduction (regarding the valence of "hate speech," and its relationship to legal regulation, social groups, harm, and feelings of hate).

One core advantage of the journalistic hate speech corpus is that it wholly excludes legal texts, and so helps us focus on non-legal uses. However, this corpus remains relatively small (by the standards of corpus linguistics). Furthermore, its restriction to journalistic texts raises the possibility that the patterns of use it reveals might not be representative of ordinary users more generally. In particular, one might plausibly think that journalists writing about hate speech are better acquainted with hate speech law than the average person. As a result, one might worry that the understanding of "hate speech" revealed by the journalistic corpus is likely to be biased in the direction of the legal understanding.[14]

To be clear, this bias does not mean that the journalistic corpus is irrelevant to understanding how "hate speech" is ordinarily understood. For one thing, news articles about hate speech can plausibly be expected to influence how ordinary people outside the journalistic profession think about hate speech. And since—as explained above—our journalistic corpus includes a diverse range of newspapers (local and national, tabloid and non-tabloid), some of which have a large circulation, it plausibly provides some insight into how a wide range of ordinary people might conceive of hate speech. For another, although it is likely that journalists are better versed in hate speech law than ordinary users, it needn't follow that the way they use "hate speech" perfectly coincides with the legal meaning of "hate speech." News articles and legal texts serve different social functions and have different intended audiences. Hence, we might expect them to diverge in the way they use some terms, including some legal terms.[15] In Sect. 5.3, for instance, we will see that, even in the journalistic corpus, accounts of the groups that can be targeted by "hate speech" sometimes diverge from the legal meaning of "hate speech."

Having said that, it is still plausible to think that journalists tend to use "hate speech" in a way that is closer to the legal meaning of "hate speech" than most

---

[13] Collocate strength is calculated with the logDice score, based on the frequencies of the node, the collocate, and the co-occurrence of the node and the collocate.

[14] We are grateful to a reviewer for pressing us on this point.

[15] According to Tobia et al. (2023, 27), people are less likely to think that legal professionals are experts about the meaning of a given term (including a given legal term) when that term features in a non-legal text (such as a newspaper) as opposed to a legal text. This constitutes prima facie evidence that legal terms may be used and understood differently in newspapers, compared to legal texts.

ordinary people do. To offset this potential bias, the second stage of our analysis therefore sought to replicate our findings from the journalistic corpus by conducting collocation and concordance analyses of "hate speech" in the general corpus. As discussed above, the general corpus is larger and includes much more diverse sources than our journalistic corpus. Consequently, examining the general corpus allowed us to check whether portrayals of hate speech in the journalistic corpus were reflected in broader patterns of use. On the whole, we found that they tended to be. In Sect. 5, we will cite evidence from both the journalistic and general corpus to ensure that our analysis of the ordinary meaning of "hate speech" is as balanced as possible.

Combining the journalistic corpus with the general corpus goes some way towards alleviating concerns about bias in our sample. But it is also important to reiterate that we do not intend this study to be the final word on what corpus analysis can say about the ordinary meaning of "hate speech." On the contrary, we intend this study primarily as a proof of concept, which demonstrates the fruitfulness of this mode of inquiry. Consequently, our hope is that this initial study will help motivate follow-up studies, which assemble and investigate other relevant corpora (for example, uses of "hate speech" on various social media platforms), and which therefore help us arrive at an even more balanced understanding of the ordinary meaning of "hate speech."

The stages of our analysis that we have just outlined focused directly on the term "hate speech." But, as explained above, the final stage of our analysis instead focused on a number of key hate-speech related terms in the UK's Public Order Act. Specifically, we focused on the terms "threatening," "abusive," and "insulting." We conducted concordance and collocation analyses relating to these terms, as well as disjunctions and conjunctions of these terms (e.g., "threatening and/or abusive"), within the general corpus. These analyses paid special attention to two things. First, the concordance line analysis explored whether, and how often, uses of these terms and locutions were hate-speech related or not (and so, whether there is public awareness of the place of insulting, threatening, and/or abusive speech within hate speech). Second, the collocation analysis allowed us to compare the semantic values of these three terms, using *Sketch Engine's* Word Sketch Difference tool, which automatically compares the collocates associated with two different terms.

Below, we examine the key conceptual and normative take-aways of this corpus study by systematically revisiting the core "folk platitudes" about hate speech introduced in Sect. 1. Details can be found in the Appendix.

5 The Ordinary Meaning of "Hate Speech"

As noted in Sect. 1, debates surrounding the meaning of "hate speech" often revolve around the following propositions: (a) that the term "hate speech" implicates a negative evaluation; (b) that it involves the idea of legal regulation; (c) that it targets groups; (d) that it tends to produce harm; and (e) that it is connected, in some important way, to feelings of hate or hatred. Though Brown is critical of at least some of these propositions (see below), he refers to these as "folk platitudes" (or common

claims) about the ordinary meaning of "hate speech" (Brown 2017b: 574–81).[16] The results of our corpus analysis speak to each of these claims.

5.1 "Hate Speech" and Evaluation

The first and least controversial thing to note is that, in ordinary discourse, the term "hate speech" characteristically involves a negative evaluation. To call something "hate speech" is to express a negative judgment of it.

This is notably manifested in the collocation data regarding the actions most commonly associated with "hate speech." For example, in the general corpus, "hate speech" is characteristically something that is "spewed," which connotes that it is undesirable.[17] This also comes across in the metaphors and similes deployed to refer to "hate speech." For one thing, collocations backed by concordance line analysis show that, in both the general and the journalistic hate speech corpus, hate speech is regularly likened to a destructive natural force (e.g., "hate speech is spreading like wildfire in social media.").[18] For another, discussions of hate speech in the general corpus routinely evoke the metaphor of war: hate speech is something we must "battle," "combat," or "fight."

Thus, corpus data about ordinary usage strongly supports the folk platitude that "hate speech" has a built-in evaluative dimension: it is inherently something undesirable, to be feared like a natural catastrophe, and possibly eliminated like an enemy on the battlefield.

5.2 "Hate Speech" and the Law

But even if "hate speech" picks out speech that is negatively evaluated, this leaves open what exactly should be done about it. Here, the second important finding is that "hate speech" is ordinarily viewed as speech that is, or at least is liable to be, legally regulated.

This is borne out, firstly, by keywords data. The top keywords for "hate speech" in the journalistic hate speech corpus reveal that this term typically appears in texts that are *about* "banning", "criminalising," or "prosecuting" hate speech. An analysis of collocation data in the general corpus (with, as usual, checks of concordance lines to provide context) supports this finding. Many of the top collocates for "hate speech" concern the legal prosecution of particular instances of hate speech (e.g.,

---

[16] Brown initially mentions just four of these (evaluation, regulation, hate, groups) but he later cites a fifth (harm).

[17] Collocate data from the general corpus, for instance, shows that that the top five objects of the verb "spew" are prototypically negative entities: namely, "venom," "lava," "hate," "vitriol" and "bile." Warriner et al.'s (2013) systematic examination of word valence corroborates this, indicating that "spew" has very low valence, on a par with "ordeal," "neurosis," and "narcissism.".

[18] Warriner et al. (2013) also show that "wildfire" has negative valence, comparable to "wasteful," "violence," and "vicious.".

"prosecute," "convict," "indict," "arrest") or with broader legislative policies surrounding hate speech (e.g., "legislation," "prohibition," "law").

These patterns of use do not necessarily mean that people generally believe hate speech morally *should* be legally actionable. But they do show an awareness of "hate speech" as a legally relevant term. Hate speech, on this understanding, is speech that is usually subject to legal regulation, and performance of which is liable to lead to prosecution.

This is not an obvious finding, for two reasons. First, although the term "hate speech" originated in legal circles, it has, as Brown observes, since gained a life of its own outside legal circles (Brown 2017a). Second, and as we have already mentioned, the specific term "hate speech" is absent from the text of many laws that are conventionally regarded as hate speech laws (including the Public Order Act of 1986). Prima facie, it was therefore by no means obvious that the ordinary concept "hate speech" would preserve, or involve, this legal dimension. Yet our corpus approach suggests that it does: the idea that hate speech is liable to legal regulation and prosecution is salient in ordinary usage.

This result is morally significant, not least because of its implications for fair notice. Indeed, it tells us at least two important things about the public's awareness and understanding of the laws to which it is subject. The first is that, although the term "hate speech" is absent from many hate speech laws, its ordinary meaning *is* relevant to how the public regards and understands these laws. Patterns of ordinary use show that "hate speech" is commonly used when referring to the regulation and prosecution of hate speech—regardless of whether the laws in question contain the term "hate speech." The second implication is more substantive. It is that there *is* widespread awareness, among the public, that hate speech is usually subject to legal regulation. Prima facie, this is good news for fair notice.

To some, however, building legal regulation into the ordinary meaning of "hate speech" comes with countervailing drawbacks. Brown, in particular, expresses the concern that doing so may lead to an excessively narrow understanding of "hate speech"—particularly if the community in question has a high threshold for considering speech to be regulatable.Hence, such a concept of "hate speech" may fail to include many "forms of speech that disproportionately harm already disadvantaged or victimised members of society" (2017b: 581).

There are several things to say in response. First, the fact that "hate speech" is closely associated to the idea of legal regulation needn't entail a narrow understanding of "hate speech." One reason for this is that we could in principle decide to endorse a low threshold for speech to be regulatable. Alternatively, we might maintain a high threshold, but determine that more speech satisfies this threshold than had previously been acknowledged (perhaps due to a greater acknowledgement of the harms that hate speech inflicts on disadvantaged members of society).

Second, a narrow conception of "hate speech" may not be as problematic as Brown suggests. It *would* be problematic if, as Brown (2017b: 581) worries, it left us without the conceptual resources needed to "identify and flag" communicative acts that disproportionately harm certain vulnerable groups. But we have other concepts that can serve this purpose. For example, we might use Mary Kate McGowan's (2019: ch.5) concept of "oppressive speech" to pick out such communicative

utterances, while leaving it open which instances of oppressive speech are regulatable—and so, which of them counts as "hate speech."

Moreover, a relatively narrow ordinary concept of "hate speech" may be positively useful from the perspective of the chilling effect. As mentioned in Sect. 2, one of the core concerns with legally regulating "hate speech" is that it might lead to excessive self-censorship. The narrower the ordinary meaning of "hate speech" is, the less likely this is to happen. Thus, the fact that ordinary uses of "hate speech" prototypically involve the idea of regulation—as our corpus study suggests—needn't be a problem, even if it were to lead to a more restricted conception of "hate speech."

So far, we have suggested that the ordinary meaning of "hate speech" is prototypically associated with the idea of legal regulation, and that, normatively speaking, this may well be a welcome result. At this point, however, one might raise a methodological concern. Our analysis up to this point has assumed that, when "hate speech" is associated with the idea of legal regulation in our corpora, this is evidence that the ordinary meaning of "hate speech" is connected to the idea of legal regulation. Yet this assumption may be too quick. It is possible, after all, that ordinary people sometimes use the ordinary concept of "hate speech," and sometimes instead use the legal concept of "hate speech." Indeed, in a series of recent experimental studies, Tobia et al. (2023) find that ordinary people are often inclined to interpret legal terms according to their legal meaning (even when these legal terms also have an ordinary meaning).[19] Accordingly, it is possible that, when "hate speech" is associated with the idea of legal regulation in the journalistic corpus and the general corpus, people are using the legal concept, not the ordinary concept. If so, then our evidence would not necessarily show that the *ordinary* meaning of "hate speech" is connected to the idea of legal regulation.[20]

This is a real concern. It is extremely difficult, empirically speaking, to distinguish (1) the possibility that the ordinary concept of "hate speech" deployed by ordinary people is associated with the idea of legal regulation, from (2) the possibility that ordinary people are simply switching seamlessly between the ordinary concept of "hate speech" and the legal concept of "hate speech."

The first thing to say in response is that this methodological concern is not specific to the corpus approach. Consider, for example, the use of dictionary definitions to ascertain ordinary meaning. Suppose a dictionary definition of "hate speech" refers to the idea of legal regulation.[21] From this, we might want to infer that the ordinary concept of "hate speech" is associated with the idea of legal regulation. But, here too, there is an alternative interpretation—namely, that this dictionary definition is informed by ordinary uses of "hate speech," and those ordinary uses switch back and forth between the ordinary concept and the legal concept.[22]

---

[19] See, e.g., Tobia et al.'s (2023, 1) assertion that "ordinary people […] regularly take terms in law to communicate *technical legal* meanings, not ordinary ones.".

[20] We are grateful to a reviewer for raising this challenge.

[21] See, e.g., the Cambridge Dictionary (2022) definition, whose sole example of "hate speech" in a sentence makes reference to the legal prosecution of hate speech.

[22] An opponent might respond that, if ordinary uses of "hate speech" refer to the idea of legal regulation so much that this is picked up by a dictionary definition, then the idea of legal regulation has arguably become part of the ordinary meaning of "hate speech." But if we grant this response, then it *also* helps

But even if the problem at hand is not *specific* to the corpus approach, it might still limit our ability to draw conclusions from our study. So we need to go further than this initial "companions-in-guilt" response. A second response is empirical. Tobia et al.'s (2023) recent experimental studies suggest that whether ordinary people understand a term according to its legal meaning, or to its ordinary meaning, is highly sensitive to the term's context. When an ordinary user is asked how a term should be understood *in a legal text*, they are very likely to think it should be understood according to its legal meaning.[23] But when the term is placed in an ordinary non-legal context (e.g., newspapers) the opposite is true—that is, the ordinary meaning tends to be favoured.[24]

This is significant, because our corpora are overwhelmingly composed of non-legal texts. The journalistic corpus, recall, is exclusively composed of newspaper articles; and although the general corpus *can* include legal texts (see Sect. 4), these represent at most an extremely small part of this corpus.[25] This provides some reason to think that our corpora are well placed to provide insight into the ordinary meaning of "hate speech," as opposed to its legal meaning. And so, it provides tentative reasons for thinking that the very strong association we find between "hate speech" and the idea of legal regulation reflects, at least partly, the ordinary meaning of "hate speech."

We emphasize that these empirical reasons are tentative. After all, Tobia et al.'s (2023) experimental surveys do not specifically ask respondents how they think "hate speech" should be understood in ordinary contexts. But their experiments nevertheless help suggest future empirical research that could complement, and reinforce the results of, the present corpus study (in line with the triangulation approach outlined in Sect. 3). Future research could use experimental methods to assess whether ordinary people are inclined to think "hate speech" should be understood according to its legal meaning, or according to its ordinary meaning, across different contexts.

The third and final response relates more closely to the normative significance of our study: even if the alternative interpretation of our results proved to be correct, this still would not undermine the normative insights we draw from our corpus analysis. To see this, consider once more the principle of fair notice. Suppose it were true that, whenever ordinary users in our corpus associate "hate speech" with legal

regulation, they are drawing on the *legal concept* of hate speech. Still the important conclusion outlined above remains: namely, that ordinary users appear to be aware that hate speech is liable to legal regulation.[26]

Thus, our point is not just that the methodological challenge under consideration is not specific to the corpus approach; nor simply that there are (defeasible) empirical reasons to think that the ordinary meaning of "hate speech" *does* have a connection to the idea of legal regulation. In addition—and crucially—our corpus is capable of providing normatively significant insights even *before* we can decisively resolve this methodological challenge.

Let us take stock. We have argued that, insofar as our corpus approach provides insight into the ordinary meaning of "hate speech,"[27] it suggests that the ordinary meaning of "hate speech" is aligned with the law in an important respect: it recognizes that hate speech is, very often, legally regulated. But even so, it is possible that what ordinary usage counts as "hate speech" differs substantially from the speech that is actually prohibited. In what follows, we therefore examine more closely the perceived *content* of hate speech.

5.3 "Hate Speech" and Groups

"Hate speech," as it is legally understood, is typically speech that targets particular social groups (Brown 2016; Waldron 2012; Gelber 2019). Our corpus analysis suggests that this is reflected in ordinary use. "Hate speech," as it is ordinarily understood, takes aim at specific groups of people.

This is visible, firstly, in the keywords data from our journalistic hate speech corpus. Overwhelmingly, "hate speech" is mentioned in journalistic texts that are about speech that targets racial and ethnic groups, religious groups, and sexual minority groups. The collocates from the general corpus broadly support this. The term "hate speech" is frequently modified by adjectives specifying which social group is being targeted (e.g., "racist," "antisemitic," "homophobic," "anti-gay," "Islamophobic," "sexist").

At first sight, this is a positive result from the standpoint of fair notice. It suggests that, in this respect, the ordinary meaning of "hate speech" is aligned with what the law actually prohibits. But this agreement might yet be superficial. It also matters whether the specific target groups included in hate speech law match the target groups specified in ordinary use.

A complication arises when tackling this question. The problem is that different hate speech laws specify different categories of social groups (Brown 2016: 276–77). To make this question tractable, therefore, we will focus predominantly

---

[26] Tobia et al. (2023, 9, 34, 59–68) are clear on this point. From the observation that ordinary users understand certain legal terms according to their legal meaning, they conclude that the way ordinary users construe this legal meaning is relevant to fair notice.

[27] This qualifier is needed because, as explained above, it is conceivable that users are switching back and forth between the ordinary and legal concepts, and our empirical data cannot conclusively rule out this possibility.

on a specific category of hate speech laws: the UK's Public Order Act of 1986. And, accordingly, we look predominantly at uses in the journalistic hate speech corpus, which comprises texts drawn exclusively from British media, and is therefore most relevant to the UK context.

The most striking observation is that there is a great deal of alignment between the two in terms of which categories of groups are included. Since 2010, the Public Order Act prohibits speech that stirs up hatred against racial groups (where this includes ethnic groups), religious groups, and groups defined by sexual orientation. As mentioned above, this is mirrored in the keywords data from the journalistic hate speech corpus. Uses of "hate speech" mostly occur in contexts involving speech that targets groups on grounds of their race and ethnicity, religious, and sexual orientation. So, the ordinary meaning of "hate speech" in the UK seems largely to match existing UK hate speech law.

There are nevertheless some notable points of departure. The journalistic hate speech corpus contains frequent references to specific gender-based hate speech.[28] For instance, the word "misogynistic" is one of the most frequent modifiers of the term "hate speech." To be clear: this does not mean that gender-based hate speech is a prototypical example of "hate speech." By analogy, the fact that "black" is a frequent modifier of "swan" does not mean that swans are prototypically black. Rather, it means the opposite—the modification is needed precisely because swans are *not* prototypically black. Still, this fact about ordinary usage does suggest that hate speech can target gender groups. And, to reiterate, this constitutes a departure from UK law.

What should we make of this departure, normatively speaking? The departure suggests, first, that the ordinary meaning of "hate speech" is more expansive than the legal meaning. This, in turn, might raise worries relating to the chilling effect. The broader the meaning of "hate speech," the greater the potential for it to induce self-censorship.

Yet this more expansive understanding may also have countervailing benefits. We have already seen that the legal meaning of "hate speech" informs its ordinary meaning. But the reverse is also possible: ordinary usage of "hate speech" can serve an *ameliorative* function, by pointing the way towards a more morally consistent account of which groups should be covered by hate speech law (on the idea of conceptual amelioration, see, e.g., Haslanger 2012; Burgess, Cappelen, and Plunkett 2020). If hate speech is prohibited because it harms particular disadvantaged groups, then one might think that this applies, not just to racist and homophobic speech, but to some instances of misogynistic speech as well.[29] Nor is it speculative to think that the ordinary meaning of "hate speech"

---

[28] Two things are worth noting here. First, references to gender-based hate speech are also observable in the general corpus. Second, this divergence between uses of "hate speech" in the journalistic corpus, and the legal understanding of "hate speech," seems to be especially strong evidence that the ordinary understanding of which groups are targeted by hate speech diverges from the legal understanding. This is because—as discussed in Sect. 4—the journalistic corpus is likely to be somewhat biased in the direction of the legal meaning of "hate speech.".

[29] See Richardson-Self (2018; 2021) for an argument to this effect.

might positively impact its legal meaning. In fact, the UK government has been considering such "expansionary proposals" (Law Commission No 348, 2014). So, even if this more expansive account of the targets of hate speech exacerbates concerns relating to the chilling effect, it does not necessarily follow that it is unwelcome overall.

Up until this point, the main mismatch we have considered has to do with types of social groups (e.g., race, religion, sexual orientation, gender). But there is a second notable difference, which has to do with subgroups within these social groups. In the journalistic hate speech corpus, "hate speech" refers overwhelmingly to speech that targets *vulnerable* or *disadvantaged* subgroups—for example, "Jews," "Muslims," "Rohingyas," the "LGBT" community, and so on. This finding also holds in the general corpus (e.g., "hate speech against religious minorities has been stepped up"). By contrast, the text of the Public Order Act—and indeed, the text of hate speech laws more generally—tends to focus on the general categories ("race," "sexual orientation"), without distinguishing between dominant and vulnerable subgroups.

This comparative narrowness may be especially desirable from the perspective of the chilling effect. A particularly strong version of this worry holds that hate speech legislation might silence criticism of dominant groups (e.g., whites, heterosexuals) by the very groups whom they oppress (e.g., people of colour, members of the LGBT community). On this view, which Nadine Strossen (2018) notably articulates, vulnerable groups might refrain from voicing their legitimate recriminations out of fear of being legally sanctioned for doing so. The narrow focus on vulnerable subgroups makes this less likely, by explicitly excluding the possibility that speech directed at dominant social groups might count as "hate speech."

In other respects, this narrowness—and the resulting mismatch between the ordinary meaning of "hate speech" and hate speech law—may nonetheless seem problematic. In terms of fair notice, members of the public might not realize that hate speech laws prohibit speech that targets dominant groups *as well as* vulnerable groups. And one might worry that this mismatch, in turn, could have a damaging impact on democratic trust. Someone who feels that they have been unfairly silenced, because the legal prohibition on hate speech is more expansive than they had initially thought, may lose faith in the democratic process as a result.

In practice, the mismatch in question may be more apparent than real. As Louise Richardson-Self (2018: 1–2) explains in the Australian context, while hate speech law is "ostensibly neutral" between dominant and vulnerable groups, in practice it has almost exclusively been applied to speech that targets vulnerable groups. Nor is this observation confined to the Australian context. There is a clear moral justification for this asymmetrical treatment: as Katharine Gelber (2019: 404) argues, speech directed at subgroups can generate harms that speech directed at dominant groups usually cannot. However, although it may not influence the application of hate speech laws, the mismatch does point to a way in which clarification of the law to bring it better into line with ordinary meaning would improve fair notice.

The broader upshot is this: When it comes to the "who" of hate speech—who are the targets of hate speech—corpus data suggests that the ordinary concept of "hate speech" aligns quite closely with hate speech law. Furthermore, to the extent that the ordinary meaning departs from hate speech law, these departures (the focus on

gender-based hate speech and the focus on subgroups) have the potential to serve an ameliorative purpose.

5.4 "Hate Speech" and Harm

"Hate speech," as the term is ordinarily understood, is negatively valanced, is prototypically considered to be regulatable speech, and usually targets certain groups. But what makes it regulatable, according to most legal philosophers, is not just that it targets certain groups. It is that it has a tendency to *harm* members of these groups (Gelber 2019: 400; Waldron 2012: ch.4; Maitra and McGowan 2012: 4–8; Brown 2017b: 579–80).

This connection between hate speech and harm is omnipresent in ordinary usage. The ordinary meaning of "hate speech," as reflected in our corpora, associates hate speech with a number of different harms. The strongest association is undoubtedly with incitement (Howard 2019). Keywords from the journalistic hate speech corpus signal that references to "hate speech" are very often about speech deployed for the purposes of "inciting violence" or "inciting hatred." Collocation data from both the journalistic and general corpora unambiguously supports this. In the latter, for example, "incite" is the single verb most often used with "hate speech" as a subject.

But the harms ordinarily associated with "hate speech" do not end there. At least two other categories of harms stand out. One salient category has to do, roughly, with violating the dignity of targets (2012: ch.4). Keywords data from the journalistic hate speech corpus, for example, suggests that "hate speech" is commonly used to characterize speech that "insult[s]," "dehumanise[s]," "belittle[s]," and "denigrate[s]." Hate speech, based on this pattern of usage, picks out speech that diminishes, or promotes the inferiority of, its targets.

Another, slightly less salient, category of harms that is manifest in ordinary use concerns the psychological effects of hate speech. For instance, collocation data from the hate speech and general corpora indicates that "hate speech" is often used when discussing speech that "offend[s]," "harass[es]," or "abuse[s]" its targets. The common idea here is that "hate speech" can psychologically "wound" its targets, by causing them to experience negative feelings (e.g., of hurt, distress, shock) (see, e.g., Delgado 1993, on "words that wound").

These examples are not meant to exhaust the harms that ordinary use associates with "hate speech." Rather, they are intended to illustrate, first and foremost, that harm does feature prominently in the ordinary meaning of "hate speech." But we can go further than this first conclusion. The particular *way* in which harm is connected to the ordinary meaning of "hate speech" has notable implications for both fair notice and the chilling effect.

In terms of fair notice, the corpus indicates a meaningful measure of alignment between what hate speech law prohibits, and the way "hate speech" is ordinarily understood. One complication here is that, as mentioned above, hate speech law varies across different countries. But despite this variation, some harms are widely recognized across different bodies of hate speech law. This is perhaps most notably the case with incitement to hatred and/or violence (Brown 2015: 26–28). For example,

the Public Order Act explicitly prohibits "stirring up hatred." Likewise, the Canadian Criminal Code 1985 outlaws "the wilful promotion of hatred against any identifiable group." The fact that incitement features so centrally in the ordinary concept of "hate speech" allows ordinary users to track, in an important respect, actual legal provisions.

Other dimensions of alignment are more specific to the UK context (to which, recall, our journalistic hate speech corpus is especially relevant). The Public Order Act specifically rules out the use of "insulting," "abusive," or "threatening" words that are likely, or intended, to stir up hatred. Ordinary usage mirrors this to a significant degree. As explained above, "hate speech" is frequently associated, in ordinary usage, with speech that "insult[s]" or "abuse[s]" its targets. Conversely, our corpus searches on the terms "abusive," "threatening," and "insulting" show that, when ordinary users employ two or more of these terms together, they are very often using them to refer to hate speech. This provides evidence that there is public awareness of key terms contained in the UK's hate speech law[30]—and, insofar as this is the case, that the public has adequate notice of these legal provisions.

What about the chilling effect? The most important upshot of our results is that the ordinary meaning of "hate speech" is not reducible to "offensive speech." Psychological offence *is* one of the negative effects commonly associated with "hate speech." But ordinary usage suggests that "hate speech" characteristically involves more than simply offence—in particular, "hate speech" prototypically involves harms such as incitement, and often, violations of dignity.

This is morally significant. As Jeremy Waldron (2012: 105) has argued, "offense, however deeply felt, is not a proper object of legislative concern." The reason, for Waldron, is that offence, including deep offence, is inevitable in any society marked by cultural and religious diversity. Indeed, in such a society, "each group's creed seems like an outrage to every other group" (Waldron 2012: 127). To prohibit speech simply on the grounds that it is offensive would therefore give rise to an excessive chilling effect: doing so would be incompatible with maintaining adequate protections for personal, cultural, and religious expression. It is a good thing, then, that "hate speech" is prototypically understood as harmful in ways that go beyond mere offensiveness.

Other aspects of ordinary usage might appear more troubling. Our corpus analysis of "threatening," "abusive," and "insulting" suggests that these three terms have quite different semantic profiles. In other words, they are ordinarily used in quite different ways (see Appendix, Figs. 1, 2, 3).

Prima facie, this may seem worrying from the perspective of the chilling effect. According to Nadine Strossen (2018), who critiques hate speech legislation, it is wrong for the state to prohibit categories of speech that are defined in vague terms. This, she suggests, is because.

---

[30] By contrast, similar searches conducted on an American corpus (the Corpus of Contemporary American English) showed exceedingly few co-occurrences of these terms—and that, when they co-occur, they are typically not related to hate speech.

when an unduly vague law regulates speech […] it inevitably deters people from engaging in constitutionally protected speech for fear that they might run afoul of the law. (2018: 69)

The problem, in our context, is that if the Public Order Act prohibits "threatening," "abusive," or "insulting," language, and if the ordinary meanings of these three terms differ from one another, the statute may seem excessively vague—and, as a result, potentially quite broad—to the public. This vagueness and breadth, Strossen might say, risks leading to excessive self-censorship.

On closer inspection, however, this worry may be overstated. The Public Order Act does not prohibit the use of "threatening," "abusive," or "insulting" speech simpliciter. It prohibits such speech *when it is likely to, or intended to, stir up hatred*. This further "incitement" clause makes the category of prohibited speech more restricted and specific—and, crucially, our corpus data suggests that the public are often aware that "hate speech" involves incitement. So, although the semantic divergence of "threatening," "abusive," and "insulting" may introduce some vagueness in public perceptions of hate speech law, patterns of ordinary usage suggest that this vagueness (and the attending self-censorship) is not as great as it might initially seem.

5.5 "Hate Speech" and Hate

Our analysis so far has assessed four dimensions of the ordinary term "hate speech": its valence; its relationship to legal regulation; whom it targets; and its connection to harm. Up to this point, though, we have said next to nothing about the last "folk platitude" Brown mentions: the idea that "hate speech" is meaningfully connected to emotions or feelings of hate.

This is one of the most vexed definitional questions surrounding "hate speech." The idea that "hate speech" bears an important relationship to hate seems commonsensical—and, accordingly, it has enjoyed fairly broad support among legal theorists (e.g., Post 2009: 123; Strossen 2018: xxiii). But Brown himself, like several other legal theorists, has categorically rejected what he calls the "myth of hate." "The ter[m] 'hate speech'," he concludes, "can be used in cases where no hate or hatred is involved" (2017a: 467; see also Waldron 2012: 34–37).

Ordinary usage suggests that there is in fact a close connection between the ordinary meaning of "hate speech," on the one hand, and emotions or feelings of hate or hatred, on the other. This is particularly visible in the keywords data from our journalistic hate speech corpus: "hatred," "religious hatred," "racial hatred," and "inciting hatred" all feature among the top keywords. What this indicates is that articles where the term "hate speech" appears also tend to be about feelings of hatred.

What might be the nature of this relationship? In principle, the term "hate speech" and hatred could be connected in different ways. Feelings of hatred could *motivate* hate speech; could be *expressed by* hate speech; and, finally, could be *produced by*

hate speech. While our corpus data does not exclude the first two, it points most clearly to the third. The fact that "inciting hatred" features as one of the top key-words for "hate speech" suggests that "hate speech" is commonly used when discussing speech that is liable to produce feelings of hatred. A closer check of 215 concordance lines supports this suggestion. Overwhelmingly, "hatred" features as a potential product of "hate speech" (e.g., "disablist and misogynist hate speech contributes to a climate of hatred").

This may appear to contradict Brown's rejection of the "myth of hate." But on closer examination, this appearance is misleading for two reasons. The first reason relates to context. The relationship between feelings of hatred and "hate speech" appeared more strongly in the journalistic hate speech corpus than in the general corpus. This is quite likely no accident. The journalistic hate speech corpus, as we have seen, is based on UK news articles. Given that hate speech laws in the UK prohibit stirring up, or inciting, hatred, it is therefore unsurprising that, *in this national context,* we should find a close association between "hate speech" and hatred.

This result leaves open the possibility that, in a *different* national context, there may be no meaningful connection between "hate speech," as ordinarily understood, and the feeling of hatred. So, our finding is in principle compatible with rejecting the "myth of hate" when considering "hate speech" from a comparative, cross-national, perspective (as Brown does).

The second reason concerns the nature of "ordinary meaning." When assessing the ordinary meaning of "hate speech," Brown is predominantly concerned with *permissible* meanings. What he shows is that it may be linguistically *permissible*—even if it may be unusual—to use the term "hate speech" in contexts entirely unrelated to the feeling of hatred. By contrast, and as discussed in Sect. 3, insofar as corpus methods are used to look at data about linguistic frequency, they provide insight into the *prototypical* (or most salient) meaning of "hate speech." What our analysis suggests, then, is that, in the UK, "hate speech" is prototypically connected to the production of hatred. This is consistent with thinking that it might nonetheless be permissible (if unusual) to use the term "hate speech" differently.

These two points show why our findings relating to the place of hate in "hate speech" are not necessarily at odds with Brown's. But, perhaps more importantly, they also illustrate two core advantages of using corpus methods to investigate "hate speech." First, corpus methods can provide insight into the ordinary meaning of "hate speech" that is tailored to specific national or subnational contexts. This is crucially important for policy guidance. To determine whether hate speech laws in a specific context satisfy requirements of fair notice, or are likely to lead to a chilling effect, we need to know about the ordinary meaning of "hate speech" *in that context.*

Second, corpus methods can provide insight into prototypical meaning. This, too, is of vital importance when thinking about policy. As Sect. 3 discussed, the prototypical meaning of "hate speech" may be more relevant than its permissible meaning to determining whether the public has sufficient notice of hate speech laws (or, for that matter, whether such laws are likely to lead to a chilling effect). What is more, Mouritsen (2011: 161) provides evidence that, while individual intuitions can

be a relatively secure guide when it comes to permissible meanings, they are highly misleading as guides to prototypical meaning. Thus, the corpus method is doubly useful: it yields information about ordinary meaning that is needed to assess hate speech policies in specific contexts; and this information would otherwise be difficult to access.

6 Conclusion

Our analysis has two core upshots for legal and philosophical investigations of hate speech. The first upshot is methodological. We have argued—and provided a proof of concept—for using corpus linguistics to ascertain the ordinary meaning of "hate speech." Looking at how ordinary speakers use "hate speech," on a large scale and in natural communicative environments, can help us assess, in ways that would otherwise be difficult to do, some of the most prominent ethical and legal objections to hate speech law: e.g., that it risks violating fair notice; that it jeopardizes democratic norms; and that it undermines free speech. Indeed, a corpus approach to "hate speech" allows us to gauge more precisely how, and to what degree, these objections really apply within a particular social community. Thus, it helps bridge the divide between abstract normative considerations, and the real-world contexts in which they are meant to apply. In doing so, it provides an empirically grounded foothold for resolving protracted disputes between advocates and opponents of hate speech law.

Our investigation also yields substantive, and normatively relevant, insights about "hate speech." One recurrent insight is that ordinary usage in the UK tracks, to a significant degree, actual hate speech law. It reflects an understanding, in particular, that hate speech is not just frowned upon, but legally regulated; that it targets certain groups; that it is harmful speech; and that, in the UK at least, it is closely (if not necessarily) associated with inciting hatred. As we have seen, this alignment is morally significant. It suggests, for instance, that the UK's Public Order Act needn't violate fair notice or erode trust in democratic norms. It also contributes to allaying fears relating to free speech—not least, because the public recognizes that prototypical "hate speech" is a narrower category than "offensive speech." The alignment is by no means total. But even where ordinary understanding departs from the law, it may serve an ameliorative purpose, for example, by encouraging a normatively appropriate broadening, in the law, of which groups can be targeted by "hate speech"; or by recommending that the law's de facto focus on speech targeting vulnerable groups be made more explicit.

These substantive conclusions are not meant to be definitive. They, like the preliminary evidence we have supplied, are defeasible. But what they illustrate is the prospect, made possible by corpus methods, of moving past merely rhetorical appeals to values such as free speech, democracy, and the rule of law, to a more serious consideration of what these values recommend in terms of real-world speech regulation.

Appendix

Journalistic hate speech corpus

Top ten keywords (word clusters) of the journalistic hate speech corpus (reference corpus: *English Web 2020 enTenTen20*) (Table 1)

Table 1 Keywords—words clusters

Item	Freq. (focus)	Freq. (reference)	Relative freq. (focus)	Relative freq. (reference)	Score
text of report	60	111	312,64,166	0,0072	311,4
hate crime	114	26,400	594,01,917	1,71,299	219,32
online hatespeech	30	0	156,32,083	0	157,32
fake news	28	1285	145,89,944	0,08,338	135,59
social media company	26	1397	135,47,806	0,09,065	125,14
freedom of speech	78	41,362	406,43,417	2,68,381	110,6
free speech	103	59,722	536,70,154	3,87,511	110,3
religious hatred	24	2648	125,05,666	0,17,182	107,57
racial hatred	24	3070	125,05,666	0,1992	105,12
freedom of expression	86	58,777	448,11,972	3,8138	93,298
hatespeech law	16	0	83,37,111	0	84,371

Keywords of the journalistic hate speech corpus portraying hate speech-related actions (single words and word clusters) (Table 2)

Table 2 Keywords—actions related with hate speech

Topics	Keywords (frequency focus corpus / score)
Provoking	Incite (93/117.53), inciting violence (12/56.56), inciting hatred (7/34.67)
Expanding	spreading hatespeech (10/53.1), propagating hatespeech (8/42.68), spreading hate (6/31.56), aggravating (6/21.06), polarise (6/19.55)
Minimising	Normalise (12/37.36), trivialise (4/19.43)
Performing hate speech	Dehumanise (7/31.27), insult (50/30.2), denigrate (9/25.78), belittle (8/22.85), terrorise (5/19.26), criticise (25/20.59), demonise (5/19.87)
Countering	Apologise (11/16.96), stop hate (6/32.14), banning hatespeech (5/27.05), prosecute (43/24.09), stop hate for profit (5/27.05), criminalise (6/25.32)/ criminalize (11/22.33)

Keywords of the hate speech corpus portraying social actors (single words and word clusters) (Table 3)

NB: Table 1 does not include nationality adjectives, in their majority references to the countries that have been associated with hate speech issues in the UK media

Table 3 Keywords—social actors associated with hate speech

Social actors	Keywords (frequency focus corpus/score)
Hate speech promoters	white supremacist (12/43.09), hate preacher (8/41.26), bigot (6/17.93), hate Fig. (5/26.66), violent extremist (5/23.36)
Sex/gender related	LGBT (64/29.1), LGBTI (10/27.05), gay people (9/26.62), transgender people (8/27.14), paedophile (7/24.07)
Ethnic/religion identified	Kikuyu (15/65.57), rohingya (22/49.81), boer (14/31.97), serb (12/29.78), Muslims (61/17.43), korean resident (5/26.57), jewish group (6/25.36), ethnic group (27/23.88), muslim woman (10/22.81)
People—politics	far-right group (5/25.16), rightwing (17/57.77), neo-nazi (7/22.54)
Social groups	hate group (8/35), rights group (19/28.95), civil rights group (7/30.2), justice group (6/28.14), racial justice group (5/26.98)
Others	white farmer (12/57.64), justice minister (7/31.48), indian savage (5/26.92), vindicator (9/43.21), censor (30/34.68), outlaw (20/20.97), campaigner (21/20.86), rapist (9/20.11)

Keywords of the hate speech corpus portraying entities related to hate speech (single words and word clusters) (Table 4)

Table 4 Keywords—entities related to hate speech

Topics	Keywords
Hate action	hate crime, hate incident, incitement, hatred, hate, sectarianism, bigotry, extremism, intolerance, hostility, intimidation, discrimination, animosity
Verbal action or material	spread of hatespeech, hate message, use of hatespeech, hateful comment, extremist speech, offensive comment, instance of hatespeech, political hatespeech, insulting word, derogatory term, hatespeech/hate-speech, defamation, blasphemy
Sex/gender	lgbt-free zone, pornography, paedophilia, homophobia, misogyny, sexism, sexual orientation, gender identity
Race/ethnicity	racial hatred, racist abuse, Holocaust denial, antisemitism, xenophobia, racism, basis of race, ethnic origin, skin colour, national origin, ethnicity, nationhood
Religion	religious hatred
Against hate	counter-speech, anti-pornography, anti-racism, anti-discrimination, anti-hatespeech, anti-defamation, moderation

Collocations for "hate speech" (grammatical patterns of modifiers, nouns modified by "hate speech", verbs with "hate speech" as object, and words associated with "hate speech" with "and"/"or" conjunctions) (Table 5)

Table 5 Collocations with "hate speech"—main patterns

Grammatical patterns	Top-ten collocates with frequency and association score
(1) *modifiers of 'hatespeech'* (143/12.6) e.g. "…disablist and **misogynist hate speech** contributes to a climate of hatred"	online (26/11.6), illegal (15/11.5), misogynist (3/9.38), anti-Muslim (3/9.23), political (4/8.71), religious (3/8.6)
(2) *nouns modified by 'hatespeech' (165/14.6)* e.g., "… British **hatespeech laws**, which work to prevent discord among ethnic and religious minorities, he added."	law (16/10.2), policy (8/10.1), bill (6/9.93), legislation (5/9.56), rule (3/8.98), case (3/8.79), crime (3/8.34)
(3) *verbs with 'hatespeech' as object (334/33)* e.g., "He said that legal action against people who **spread hatespeech** had to be conducted without violating citizens' rights to express themselves…"	spread (19/10.5), constitute (13/10.1), combat (13/10), tackle (11/9.74), remove (11/9.59), use (13/9.56), propagate (8/9.42), direct (8/9.36), define (8/9.35), address (8/9.35)
(5) *Nouns with which 'hatespeech' shares and/or constructions* (221/19.5) e.g., "… there was consensus that the fight against **hatespeech and hate crimes** is an important matter, both in Turkey and around the world."	crime (16/10.9), news (9/10,2), incitement (9/10.2), abuse (9/10.2), extremism (8/10.1), misinformation (8/10.1), violence (9/9.82), harassment (6/9.67), content (5/9.39), speech (5/9.28)

General corpus (*English Web 2020 enTenTen20*)

NB: The examples included in the tables to illustrate the grammatical patterns were obtained with the GDEX (*Good Dictionary Example*) tool of *Sketch Engine*, which provides examples that would serve dictionary or teaching purposes.

Collocations of the general corpus reflecting actions related to hate speech (grammatical patterns of verbs with "hate speech" as subject and object) (Table 6)

Table 6 Collocates—actions related to hate speech

Grammatical pattern	Top-ten collocates with frequency and association score
(1) *verbs with "hate speech" as subject* (4484/11.2) e.g., "That type of **hate speech fuels** the divisive machine." e.g., "The letters are rife with the most disturbing **hate speech targeting** females."	incite (28/5,88), spew (13/4,22), victimize (7/3,8), target (118/3,18), fuel (29/2,49), harm (9/2,28), offend (7/2,13), circulate (8/1,32), thrive (9/1,02), spread (23/0,91)
(2) *verbs with "hate speech" as object* (13,249/33.2) e.g., "The conference participants will discuss national minority protection and promotion, protection of LGBTI rights and **combating hate speech**" e.g., "Judicial experts and politicians cited these articles to highlight the difficulties of **criminalizing hate speech**"	spew (122/6,29), criminalize (86/5,84), counter (313/5,47), combat (432/5,09), spout (35/4,83), curb (140/4,81), condone (40/4,66), outlaw (52/4,61), criminalise (12/4,48), propagate (67/4,48)

Collocations of the general corpus reflecting social actors associated with hate speech as victims (grammatical patterns of modifiers of "hate speech" and "hate speech" followed by *against* prepositional phrase) (Table 7)

Table 7 Collocates—social actors associated with hate speech

Grammatical pattern	Top-ten collocates (frequency / association score)
(1) *modifiers of "hate speech"* (5203/13.1) e.g. "Does it protect **racist hate speech** as well as political dissent?" e.g., "Thus, public affirmation of Christian sexual morality is judged to be **homophobic hate speech** that must be silenced."	racist (265/5), antisemitic (58/5), homophobic (62/4,83), anti-Semitic (75/4,75), anti-gay (42/4,3), gender-based (44/4,12), sexist (42/4,11), Islamophobic (25/4,02), xenophobic (28/3,91), anti-semitic (23/3,85)
(2) *"hate speech" against…* (754/1.89) e.g., "The **hate speech against** religious **minorities** has been stepped up" e.g., "Its leader has engaged in direct **hate speech against Muslims**"	non-Moslem (20/9,51), migrant (24/8,16), Roma (20/7,83), refugee (18/7,77), Muslim (52/7,27), minority (50/7,22), gay (19/7,16), Rohingya (10/6,98), homosexual (14/6,96)

Collocations of the general corpus reflecting hate speech-countering actions and measures (grammatical patterns of the prepositional phrases *for, on* and *against* in pre-modifying position) (Table 8)

Table 8 Collocates—hate speech countering actions and measures

Grammatical pattern	Top-ten collocates with frequency and association score
(1) … *For "hate speech"* (1049/2.63) e.g., "His subsequent **conviction** for **hate speech** was later overturned on appeal." e.g., "In Britain you can be **arrested** for **hate speech** because it affects "community cohesion"	prosecute (41/6,28), infamous (7/5,32), convict (13/4,96), prosecution (14/4,59), tolerance (22/4,26), investigate (10/4,19), indict (7/4,15), ban (8/3,99), conviction (12/3,7), arrest (25/3,52)
(2) … *on "hate speech"* (959/2.4) e.g., "This policy includes a **prohibition** on **hate speech** and protects our users" e.g., "The **crackdown** on **hate speech** and white supremacist content came after YouTube declined to delete homophobic videos targeting a Vox reporter."	Conduct (9/7,39), Action (25/5,72), legislation (25/4,98), standard (22/4,66), policy (97/4,54), crackdown (29/4,52), prohibition (20/4,45), rule (41/4,33), border (19/4,29), law (63/4,09)
(3) … *against "hate speech"* (800/2.01) e.g., "France has strict **laws** against **hate speech**" e.g., "We do that, so why not analogously take **measures** against **hate speech**?"	guideline (7/8,05), policy (71/7,72), law (165/7,49), rule (43/7,42), legislate (8/7,27), legislation (11/7,1), Action (7/5,59), prohibition (24/5,54), stance (13/5,49), measure (14/4,52)

Some concordances reflecting a representation of hate speech as a natural force and in terms of war:

– volunteer groups to **fight** hate speech **spreading at "lightning speed"** […] (business-standard.com)
– [war metaphor and natural forces metaphor]
– The **battle against hate speech** is a universal and international one, and while geoblocking helps to **protect** people in certain countries from accessing and seeing hateful content, it does not **combat** the issue as a whole. (media-diversity.org)
   [war metaphor]
– Hate speech is **spreading like wildfire** in social media. We must **extinguish** it, […] (news.un.org)
   [natural forces metaphor and war metaphor]

Concordance analysis for the patterns "threatening and/or abusive" and "threatening and/or insulting."

NB: The first 100 concordances of the patterns "threatening and/or abusive" and "threatening and/or insulting" were coded according to whether (1) the collocates are associated with hate speech (according to the Public Order Act definition), and (2) the context of the concordance is legal. Online sources of the concordances were checked to help tagging the context of use in those cases where it could not be determined by the co-text or the name of the source alone.

Concordances of the pattern "threatening and/or abusive"

Tables 9, 10.

Table 9 Concordances of the grammatical pattern "threatening and/or abusive"

Hate speech related			Context of use			
Yes	No	?	Legal	Offence	No legal	?
89	10	1	76	13	11	2

Table 10 Correspondences between references to "hate speech" and context of use for the grammatical pattern "threatening and/or abusive"

Hate speech related	Context	Concor-dances	Examples
Yes	Legal	76	… not to use the services we provide via the Site to: </s> <s> Transmit any information or do anything that is unlawful, harmful, **threatening, abusive**, harassing, tortious, defamatory, vulgar, obscene, libellous, invasive of another's privacy, hateful
	Offence	10	Yesterday, three teenage boys—aged 15, 16 and 17—admitted using **threatening or abusive** words or behavior against the two women shortly before trial
	No legal	3	He said those who made **abusive and threatening** posts about the study were "fanatics opposed to the very values of a liberal society"
No	Offence	3	… Lady Scott sentenced a man to 6 years imprisonment after he pled guilty to behaving in an **abusive and threatening** manner, abducting and assaulting his family and attempting to cause an explosion in the flat where they were detained
	No legal	7	"All too often her reaction to any discussion at all which she does not understand or does not like is to become **threatening and abusive**."
?	No legal	1	ESPECIALLY when they are in violation of the "calling anonymously" section of the ACT, as well **as threatening, abusive** behavior. </s> <s> NOTE: THEY ARE NOT DIRECTLY AFFILIATED WITH CITIBANK AND THUS CAN NOT GET YOUR CREDIT HISTORY

Concordances of the pattern "threatening and/or insulting"

Tables 11 and 12

Table 11 Concordances of the grammatical pattern "threatening and/or insulting"

Hate speech related		Context of use		
Yes	No	Legal	Offence	No legal
74	26	48	13	39

Table 12 Correspondences between references to "hate speech" and context of use for the grammatical pattern "threatening and/or insulting"

Hate speech related	Use	Concor-dances	Examples
Yes	Legal	47	Rudeness, profanity, **threatening, insulting** posts, personal attacks, defamatory or inflammatory posts will not be tolerated and are a breach of the […]
	Offence	11	Prosecutor Anwen Walker said the sign was **insulting, threatening** and in breach of public order legislation
	No legal	16	Yes, yes, words have meaning, and we shouldn't have to put up with **insulting, threatening** behaviour, or discomfortingly loud noise of any discription. (sic.) But there are already laws for that
No	Legal	1	Aggression: Frowning, snarling, baring teeth, staring, with redden face, rigid body, clenched fists, and large, **threatening and insulting** gestures, you display unexpectedly sudden movements, intrude on someone's space, and deliver […]
	Offence	2	Numerous testimonies indicated that the soldiers' language to women was rude, **insulting, threatening**, and vulgar
	No legal	23	Remove shaming, shouting, ignoring, **threatening, insulting**, bullying and spanking from your parenting tool kit

Uses of "threatening", "abusive" and "insulting"

NB: The contrasting collocates of the three descriptors ("threatening", "abusive" and "insulting") were examined with the Word Sketch Difference tool of *Sketch Engine*. Figures 1, 2, 3 below provide the collocates for the grammatical pattern "… and/or …", which allows us to identify characteristics associated with the concepts.
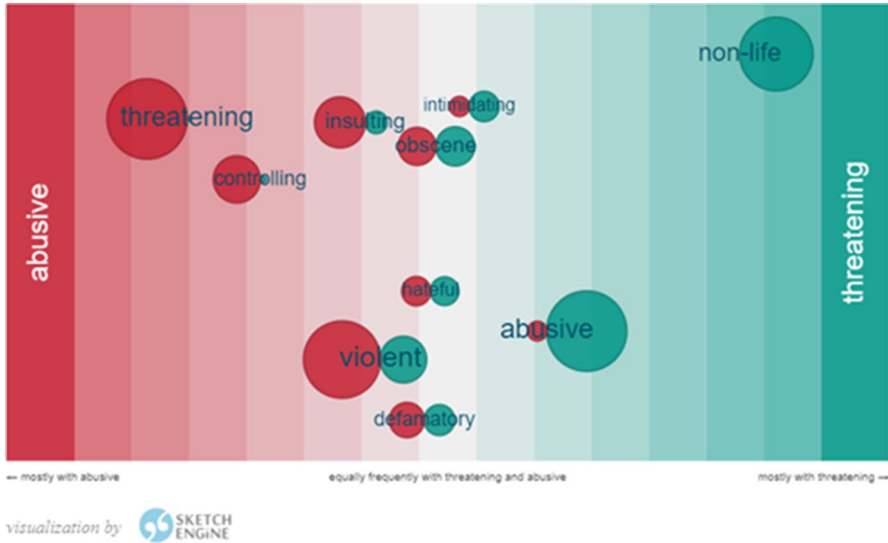
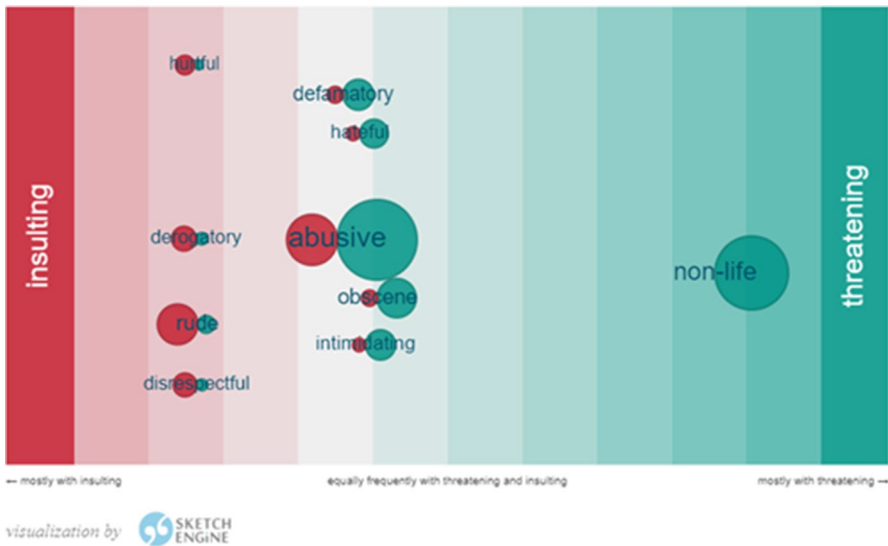**Fig. 1** "Threatening" and "abusive" in "and/or X" pattern

Fig. 2 "threatening" and "insulting" in "and/or X" pattern
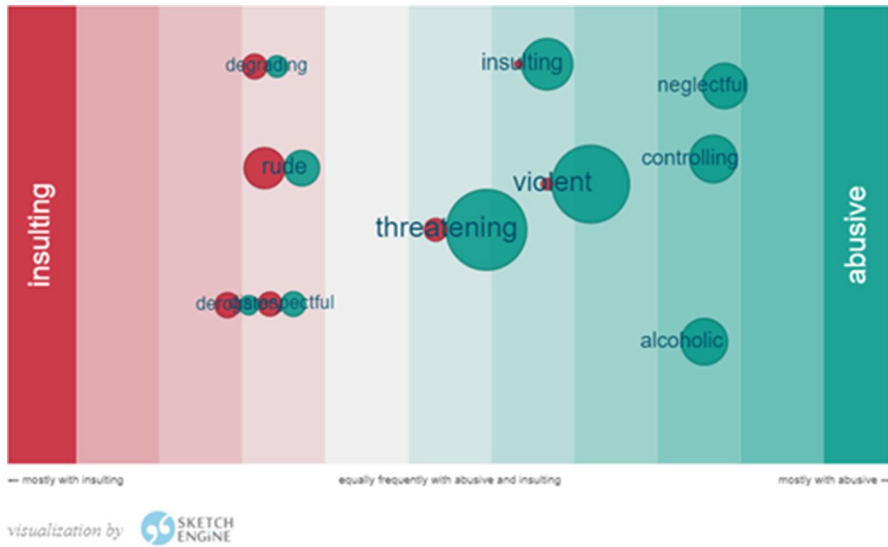
Fig. 3 "Insulting" and "abusive" in "and/or X" pattern

The figures show the top 10 collocates. In the diagrams, collocates are distributed horizontally according to their association with the term considered; the circles show whether the collocates present more or fewer hits with the term.

Declarations

Conflict of interest The authors have no competing interests to declare.

References

Abboud, Leila. "French Presidential Candidate Zemmour Convicted of Hate Speech." *Financial Times.* January 17, 2022.

Alkiviadou, Natalie, Jacob Mchangama, and Raghav Mendiratta. "Global Handbook on Hate Speech Laws." *The Future of Free Speech*, 2020.

Baker, Paul. *Using Corpora in Discourse Analysis*. London/New York: Continuum, 2006.

Brown, Alexander. *Hate Speech Law*. Abingdon: Routledge, 2015.

Brown, Alexander. "The 'Who?' Question in the Hate Speech Debate: Part 1." *Canadian Journal of Law & Jurisprudence* 29 (2016): 275–320.

Brown, Alexander. "What Is Hate Speech? Part 1: The Myth of Hate." *Law and Philosophy* 36 (2017): 419–68.

Brown, Alexander. "What Is Hate Speech? Part 2: Family Resemblances." *Law and Philosophy* 36 (2017): 561–613.

Burgess, Alexis, Herman Cappelen, and David Plunkett, eds. *Conceptual Engineering and Conceptual Ethics*. Oxford: Oxford University Press, 2020

Culpeper, Jonathan. "Impoliteness and Hate Speech." *Journal of Pragmatics*, 2021, 1–8.

Delgado, Richard. "Words That Wound." In *Words That Wound*, edited by Mari Matsuda, 89–110. Boulder: Westview Press, 1993.

Futtner, Nora and Natalia Brusco. "Hate Speech Is On The Rise." *Geneva International Centre for Justice.* March 2021. Available at: https://www.gicj.org/images/2021/Hate_Speech_is_On_the__Rise-FINAL_1.pdf

Gelber, Katharine. "Differentiating Hate Speech: A Systemic Discrimination Approach." *Critical Review of International Social and Political Philosophy* 24 (2019): 393–414.

Gries, Stefan. "Corpora and Legal Interpretation." In *Routledge Handbook of Forensic Linguistics*, edited by Alison Johnson, 628–43. London: Routledge, 2020.

Haslanger, Sally. *Resisting Reality*. Oxford: Oxford University Press, 2012.

"Hate Speech." *Cambridge English Dictionary.* Available at: https://dictionary.cambridge.org/dictionary/english/hate-speech

Howard, Jeffrey. "Dangerous Speech." *Philosophy & Public Affairs* 47 (2019): 208–54.

Howard, Jeffrey. "Terror, Hate, and the Demands of Counterspeech." *British Journal of Political Science* 51 (2021): 924–39.

Law Commission. "Report n.348—Hate Crime: Should the Current Offences Be Extended?" May 2014.

Lee, Thomas, and Stephen Mouritsen. "Judging Ordinary Meaning." *Yale Law Journal* 127 (2017): 788–879.

Lee, Thomas, and Stephen Mouritsen. "The Corpus and the Critics." *University Chicago Law Review* 88 (2021): 275–366.

Lepoutre, Maxime. *Democratic Speech in Divided Times.* Oxford: Oxford University Press, 2021.

Maitra, Ishani, and Mary Kate McGowan. "Introduction and Overview." In *Speech and Harm*, edited by Ishani Maitra and Mary Kate McGowan, 1–23. Oxford: Oxford University Press, 2012.

McGowan, Mary Kate. *Just Words*. Oxford: Oxford University Press, 2019.

Mouritsen, Stephen. "Hard Cases and Hard Data." *Columbia Science & Technology Law Review* 13 (2011): 156–205.

Post, Robert. "Hate Speech." In *Extreme Speech and Democracy*, edited by Ivan Hare and James Weinstein, 123-138. Oxford: Oxford University Press, 2009.

Richardson-Self, Louise. *Hate Speech Against Women Online*. London: Rowman & Littlefield, 2021.

Richardson-Self, Louise. "Offending White Men." *Feminist Philosophy Quarterly* 4 (2018): 1–24.

Solan, Lawrence, and Tammy Gales. "Finding Ordinary Meaning in Law." *International Journal Legal Discourse* (2016): 253–76.

Strossen, Nadine. *HATE: Why We Should Resist It with Free Speech, Not Censorship*. Oxford: Oxford University Press, 2018.

Sytsma, Justin. "Ordinary Meaning and Consilience of Evidence." In *Advances in Experimental Philosophy of Law*, edited by Stefan Magen and Karolina Prochownik. London: Bloomsbury, forthcoming.

Sytsma, Justin, Roland Bluhm, Pascale Willemsen, and Kevin Reuter. "Causal Attributions and Corpus Analysis." In *Methodological Advances in Experimental Philosophy*, edited by Eugen Fischer and Mark Curtis, 209–38. London: Bloomsbury Academic, 2019.

Tobia, Kevin. "Testing Ordinary Meaning." *Harvard Law Review* 134 (2020): 726–806.

Tobia, Kevin. "The Corpus and the Courts." *University of Chicago Law Review*, (2021): 1–36.

Tobia, Kevin, Brian Slocum, and Victoria Nourse. "Ordinary Meaning and Ordinary People." *University of Pennsylvania Law Review* 171 (2023): 1–114.

Waldron, Jeremy. *The Harm in Hate Speech*. Cambridge: Harvard University Press, 2012.

Warriner, Amy, Victor Kuperman, and Marc Brysbaert, "Norms of Valence, Arousal, and Dominance for 13,915 English Lemmas." *Behavioral Research* 45 (2013): 1191–207.

Xiao, Richard. "Collocation." In *The Cambridge Handbook of English Corpus Linguistics,* edited by Douglas Biber and Randi Reppen, 106-124. Cambridge: Cambridge University Press, 2015.