

Hate speech and speech-act theory.

How to express and restrain hate in language use

Jörg Meibauer (Mainz)

Abstract

Hate speech is analyzed as a kind of insulting, that is an expressive speech act aiming at triggering a certain feeling in the addressee. In particular, it is shown that aggressive emotions like hate or contempt can also be signaled by other linguistic devices besides ethnic slurs, for instance, prosodical, morphological, and syntactic patterns. These devices are potential markers of hate speech that a hate speaker uses, and a target understands, and may be conceived of as markers of illocutionary force. Furthermore, the analysis can cover indirect cases of hate speech. Special attention is drawn to the occasionally overlooked fact that potential hate speakers may engage in self-restraint: They can use weaker terms, they can use more indirect strategies, they can retract or deny their hate speech, and last but not least, they can refrain from using hateful speech act altogether.

1 Introduction

Like other pragmatic terms, *hate speech* is unlikely to have a single correct definition that is superior to all other possible definitions. Roughly, one can distinguish between broad definitions and narrow definitions. Broad definitions integrate as much as possible of what one wants to put under the term. They tend to enumerate such aspects, for example, by providing a list of forms of expression and a list of personal characteristics (of the targeted group) that may be relevant in the production of hate speech. Narrow definitions, on the other hand, refer to a specific framework and attempt to define hate speech within that framework. In this sense, I aim to provide a speech-act theoretical analysis of hate speech.

By and large, one can distinguish three strands in the linguistic analysis of hate speech. First, there are lexical analyses of the meaning and use of ethnic slurs (for instance, Pullum 2018; Technau 2018). Second, there are speech-act approaches focusing on the discriminatory, subordinating, and oppressive/exercitive character of hate speech (for instance, Graumann/Wintermantel 2007; Maitra 2012; McGowan 2012). Third, hate speech is studied from the perspective of corpus analysis, taking authentic cases of online and offline hate speech into account (for instance, Jaki/Geiger 2023; Jaki/de Smedt 2019). Moreover, hate speech has been considered in the context of verbal aggression, pejoration, and, impoliteness (cf. Bonacchi 2017; Culpeper 2021; Finkbeiner/Meibauer/Wiese 2016b). An overview provides Guillén-Nieto (2023). In this article, an integrated view is put forward. Hate speech is analyzed as a kind of insult, an expressive speech act aiming at triggering a certain feeling in the addressee.

Assuming that the main tasks of speech-act theory are the definition of speech-act types, their classification, and the explanation of their assignment, the following basic assumptions suggest themselves. Insults are speech-act types in their own right; they belong to the class of expressives; they can be recognized by certain illocutionary indicators and/or the context of use. Hate speech should also have these characteristics if it is counted among the insults. What needs to be discussed is whether hate speech is a separate speech-act type, belongs to a separate speech-act class, and has very specific illocutionary indicators and contexts of use. The question is therefore whether there is an argument for or against subsuming hate speech under insults.

2 Speech-act theoretical frameworks

Though there are numerous speech-act theories on the market, I will take the framework developed by Searle (1969, 1979a, 1979b) and Vanderveken (1990) as my point of departure. By and large, speech-act theory, as a subdiscipline of linguistic pragmatics, has the following theoretical agenda.

First, it must define the speech-act types that are relevant for the language use of a speech community. Some of these speech-act types may be universal, for instance the assertion; some of them may be language-specific. The definition of a speech-act type usually works with a set of necessary and sufficient conditions for the successful realization of that speech-act type. With respect to the issue of hate speech, we have to ask whether hate speech constitutes a separate speech-act type (like, for instance, promising).

Insulting is a social category that is relevant for most societies. This becomes clear when one looks at the law (cf. Frank 2023; Fuchs-Kreiß/Oğlakcıoğlu 2026; Oğlakcıoğlu 2023). For instance, in German law, ‘insulting’ (*Beleidigung*) is criminalized in §185 StGB (*(Strafgesetzbuch)* ‘the German criminal law code’). As Hilgendorf (2008) explains, insulting is directed against ‘reputation’ (Germ. *Ehre*), that is, the social esteem someone enjoys, or ‘human dignity’ (*Menschenwürde*) (cf. also Frank 2023). There is a tendency in jurisprudence to favor human dignity as the most basic norm. Thus, according to criminal law, linguistic expressions are offensive if they represent a ‘manifestation of disrespect or disregard’ (*Kundgabe von Miss- oder Nichtachtung*). It is therefore a question of the respect due to another person (cf. Hilgendorf 2008: 407). Related paragraphs concern ‘incitement to hatred’ (*Volksverhetzung*, §130 StGB) and ‘inciting insult’ (*Volksverhetzende Beleidigung*, §192a StGB). Thus, German law has three categories of insulting acts. While §185 covers insulting in a general way (for instance, saying *You bloody idiot!* to a teacher at school), §130 concerns insulting of national, racial, religious groups, and groups defined by their ethnic origin, or members of these groups. Part of §130 is also the prohibition of Holocaust Denial. The recently introduced §192a targets hate messages that are addressed directly to the people concerned. These were not covered by the existing criminal provisions, that is, §130 and §185. In most cases, there is no incitement to hatred (§130) because the message is not publicly disseminated. For an insult to be punishable (§130), a concrete reference to the person concerned is required. In sum, then, the law acknowledges the connection between insulting and hate speech, and implements the aforementioned categories, with insulting as a super category. That insulting is the most prototypical category, is in line with the postulation of a potential “right to insult” (Clooney/Webb 2017).

Milić (2018: 546f.) criticizes the speech-act approach by Meibauer (2014) because one would have to assume that insulting could occur only as double speech acts. I think that he has speech acts like *You are X!* (with *X* being a slur) in mind, in which we have indicators of asserting and indicators of insulting. But this should turn out as no problem, if one includes the notion of an indirect speech act that implies the dominance of one indicator over the other (cf. Searle 1979b; Meibauer 2019). Thus, his conclusion that “it is safe to assume that acts of insult are not illocutionary acts *per se* but rather a phenomenon which *results* from performing illocutionary acts” (Milić 2018: 546, italic in the original), seems to be premature.¹

Why not assume that “hate speech” does qualify for a separate speech-act type? The reason for rejecting this possibility is that one would not assume categories like “love speech” or “X-speech” (with *X* being an emotion). There is a connection between hate speech and the emotion of hate, but this connection is implemented via the meaning of illocutionary indicators.

Researchers are, of course, free to choose other speech-act etiquettes, such as derogation, demeaning, or denigration.² This may have the advantage to avoid speaking of “insulting” or “hate speech” altogether, while stressing the pejorative attitudes going together with the focused act. However, the task of providing a speech-act theoretical definition still remains.

Second, a speech-act theory must offer a taxonomy of speech-act types, that is, explain how the single speech-act types are related to each other. For instance, one may argue that warnings belong to the class of directives, while greetings belong to the class of expressives. Can we subsume hate speech/insulting under a super category like directives or expressives? Or do we need a separate class of hate-speech acts? Setting aside the Austinian class of exercitives (cf. McGowan 2009, 2012) that is scattered and ill-defined, there are several choices on the table.

Granted that hate speech is a subtype of the speech-act type of insulting, we must ask in which speech-act class insulting belongs. I think that insulting is a member of the expressives. According to Searle (1979a: 15), “the illocutionary point of this class is to express the psychological state specified in the sincerity conditions about a state of affairs in the propositional content” and he mentions the speech-act verbs *thank*, *congratulate*, *apologize*, *condole*, *deplore*, and *welcome*.³ The psychological state of an insulter is that he wants the target to feel insulted. The insulter may or have not an individual feeling of hate or contempt. Like in greetings (for instance, saying *hello*), it depends on whether the society takes an utterance as conventionally signaling the respective state of the mind.⁴ What makes the expressives as the speech-act theoretical home of insulting so appealing is that so-called “bad words” (Sosa 2018) (‘Beleidigungswörter’, cf. Technau 2018) with their typical expressive meaning occur in insults. *Idiot*, *fatso*,

¹ While he is sympathetic of an “adverbialist” approach – some act counts as insult if something is said insultingly – he leaves open that “some insulting acts”, for instance “slurring”, may be considered as “speech acts of their own” (Milić 2018: 546, Fn. 14). However, it is not clear why there should exist a speech act of slurring but not a speech act of insulting.

² For instance, Liu (2020) chooses “derogation” as a separate speech-act type.

³ For a thorough critical discussion of the class of expressives, cf. Finkbeiner (2019).

⁴ This explains why it is possible that a target is insulted without feeling insulted (cf. Milić 2018 for this distinction).

Nazi (for Germans) are such expressive (pejorative) words and they function, as I will argue, as indicators of illocutionary force (cf. Meibauer 2016; Technau 2018; Liu 2020).

Orlando/Saab (2020: 612, italic in the original) postulate that so-called “insultive speech acts” are a subcategory of expressives: “Given a speaker *S*, a sentence *t* containing a slur ‘g’ and a hearer *H*, in uttering *t* *S* insults *H* if and only if *S* intends that the utterance of *t* expresses her contempt for the group *G* (namely, *S* intends that the utterance of *t* expresses her active endorsement of the semantic stereotype of ‘g’).” This definition seems to exclude insults/illocutionary markers like *idiot* or *arsehole*, while it includes slurs like *fatsso* or *spic*. So, the definition seems to target slurring without comprising other kinds of insulting. In this definition, insulting consists in expressing contempt for the targeted group. But this seems to narrow since other feelings may also play a role, for instance, the different types of hate or disgust, etc.

For Liu (2020), hate-speech acts instantiate *derogations*. Slurs are illocutionary force indicating devices that signal these illocutions. Derogations have a propositional condition that does not demand propositions but has at its content “persons or groups of people” (Liu 2020: 1055). The sincerity condition requires “that the speaker must have negative attitudes (for instance, contempt, disrespect, etc.)” towards the addressees (ibid.). (Here, a distinction between second-person utterances and third-person utterances would be relevant.) The essential condition points to the class membership of derogations: they are declaratives. This means that the hate speaker “must intend to enforce a norm against the target, unless the speaker speaks with the institutional authority of the office he or she occupies” (ibid.).

However, this is not plausible. In Searle’s (1979a) taxonomy, declaratives bring about a certain state of affairs, which is nicely put into the slogan “saying makes it so” (Searle 1979a: 16). But when I say *Roman is a polack*, I do not bring about the state that Roman is a polack. I certainly want to implement a norm against “polacks” (Roman) but this is not backed up by the conventions of language use. Whether individual speakers can arrogate to themselves an authority that institutions otherwise have is doubtful (but cf. Maitra 2012).

Third, and most demanding, we have to explain how the production and comprehension (assignment) of speech acts works. For instance, how can addressees find out that they are the target of a hate speaker? A typical answer is to postulate so-called illocutionary indicators (or markers) that signal (in a specific context) the force of a hate-speech act. Illocutionary indicators can be found on the linguistic levels of phonology and prosody, punctuation, morphology, syntax, semantics, lexicology, and pragmatics (cf. Finkbeiner/Meibauer/Wiese 2016a; Knoblock 2022; Meibauer 2022, 2026).⁵ While ethnic slurs are mostly discussed in the philosophy of language, it is important to keep in mind that there are many other types of slurs. “A sampler of slurs” is provided by Bach (2018: 74f.). It goes without saying that also the context can figure as an indicating factor.

⁵ A selection of such devices will comprise pejorative prosody (cf. Sendlmeier/Steffen/Bartels 2016), pejorative word formation (cf. Meibauer 2013), demonstratives (cf. Averintseva-Klisch 2016), generic sentences (cf. d’Avis 2016; Leslie 2017), certain syntactic constructions (cf. d’Avis/Meibauer 2013), and numerous lexical items (cf. Technau 2018).

Because slurs like *polack* have a double nature oscillating between a denotative component (they refer to Polish people) and a pejorative component (Polish people are denigrated), several researchers have assumed that this projects to the level of speech acts. Thus, Tenchini/Frigerio (2016), Camp (2018) and Liu (2020) assume double speech-act analyses.

An explicit double speech-act analysis of ethnic slurs has been put forward by Tenchini/Frigerio (2016). For them, one speech act is the expressive speech act of demeaning or denigrating, and the other speech act is an assertion, question or whatever. Similarly, Liu (2020) assumes that one speech act can comprise several direct illocutionary acts. So, in *Yao is a chic*, two acts are realized, an assertion and a derogation. Also, Camp (2018) endorses a “dual act” analysis, arguing that slurs have a denotative and a perspectival component that lend themselves to associated acts.

However, the assumption of the simultaneous realization of two or more illocutions is problematic. Numerous arguments were put forward for the claim that only one illocution (the primary, indirect illocution) matters for the hearer’s uptake (cf. Meibauer 2019; Searle 1979b). For instance, *There are many polacks in my street* is secondarily asserting (via the declarative sentence type) while its primary force is insulting (via the use of *polacks*). Thus, the illocutionary force indicator signaling the insult (hate speech) dominates the other indicator.

Finally, hate-speech messages may be conveyed via conversational implicature. Though conversational implicatures (for instance, with respect to dogwhistling and the operation of code words, cf. Henderson/McCready 202; Khoo 2017; Saul 2018, 2024) are not speech acts themselves, the relevant pragmatic inferences are nevertheless based on speech acts. If one adopts a speech-act approach to hate speech, this interface should be modelled, too.

3 Hate speech as insulting

Among others, Anderson (2018), Meibauer (2016), Milić (2018), Orlando/Saab (2020), and Tenchini/Frigerio (2016) endorse the analysis of hate speech as an act of insulting. Yet, as far as I see, my proposal is the only one giving an explicit definition of insulting in terms of felicity conditions (rules). Most researchers, it seems, are afraid of giving such a definition and think about insults in more general terms (for instance, Anderson 2018).

Searle (1969) hypothesized that there are linguistic means that signal the illocutionary force of an utterance to the hearer. According to him, the most prominent illocutionary indicators (or “illocutionary force indicating devices”, or IFIDs for short) were the explicit performative constructions, next to the sentence type, verbal mood, modals, verbs expressing an attitude, modal particles, and prosody or punctuation.

Let us take an example. If someone addresses (1) to Bob, who is the proud owner of a yellow Hummer, Bob probably will feel insulted.

(1) I really hate these damn SUV drivers!

Several indicators of hate speech are built into this utterance: the construction *I hate X*, directly expressing the emotion of hate, which is intensified by the adverb *really*, the demonstrative

construction *these X*, signaling distance, together with the expressive device *damn*, targeted as the categorized class of SUV drivers.⁶

Pejorative elements are connected to the system of emotions. Brown (2017a, 2017b) argues at length that hate speech having to do with “hate” is a “myth”. According to him, other emotions such as mere contempt are sufficient. Hate researchers like Sternberg/Sternberg (2008) subclassify hate, assuming that one can distinguish between cool hate (disgust), hot hate (anger/fear), cold hate (devaluation/diminution), boiling hate (revulsion), simmering hate (loathing), seething hate (revilement), burning hate (need for annihilation). Vendrell Ferran (2024) differentiates between just four types of hate. The ANCODI hypothesis entails that a mixture of anger, contempt and disgust leads to hateful intergroup behavior (cf. Matsumoto/Hwang/Frank 2017).⁷ How the composition of pejorative indicators mirror a certain complex emotional state (with a specific mixture of emotions) of an individual, is presently unknown.

While the hate speakers express (part of) their emotional state, it is reasonable to assume that they want to influence the emotional state of their addressees. This is in line with the general characterization of insulting put forward by Neu (2008: 31, italic in the original), who portrays insulting as an attempt to trigger in the addressee a certain feeling, namely the feeling of being insulted: “Perhaps the most adequate general characterization of feeling insulted is that it is kind of pain cause by disappointed expectations of attention and respect. The relevant expectations (...) are normative rather than predictive, involving as they do considerations about how we *should* be thought of and treated; and the disappointment is a matter of shock rather than simple surprise.”

It is reasonable to assume, then, that in insulting, the speaker tries to trigger in the addressee the feeling of being insulted.

The constitutive rule (essential condition) captures the essential force (or point) of a speech act type. My proposal is to rely on Neu’s insight that the insulter wants to trigger a certain feeling in the victim:

- (2) Constitutive rule
Counts as an undertaking to the effect that H feels insulted.

What a feeling of being insulted really amounts to, has to be explained by emotional psychology; this is not the task of speech-act theory.

What about restrictions on the content of an insult? Concerning the content, we can draw a distinction between propositional content *p* that can be evaluated in terms of truth and falsity (truth-conditional content) and expressive modifications *e* of that content. Ethnic slurs (partly) possess expressive (evaluative) content.

It can be left open how *e* is to be specified linguistically. There are a number of semantic and pragmatic approaches to the explanation of *e*. For example, presupposition, conventional impli-

⁶ Note that Bob is not directly addressed, so that pragmatic inferences come also into play. (The speaker knows that Bob owns a SUV, so that Bob is a member of the targeted group.)

⁷ An excellent overview is given by Fischer et al. (2018). For the emotion of disgust, cf. Katzir/Hoffmann/Liberman (2019); for the emotion of contempt, cf. Fischer/Giner-Sorolla (2016); Jeshion (2018).

capture, and conversational implicature have been proposed as carriers of meaning (cf. Hess 2021). Other approaches treat *e* as a stereotype (cf. Croom 2013; Orlando/Saab 2020), as a perspective (cf. Camp 2018), or emphasize the social dimension of *e* (cf. Lepore/Stone 2018; Nunberg 2018). What matters here is only the difference between propositional and expressive meaning, which seems well founded (cf. Potts 2007).

It is obvious that insulting may happen only with *p*, only with *e*, and with a combination of both:

- (3) a. You are looking like your grandmother. [context: addressee hates her grandmother]
 b. Fatso! [context: addressee has a high body mass index]
 c. Dr Smith, that asshole, has just called.

In (3a), it is only the context that makes the utterance insulting. In (3b), a (non-ethnic) slur exhausts the utterance and makes it an insult. And in (3c), we face a combination of a propositional content (*that Dr Smith has just called*) and the expressive element *that asshole*.

Insults can refer to past and future events, so there is no need to specify this in the propositional content role. All we need is the postulate of something that is functional for bringing about the feeling insulted.

- (4) Propositional content rule
 What is to be expressed is any proposition *p* or expressive meaning *e* functional for insulting.

Let us turn to the preparatory rules (preparatory conditions) that capture special aspects of contextual fit.

The insulters may have particular reasons for their insulting, and occasionally even good reasons.⁸ However, there is also known that insulting may happen just for fun. Hate speech may happen without any racist or sexist motive of the hate speaker, just because of stirring up the feelings of others. You are really upset when you are insulted in a general way, out of the blue, without the slightest clue why you are the target.

Speech acts should not express what is obvious. For instance, when what is promised would happen independently of the promise, it is no use to give a promise at all. The rule of non-obviousness captures this fact. However, in insulting, non-obviousness plays no role. It may be obvious that the target is already insulted, but nevertheless, the insulters may be going on with their insults. Conversely, the target, being used to being insulted, may acquire a habit of bouncing off the insults. Nevertheless, he or she is insulted in fact.

Thus, I propose the following preparatory rules:

- (5) Preparatory rules
 First preparatory rule: S need not have a particular motive for insulting H.
 Second preparatory rule: It is neither obvious to S nor to H that H will feel insulted by *p/e*.

The sincerity rule concerns the psychological attitude of the speakers towards their speech act. Insulting, as we posited above, is connected to the intention of the speaker to evoke a feeling of being insulted in the addressee. The speaker wants to achieve this effect by uttering his or

⁸ Think of hate speech against the bourgeoisie, apparatchiks, oligarchs, dictators, etc.

her insult. In a way, he is committed to this aim (cf. the concept of expressive commitment in Hess 2020).

Therefore, I assume the following sincerity rule for insults:

- (6) Sincerity Rule
S wants H to feel insulted.

An obvious question is whether we need special rules/conditions for those speech acts that qualify as hate speech. In principle, it would be possible to put such qualifications in additional preparatory rules. For instance, one could draw a distinction between insults like *Idiot!* (sometimes called “slander”) and slurring, slurring in turn being subcategorized into non-ethnic slurring (*Fatso!*) and ethnic slurring (*Spic!*) (cf. Meibauer 2016). However, while it is viable to analyze insulting words as indicators of illocutionary force, I would not go so far as assuming a separate illocutionary force according to the specific indicator.

It is known that insult expressions have different degrees of strength. For example, *kraut* is not as strong as *Nazi pig*. The strength of hate expressions in a community of speakers can be determined empirically, through questioning. The perceived strength can change, for example through increased appropriation or taboo. Searle (1979a: 5) mentions as a classification dimension “differences in the force or strength with which the illocutionary force is presented” and gives as an example *suggest* vs *insist*. Of course, one cannot exclude the possibility that a lexicalization of illocutionary verbs denoting insults with different strengths has taken place in a language (cf. for example, *anmachen* vs. *heruntermachen*). Here I want to argue that the use of indicators of different strength does not necessarily have to filter through to a corresponding classification.

Another classification criterion in Searle (1979a: 5) are “differences in the status or position of the speaker and hearer as these bear on the illocutionary force of the utterance”, for instance, order/command (by a general) vs. suggestion/proposal (by a private). Could it be that the authority criterion also plays a role in hate-speech acts? Indeed, it makes a difference whether hate speech is exercised by a powerful institution or directed against the powerful from a powerless position. However, for one to speak of “subordinate” acts, it is sufficient that a speaker ascribes authority to himself or herself in a certain situation or that this authority is recognized (cf. Maitra 2012). Moreover, there is also hate speech in situations where neither party can claim authority. Therefore, in my opinion, “authority” is not a suitable classification criterion for hate speech.

I have assumed that the actual illocutionary goal of the insulter is that the insulted person feels insulted (cf. Neu 2008). In his critique of Meibauer (2014), Milić criticizes it for being subject to the same objections that apply to the attitudinal view (speaker-centered view) and the subjectivist view. The attitudinal view states that the insulter expresses his attitude towards the victim. However, expressing one’s attitude is neither a necessary nor a sufficient condition for insulting. The subjective view states that an insult occurs when the addressee feels insulted. Milić (2018: 547) argues for the objective approach, which he defines as follows: “[...] a linguistic act x counts as an insult if and only if (i) x is recognised as demeaning by the standard of the relevant social group at t and (ii) x is demeaning when addressed at A.” This definition is intended to take account of the fact that subjective feeling insulted can vary greatly, even

among the group of people affected. For this purpose, a “judge” is introduced – similar to what has been proposed for the predicates of taste – who judges (the degree of) demeaning.

We have to distinguish three cases: (i) feeling insulted but not being insulted (for instance, A feels insulted because B hasn’t greeted her; however, it turns out that B didn’t wear her glasses; cf. Anderson 2018). (ii) being insulted but not feeling insulted. B doesn’t feel insulted (she is used to it, etc.) but – according to the law, the relevant social group, etc. – she is insulted in fact. (iii) Finally, there is the case of feeling insulted plus being insulted and this is the standard case to be modeled by speech-act theory. I do not see why Milić denies the attributed attitude of the speaker to express their negative emotion (hate, contempt, disgust...) towards the target. It seems weird to assume that a speech act report like (7) is viable.

- (7) # In acting insultingly, the speaker has no intention to express a demeaning attitude towards the relevant social group and A.

The intentionality of hate speakers is important. Targets regularly assume that there is a speaker intention behind slurs. With respect to hearer uptake, it is important to assume that hearers cannot autonomously decide how a speech act is to be understood (cf. McDonald 2021). The hate speaker relies on illocutionary indicators that conventionally and directly signal the insult. Indeed, this seems the reason for their existence. The target – in line with possible “judges” – understands this speaker intention, and except the case (ii) will feel insulted.

4 Self-restraint, commitment, deniability

When the former Ukrainian Ambassador in Berlin, Andrij Melnyk, called the former German Chancellor Olaf Scholz *beleidigte Leberwurst* (lit. ‘insulted liverwurst’; ‘prima donna’), then this was potentially insulting. We do not know whether Scholz felt insulted but we know that *beleidigte Leberwurst* is a pejorative predicate. Accordingly, several politicians complained about the Ambassador’s behavior and called on him to apologize for this choice of words. The question is, then, why Melnyk did not engage in self-restraint.

With respect to hate speech (as a case of insulting), we may ask under what conditions potential hate speakers are willing to engage in self-restraint. In informal terms, Thompson (2019: 660) points out, it would be possible that “they hold back, remain silent, bite their tongue, swallow their words”. Moreover, they could use weaker terms, or they could use more indirect strategies. Not much is known about how potential hate speakers reason about these possibilities.

According to my proposal, the second preparatory condition for insulting was “It is neither obvious to S nor to H that H will feel insulted by p/e.” However, one can demand of the speakers that they calculate possible effects of their offensive utterances on their addressees. Thompson (2019) draws a distinction between intentionality and foreseeability of hate speech. On the basis of these two variables, he presents four categories. Of these four categories, the categories of (i) “intentional but futile” (+intentional, –foreseeable) and (ii) “not intentional but reasonably foreseeable” (–intentional, +foreseeable) are the most interesting.⁹ Category (i) relates to cases in which the speaker intended but “should have foreseen that their act would *not* stir up hatred”

⁹ The remaining categories are of the type [+intentional, +foreseeable] – this is the standard case of hate speech/insulting – and [–intentional, –foreseeable].

against its object (Thompson 2019: 664). Category (ii) relates to cases in which the speaker did not intend but “could reasonably foresee that their act would stir up hatred against its object” (Thompson 2019: 663). In both cases, the speaker should engage in self-restraint. With respect to (i), from the point of view of speaker intention, the hate speaker should have chosen more offensive or aggressive slurs. If, however, the hate speaker learns that their enterprise runs into the nothing, then there is a chance that they will renounce their efforts. With respect to (ii), it would be very welcome if the hate speakers reflect about unintended effects of their speech, with the result of refraining from risky statements in future.

The sincerity rule for insulting requires that “S wants H to feel insulted”. Thus, strictly speaking, if such an intention is lacking, we cannot speak of an act of insulting, though the addressee may feel insulted in fact (cf. category (ii) in Thompson’s taxonomy). If insulting is intended, the sincerity rule will apply, and the speaker, so I will argue, is committed to that intention. Saying “You are a dirty X, though I do not want that you will feel insulted by my utterance” is impossible. The assumption that every speech act comes with a specific commitment, has been elaborated in a number of approaches, most notably by Geurts (2019) (cf. Boulat 2023). Also, Hess (2020) argues that in expressives, the speaker is committed to the expressive content of his or her speech act. The linguistic task of the illocutionary indicator of insulting is to signal unanimously that the speakers are committed to their malicious intent.¹⁰ Even if hate speech is only spread, for example by liking or retweeting, it can be argued that the actor in question is guilty (cf. Marsili 2021; Reinbacher 2020).

Several reactions of an addressee are possible: Thus, the target can be violated (because they feel insulted), they can bounce off the hate speech, and they can engage in counter hate speech. Often, a complaint is filed, or an apology is requested. One can also be sentenced to an official apology. If the offender does not apologise, he will follow various denial strategies.

Though the speakers are committed to the expression of their negative emotions and cannot deny them, there are a number of repair strategies available. Boogaart/Jansen/van Leeuwen (2021) distinguish between four kinds of denying literal meaning: The speakers can deny having uttered certain words altogether; they can provide another “literal” meaning; they can appeal to a non-literal interpretation (conversational implicature), and finally, they can come up with excuses. Moreover, the insulters may retract their insults (cf. Caponetto 2024). In the ideal case, hate speakers will even unlearn their hate vocabulary.

5 Conclusion

In order to conceptualize the phenomenon of hate speech, an attempt has been made to situate it firmly in classical speech-act analysis. Speech-act theory has the aim to define speech-act types, put them into appropriate classes, and explain speech-act assignment via the force of illocutionary indicators. Hate speech has been conceived of as a type of insulting, in which expressions play a role that are functionalized for the demeaning and denigrating of groups or individuals belonging to those groups. Insulting, in general, aims at triggering the feeling of being insulted in the addressee. Much energy has been put in the analysis of the pejorative

¹⁰ Appropriated or perspectivized uses (cf. Camp 2018) are, of course, an exception.

meaning of slurs. It is important, though, to not neglect the whole architecture of pejoration that is part of most languages. This architecture allows the language users to be aggressive and offensive against individuals and groups. Evidently, not all aggressive or insulting utterances are prohibited, and freedom of speech is an important value in a liberal society. Yet, the society must be careful not to allow the permanent and criminal exchange of hate messages. A still underrated topic in the academic discourse is self-restraint. This is indeed a possibility of human behavior. Even if someone feels that he or she hates X, and has the freedom to insult, it is possible to engage in self-restraint, that is, communicate in a way closely observing norms of the law, ethics, and morality.

References

- Anderson, Luvell (2018): “Philosophical investigations of the taboo of insult”. In: Allan, Keith (ed.): *The Oxford Handbook of Taboo Words and Language*. Oxford, Oxford University Press: 233–246.
- Averintseva-Klisch, Maria (2016): “Demonstrative pejoratives”. In: Finkbeiner Rita/Meibauer, Jörg/Wiese, Heike (eds.): *Pejoration*. Amsterdam, Benjamins: 119–144.
- Bach, Kent (2018): “Loaded Words: On the Semantics and Pragmatics of Slurs”. In: Sosa David (ed.): *Bad words*. Oxford, Oxford University Press: 60–76.
- Bonacchi, Silvia (ed.) (2017): *Verbale Aggression. Multidisziplinäre Zugänge zur verletzenden Macht der Sprache*. Berlin/Boston: De Gruyter Mouton.
- Boogaart, Ronny/Jansen, Henrike/van Leeuwen, Maarten (2021): “‘Those are Your Words, Not Mine!’ Defence Strategies for Denying Speaker Commitment”. *Argumentation* 35: 209–235.
- Boulat, Kira (2023): *The Pragmatics of Commitment*. Frankfurt a. M.: Lang.
- Brown, Alexander (2017a): “What is hate speech? Part I: The Myth of Hate”. *Law and Philosophy* 36/4: 419–468.
- Brown, Alexander (2017b): “What is hate speech? Part II: Family resemblances”. *Law and Philosophy* 36/5: 561–613.
- Camp, Elizabeth (2018): “A Dual Act Analysis of Slurs”. In: Sosa, David (ed.): *Bad Words. Philosophical Perspectives on Slurs*. Oxford, Oxford University Press: 29–59.
- Caponetto, Laura (2024): “‘Actually, Scratch That!’: A Tour into the Illocutionary Fabric of Retraction”. In: Zeman, Dan/Hîncu, Mihai (eds.): *Retraction Matters*. Cham, Springer: 119–138.
- Clooney, Amal/Webb, Philippa (2017): “The Right to Insult in International Law”. *Columbia Human Rights Law Review* 48/2: 1–55.
- Croom, Adam J. (2013): “How to do things with slurs: Studies in the way of derogatory words. Slurs”. *Language & Communication* 33: 177–204.
- Culpeper, Jonathan (2021): “Impoliteness and Hate Speech: Compare and Contrast”. *Journal of Pragmatics* 179: 4–11.
- d’Avis, Franz (2016): “Pejoration, normalcy conceptions and generic sentences”. In: Finkbeiner Rita/Meibauer, Jörg/Wiese, Heike (eds.): *Pejoration*. Amsterdam, Benjamins: 103–118.
- d’Avis, Franz/Meibauer, Jörg (2013): “*Du Idiot/Din Idiot!* Pseudo-vocative constructions and insults in German (and Swedish)”. In: Noel, Patrizia/Aziz, Hanna/Sonnenhauser, Barbara (eds.): *Vocative! Addressing between System and Performance*. Berlin/NewYork, De Gruyter Mouton: 113–141.

- Finkbeiner, Rita (2019): “Expressive Sprechakte *revisited*”. In: Meier, Simon et al. (eds.): *50 Jahre Speech Acts. Bilanz und Perspektiven*. Tübingen, Narr: 129–151.
- Finkbeiner, Rita/Meibauer, Jörg/Wiese, Heike (2016a): “What is pejoration, and how can it be expressed in language?” In: Finkbeiner Rita/Meibauer, Jörg/Wiese, Heike (eds.): *Pejoration*. Amsterdam, Benjamins: 1–18.
- Finkbeiner, Rita/Meibauer, Jörg/Wiese, Heike (eds) (2016b): *Pejoration*. Amsterdam: Benjamins.
- Fischer, Agneta/Giner-Sorolla, Roger (2016): “Contempt: Derogating Others while Keeping Calm”. *Emotion Review* 8/4: 346–357.
- Fischer, Agneta et al. (2018): “Why We Hate”. *Emotion Review* 10/4: 309–320.
- Frank, Annika (2023): *Die Beleidigung: Diskurse um Ehre, Respekt und Integrität im Kontinuum zwischen Alltag und Recht*. Berlin: Schmidt.
- Fuchs-Kreiß, Julia/Oğlacioğlu, Mustafa Temmuz (eds.) (2026): *Hassrede an der Schnittstelle von Linguistik und Justiz*. Berlin/Boston: De Gruyter Brill.
- Geurts, Bart (2019): “Communication as commitment sharing: speech acts, implicatures, common ground”. *Theoretical Linguistics* 45/1–2: 1–30.
- Graumann, Carl Friedrich/Wintermantel, Margret (2007): „Diskriminierende Sprechakte. Ein funktionaler Ansatz“. In: Hartmann, Steffen K./Krämer, Sybille/Kuch, Hannes (eds.): *Verletzende Worte. Die Grammatik sprachlicher Missachtung*. Bielefeld, Mentis: 147–178.
- Guillén-Nieto, Victoria (2023): *Hate Speech. Linguistic Perspectives*. Berlin/Boston: De Gruyter.
- Henderson, Robert/McCready, Elin (2023): *Signalising without Saying: The Semantics and Pragmatics of Dogwhistles*. Oxford: Oxford University Press.
- Hess, Leopold (2020): “Slurs and Expressive Commitments”. *Acta Analytica* 36: 263–290. doi.org/10.1007/s12136-020-00445-x.
- Hess, Leopold (2021): “Slurs: Semantic and pragmatic theories of meaning”. In: Stolmaszczyk, Piotr (ed.): *The Cambridge handbook of the philosophy of language*. Cambridge, Cambridge University Press: 450–466.
- Hilgendorf, Eric (2008): „Beleidigung – Grundlagen, interdisziplinäre Bezüge und neue Herausforderungen“. *Erwägen – Wissen – Ethik* 19/4: 403–412.
- Jaki, Sylvia/De Smedt, Tom (2019): “Right-wing German hate speech on Twitter: Analysis and automatic detection”. arXiv:1910.07518 [11.06.2025].
- Jaki, Sylvia/Geiger, Stefan (eds.) (2023): *Digitale Hate Speech. Interdisziplinäre Perspektiven auf Erkennung, Beschreibung und Regulation*. Stuttgart: Metzler.
- Jeshion, Robin (2018): “Slurs, dehumanization, and the Expression of Contempt”. In: Sosa, David (ed.): *Bad Words. Philosophical Perspectives on Slurs*. Oxford, Oxford University Press: 77–108.
- Katzir, Maayan/Hoffmann, Matan/Liberman, Nira (2019): “Disgust as an essentialist emotion that signals nonviolent outgrouping with potentially low social costs”. *Emotion* 19/5: 841–862.
- Khoo, Justin (2017): “Code Words in Political Discourse”. *Philosophical Topics* 45/2: 33–64.
- Knoblock, Natalia (ed.) (2022): *The Grammar of Hate. Morphosyntactic Features of Hateful, Aggressive, and Dehumanizing Discourse*. Cambridge: Cambridge University Press.

- Lepore, Ernie/Stone, Matthew (2018): “Pejorative tone”. In: Sosa, David (ed.): *Bad Words. Philosophical Perspectives on Slurs*. Oxford, Oxford University Press: 132–154.
- Leslie, Sarah-J. (2017): “The Original Sin of Cognition: Fear, Prejudice, and Generalization”. *Journal of Philosophy* 114/8: 393–421.
- Liu, Chang (2020): “Slurs as Illocutionary Force Indicators”. *Philosophia* 49/3: 1051–1065.
- Maitra, Ishani (2012): “Subordinating Speech”. In: Maitra, Ishani/McGowan, Mary K. (eds.): *Speech and Harm. Controversies over Free Speech*. Oxford, Oxford University Press: 94–120.
- Marsili, Neri (2021): “Retweeting: its linguistic and epistemic value”. *Synthese* 198: 10457–10483.
- Matsumoto, David/Hwang, Hyisung C./Frank, Mark G. (2017): “Emotion and Aggressive Intergroup Cognitions: The ANCODI Hypothesis”. *Aggressive Behavior* 43/1: 93–107.
- McDonald, Lucy (2021): “Your word against mine: the power of uptake”. *Synthese* 199: 3505–3526.
- McGowan, Mary Kate (2009): “Oppressive Speech”. *Australasian Journal of Philosophy* 87/3: 389–407.
- McGowan, Mary Kate (2012): “On ‘Whites Only’ Signs and Racist Hate Speech: Verbal Acts of Racial Discrimination”. In: Maitra, Ishani/McGowan, Mary Kate (eds.): *Speech and Harm. Controversies over Free Speech*. Oxford, Oxford University Press: 121–148.
- Meibauer, Jörg (2013): “Expressive Compounds in German”. *Word Structure* 6/1: 21–42.
- Meibauer, Jörg (2014): “Bald-faced lies as acts of verbal aggression”. *Journal of Language Aggression and Conflict* 2/1: 127–150.
- Meibauer, Jörg (2016): “Slurring as insulting”. In: Finkbeiner, Rita/Meibauer, Jörg/Wiese, Heike (eds.): *Pejoration*. Amsterdam/Philadelphia, Benjamins: 145–165.
- Meibauer, Jörg (2019): “What is an indirect speech act? Reconsidering the literal force hypothesis”. *Pragmatics & Cognition* 26/1: 61–84.
- Meibauer, Jörg (2022): *Sprache und Hassrede*. Heidelberg: Winter.
- Meibauer, Jörg (2026): „Linguistische Sicht auf Hassrede“. In: Fuchs-Kreiß, Julia/Oğlakcioğlu, Mustafa Temmuz (eds.): *Hassrede an der Schnittstelle von Linguistik und Justiz*. Berlin/Boston, De Gruyter Brill: 7–24.
- Milić, Ivan (2018): “What Counts as an Insult?” *Acta Analytica* 33: 539–552.
- Neu, Jerome (2008): *Sticks and Stones. The Philosophy of Insults*. Oxford: Oxford University Press.
- Nunberg, Geoffrey (2018): “The social life of slurs”. In: Fogal, Daniel/Harris, Daniel/Moss, Matt (eds.): *New Work on Speech Acts*. Oxford, Oxford University Press: 237–294.
- Oğlakcioğlu, Mustafa Temmuz (2023): *Strafbare Sprechakte*. Tübingen: Mohr Siebeck.
- Orlando, Eleonora/Saab, Andrés (2020): “Slurs, stereotypes, and insults”. *Acta Analytica* 35: 599–621.
- Potts, Christopher (2007): “The expressive dimension”. *Theoretical Linguistics* 33/2: 165–197.
- Reinbacher, Tobias (2020): „Die ‚Weiterverbreitung‘ von Hate Speech in sozialen Medien – Fragen der Beteiligung an einer gemäß §185 StGB strafbaren Beleidigung“. *JuristenZeitung (JZ)* 75/11: 558–563.

- Saul, Jennifer M. (2018): “Dogwhistles, Political Manipulation, and Philosophy of Language”. In: Fogal, Daniel/Harris, Daniel/Moss, Matt (eds.): *New Work on Speech Acts*. Oxford, Oxford University Press: 360–383.
- Saul, Jennifer M. (2024): *Dogwhistles and Figleaves. How Manipulative Language Spreads Racism and Falsehoods*. Oxford, Oxford University Press.
- Searle, John R. (1969): *Speech Acts. An Essay in the Philosophy of Language*. Cambridge, Cambridge University Press.
- Searle, John R. (1979a): “A taxonomy of illocutionary acts”. In: Searle, John (ed.): *Expression and Meaning. Studies in the Theory of Speech Acts*. Cambridge, Cambridge University Press: 1–29.
- Searle, John R. (1979b): “Indirect Speech Acts”. In: Searle, John (ed.): *Expression and Meaning. Studies in the Theory of Speech Acts*. Cambridge, Cambridge University Press: 30–57.
- Sendlmeier, Walter/Steffen, Ines/Bartels, Astrid (2016): “Pejorative Prosody”. In: Finkbeiner Rita/Meibauer, Jörg/Wiese, Heike (eds.): *Pejoration*. Amsterdam, Benjamins: 21–40.
- Sosa, David (ed.) (2018): *Bad Words. Philosophical Perspectives on Slurs*. Oxford, Oxford University Press.
- Sternberg, Robert J./Sternberg Karin (2008): *The Nature of Hate*. Cambridge, Cambridge University Press.
- StGB: Fischer, Thomas/Anstötz, Stephan/Lutz, Hans-Joachim (2025): *Strafgesetzbuch mit Nebengesetzen*. 72. Aufl. München: Beck.
- Technau, Björn (2018): *Beleidigungswörter. Die Semantik und Pragmatik pejorativer Personenbezeichnungen*. Berlin/Boston: De Gruyter Mouton.
- Tenchini, Maria Paola/Frigerio, Aldo (2016): “A multi-act perspective on slurs”. In: Finkbeiner Rita/Meibauer, Jörg/Wiese, Heike (eds.): *Pejoration*. Amsterdam, Benjamins: 167–185.
- Thompson, Simon (2019): “Hate Speech and Self-Restraint”. *Ethical Theory and Moral Practice* 22: 657–671.
- Vanderveken, Daniel (1990): *Meaning and Speech Acts. Vol. 1: Principles of Language Use*. Cambridge, Cambridge University Press.
- Vendrell Ferran, Ingrid (2024): “Hate: toward a Four-Types Model”. *Review of Philosophy and Psychology* 15: 441–459.