



UNIVERSITY OF HELSINKI

# In-group Hate Speech: A Speech Act Theoretical Investigation of Transphobic Hate on 4chan's Imageboard /lgbt/

Master's Program in English Studies

Master's thesis

Author:

Miro Kotajärvi

Supervisor:

Joseph Flanagan

7.4.2024

Helsinki

**Faculty:** Faculty of Arts

**Degree programme:** Master's Program in English Studies

**Study track:**

**Author:** Miro Kotajärvi

**Title:** In-group Hate Speech: A Speech Act Theoretical Investigation of Transphobic Hate on 4chan's Imageboard /lgbt/

**Level:** Master's Thesis

**Month and year:** May 2024

**Number of pages:** 45

**Keywords:** Hate speech, transphobia, 4chan, corpus linguistics, discourse analysis

**Supervisor:** Joseph Flanagan

**Where deposited:** Helsinki University Library

**Additional information:**

**Abstract:**

This study combines corpus linguistic methods and discourse analysis to examine hate speech produced by transgender people targeting other transgender people. Open access online discussion forums have enabled communities to form around increasingly fringe and, at times, violent interests, thus making online hate speech difficult to define and detect. The aim of this study is to highlight these fringe forms of hate speech and how online genres influence the production of such language.

The corpus data for this study consists of posts from the website 4chan and its imageboard /lgbt/. 4chan has been investigated extensively by other researchers of hate speech, because of its anonymous userbase and its associations with the alt-right, online political extremism and anti-LGBT attitudes. Thus, the imageboard /lgbt/ whose membership consists mostly of transgender people proves a productive avenue of inquiry into in-group directed hate speech.

To establish a statistically informed hierarchy of the transgender lexicon of /lgbt/, a list of terminology referring to transgender people was parsed out from the corpus data. The frequencies of these lemmas were compared and contrasted, after which the lemmas were subjected to more intimate, qualitative analysis to determine whether the language on /lgbt/ sufficiently resembles hate speech.

The results suggest that, not only is violent language targeting transgender people frequent on /lgbt/, but that it is increasing over time. Furthermore, the corpus data reveals that, instead of general and neutral terminology, hostile genre-specific jargon targeting transgender people is increasingly preferred within the discourse community of /lgbt/, thus suggesting that in-group directed hate speech is a key communicative goal of /lgbt/.

# Table of Contents

<b>1. INTRODUCTION</b>	<b>1</b>
<b>2. THEORETICAL BACKGROUND AND PREVIOUS RESEARCH</b>	<b>4</b>
<b>2.1. Speech act theory</b>	<b>5</b>
2.1.1. Locutionary force	6
2.1.2. Illocutionary force	6
2.1.3. Perlocutionary force	7
2.1.4. A speech act theoretical framework of hate speech	7
<b>2.2. The perceptual approach and critical perspectives</b>	<b>8</b>
2.2.1. Covert hate speech	9
2.2.2. Unsuccessful hate speech	10
2.2.3. In-group hate speech	12
<b>2.3. Genre and discourse communities</b>	<b>14</b>
<b>3. MATERIALS AND METHODS</b>	<b>15</b>
<b>3.1. Materials and 4chan</b>	<b>16</b>
3.1.1. /lgbt/	16
3.1.2. Overview of corpus data	17
<b>3.2. Methodology</b>	<b>18</b>
<b>4. RESULTS AND ANALYSIS</b>	<b>19</b>
<b>4.1. Transgender lexicon of /lgbt/</b>	<b>20</b>
<b>4.2. Genre-specific lexicon of /lgbt/</b>	<b>22</b>
4.2.1 Semantic variation of genre-specific lexicon	23
4.2.2. “AGP”	23
4.2.3. “Hon”	28
4.2.4. Genre as a mediator of hate	32
<b>5. DISCUSSION</b>	<b>35</b>
<b>6. CONCLUSION</b>	<b>38</b>
<b>REFERENCES</b>	<b>41</b>

# 1. Introduction

There are few phenomena that evoke as many distinct perspectives and conceptual paradigms as the concept of hate speech. Some perspectives understand the phenomenon as verbal attacks against a person's or persons' protected characteristics, such as sex, gender, sexual identity, ethnicity or religion (see for example, Burnap & Williams, 2016; Baider, Assimakopoulos & Millar, 2017). Other paradigms take a more general approach by defining the concept as an extension of impoliteness or incitement to hatred and violence (see for example Culpeper, 2021). Accompanying these aforementioned conceptual paradigms, is often an assumption of malicious intent on the speaker's part, and some perceived harm on the interlocutor's part (Culpeper, 2021: 6). At its broadest sense however, some researchers have even argued that the concept of hate speech does not require hostile intention or direct harm to be caused, because speech, whether accompanied by malicious intentions or not, "can alter what is considered permissible or normal" (Culpeper, 2021: 6; Langton, 2018a) and because such alterations might not directly harm any specifiable person. That is to say, speech that reinvigorates tropes of inequality, oppression or violence may be hate speech, even without the intention to be so, or without causing direct harm "in a way that is analogous to pollution" (Simpson, 2013: 706).

Secondly, as a legal matter, the concept has been approached in diverse ways by different countries, regions and legal scholars. For example, while European countries have tended to be more affirmative towards hate speech laws, the United States has been considerably more hesitant to enact similar regulation (Sellars, 2021: 5). When the American legal scholar Marie Matsuda first used the term "hate speech" in her seminal article *Public Response to Racist Speech: Considering the Victim's Story* (Matsuda, 1989), her goal was to centralise the ways in which the legal system of the United States had failed to provide adequate "means of seeking redress, civil or criminal" (Brown, 2017a: 424). By highlighting instances of racist language that were "not easily actionable under current laws", Matsuda's aim was to "show the limitations of the legal concept [of] hate speech in the United States, as a form of protected speech" (Brown, 2017a: 424). Continuing in Matsuda's footsteps, in this study, my interest is not to grasp *the essence* of hate speech, nor to review the totality of the research done on the phenomenon, but to tease out the fuzzy borders of the concept of hate speech, in order to bring to light instances of language that may yet be understood as hate

speech, yet which under the contemporary research paradigm, might be declared something else.

In the Web 2.0 era of user-generated content online, controlling the public production of hate speech has become increasingly difficult (George & Scerri, 2007). Additionally, in concordance with this development, academic interest on the phenomenon has also grown, and new terminology to describe the complex dynamics of different forms of hate speech, such as overt or explicit and covert or implicit hate speech, have been established. In this study, my aim is to add further nuance to the discussion around hate speech and its diverse forms, by providing evidence for a phenomenon I shall call *in-group hate speech*. *In-group hate speech*, as I shall outline it in section 2.2.3 below, is hate speech produced by and against members of the targeted community. Therefore, the implicit assumption by many researchers that hate speech is either unwillingly received hostility or “favo[u]ritism<sup>1</sup> toward the in[-]group and discrimination against the out[-]group in order to maintain a positive social identity” (Leets, 2002: 342) is reversed. By utilising both quantitative methods and pragmatic analysis, I will argue that the popular understanding of hate speech as speech that harms its target is *not* a necessary condition of hate speech and that there exist forms, or subcategories, of hate speech where harm is either not directly caused or is otherwise anticipated by its target audience. Concordant with this aim, the research questions (hereafter referred to as RQs) for the study are as follows:

RQ1: What kind of general and genre-specific terminology is used to refer to transgender people on /lgbt/?

RQ2: How frequently is this terminology used and how is this terminology modified?

RQ3: What do the frequencies and modifiers of this terminology reveal about the attitudes of /lgbt/ towards transgender people?

RQ4: Can the hostile language of /lgbt/ be understood as hate speech?

To do this, I will examine the imageboard, /lgbt/ on the website 4chan. 4chan is a site already well recognised by previous research to be “linked to the so-called alt-right movement” and as “the birthplace of conservative or even outright hateful and racist memes” (Rieger et al., 2021: 1). However, while 4chan has been studied extensively with numerous research interests, the subforum /lgbt/ has yet to receive

---

<sup>1</sup> Text in brackets in this and all subsequent examples denote my additions or corrections.

similar attention by hate speech researchers. As a subforum whose membership consists mostly of transgender people, on a forum determined to be hostile towards the LGBT community (Rieger et al., 2021: 7), /lgbt/, as a subforum of 4chan, represents an interesting intersection between a genre built on hostility towards transgender people and transgender people with a resentment towards other transgender people. Furthermore, while researchers in the social sciences have established terminology to describe the positive pattern of language reclamation (see for example Jeshion, 2020), little research has been done on the negative language production of minority groups to refer to themselves and other members of their community. Additionally, previous studies on 4chan have mostly utilised automated approaches that have focused “on explicit “hate terms,” thus being unable to account for more subtle or covert forms of hate” (Rieger et al., 2021: 2). So, to address these research gaps, a synthesis of a quantitative corpus linguistic approach and qualitative discourse analysis should be able to shed light on the contours of in-group directed hate speech. By familiarising oneself with the genre of 4chan and /lgbt/, a genre-specific lexicon of transgender terminology can be produced. This lexicon can then be used, in collaboration with a corpus linguistic approach, to establish a quantitatively convincing hierarchy of language, used by the discourse community of /lgbt/, to refer to other transgender people. From this data, more intimate analysis of the types of exchanges and semiosis between speakers and interlocutors is made possible. This is key to the operative theory of hate speech enforced in this text, which is one based on Austin’s (1975) speech act theory and Brown’s (2017a; 2017b) benchmark inquiry into the popular concepts of hate speech. I will argue, in accordance with Brown, that hate speech is:

a family of types of speech act including but not limited to insulting, disparaging, degrading, humiliating, misrecognising, disheartening, harassing, persecuting, threatening, provoking, inciting hatred, discrimination or violence, and justifying or glorifying discrimination or violence (Brown, 2017b: 600)

That is to say, no necessary condition, such as the requirement that harm must be caused for hate speech to be hate speech, accurately portrays what we truly mean by the notion. There are instances of hostile language, that do not cause harm that may yet be understood as hate speech and, in this study, it is my aim to provide evidence for such speech acts. Furthermore, with the findings of this study, it is my aim to

illustrate the necessity for more rigorous, explicit exploration of the concept of hate speech by researchers working in the field.

In section 2 below, I will outline some of the definitions of hate speech given by previous research and explore their limitations in capturing the totality of hate speech. I will adopt a speech act theoretical perspective to establish a framework of numerous hate speech dynamics that can all be conceived as hate speech, yet which differ greatly in their locutionary, illocutionary and perlocutionary appearance. In section 2, I will also highlight the importance of genre in approaching novel forms of hate speech like *in-group hate speech*. Subsequently, in section 3, I will explore the genre of online messaging forums, 4chan and /lgbt/, as an integral element of the hate speech presented in this study. I will also lay out the methodological design of this study in greater detail, by highlighting the benefits of synthesising quantitative corpus linguistic approaches with discourse analysis. In section 4, results and analysis, I will first go over the transgender lexicon of /lgbt/ to establish a hierarchy of lemmas used to refer to transgender people. Thereafter, I will explore the modifiers of key genre-specific lemmas in-context to answer RQs 1 and 2. To answer RQ3 and RQ4, I will further explore the semantic categories, preferences and backgrounds of the genre-specific transgender lexicon of /lgbt/ and the responses this hostile language has received. In section 5, I return to the theoretical underpinnings of this paper and review how previous research reflects the findings of this study. Finally, in section 6, I conclude by revisiting the findings of this study, their limitations and avenues for future research.

## 2. Theoretical background and previous research

For many, the term “hate speech” seems intuitively easy to grasp and instances of hostile speech can easily be categorised as hate speech due to this intuition. A closer examination of the concept of hate speech reveals a remarkably more nuanced discussion, however. In legal literature on hate speech, there are at least three distinct definitions allotted to hate speech:

we can define hate speech in terms of the harms it will cause—physical contingent harms like violence or discrimination; or we can define hate speech in terms of its intrinsic properties—the kinds of words it uses; or we can define hate speech in terms of its connection to principles of dignity (Herz & Molnar, 2012: 31)

The three types of hate speech definitions as noted by Herz and Molnar are therefore: (i) harm, (ii) linguistic content and (iii) dignity. Each of these definitions have their share of conceptual upshots and downsides, and some legal frameworks could certainly work towards “a hybrid definition by combining the ways mentioned” (Anderson & Barnes, 2022: para. 1). However, as this paper is not concerned with defining *the essence* or universal quality of hate speech, and rather teasing out the fuzzy borders of language that might yet be understood as hate speech, in this section, I will particularly focus on the first category: (i) harm, and its ramifications to the concept of hate speech. I will leave it to others to explore the other conceptualisations in greater detail.

In this section, I will contest the requirement of harm to the concept of hate speech, by outlining instances of language that ostensibly cause no harm, yet nevertheless seem to intend it. Furthermore, by outlining a theoretical basis for a category of intragroup hate speech, I reject the implicit assumption of the colloquial concept of hate speech that limits the phenomenon to a dynamic between distinct, disparate social groups or as tribalism that “walls off one group from another” (Langton, 2018b: 144), and merely understand it as hostile speech acts between speakers and interlocutors. In the following sections, I will flesh out this view and build a speech act theoretical framework of key hate speech dynamics while arguing for a novel form of hate speech I shall call *in-group hate speech*. To accomplish this and shed light to how such hate speech operates, I will also need to explore genre as a mediator of hate speech which I will tackle in the final section of this chapter.

## 2.1. Speech act theory

Speech act theory, ushered into popular knowledge by J. L. Austin in his posthumous book *How to do things with words* (Austin, 1975), understands utterances as sets of speech with three acts: locutionary, illocutionary and perlocutionary acts. Austin’s theory is particularly important to the study of hate speech, since it verbalises many of the dynamics at play with the phenomenon. For example, Herz and Molner’s (2012) legal exploration of hate speech highlights at least two categories of hate speech (i) harm and (ii) linguistic content which are in direct connection with Austin’s notions of perlocutionary and locutionary force of language. Let me now explore these notions in greater detail.

### 2.1.1. Locutionary force

By locutionary acts, I understand Austin (1975: 95) to mean the material product of language at face value. The locutionary act is “the performance of an act *of* saying something” (Austin, 1975: 99). Austin further distinguished phonetic, phatic and rhetic acts from one another, whereby the phonetic act “is merely the act of uttering certain noises”, the phatic act “is the uttering of certain vocables or words, [...] belonging to and as belonging to, a certain vocabulary”, and the rhetic act “is the performance of an act of using those vocables with a certain more-or-less definite sense and reference” (Austin, 1975: 95). For the purposes of this study, my analysis and the examples examined in that analysis will mostly centre sensible locutionary acts Austin would categorise as rhetic acts. Much of hate speech research done in the field of linguistics has naturally focused on the locutionary act. In particular, quantitative approaches utilising certain lists of hostile words or hate terms have been particularly popular to explore hate speech online (see for example, ElSherief et al., 2021; Teh et al., 2018).

### 2.1.2. Illocutionary force

Illocutionary acts, as I understand Austin, denote the intention behind the locution. Therefore, the illocutionary act is “the performance of an act *in* saying something” (Austin, 1975: 99). To illustrate this distinction between locutionary and illocutionary acts, the utterance: “can you close the window?”, for example, has the locutionary force of a question, yet may in fact function as a directive on the illocutionary level, as in: “close the window, please”. If your coworker were to ask this of you, and you simply answered them by stating that: “yes, I am able to close the window”, without doing anything in order to either comply with your coworker’s request, or deny said request, your coworker would have good reason to question your comprehensive faculties.

The concept of illocutionary force has been particularly influential in the field of hate speech research. The notion of “constitutive harm”, that is, “harms that occur *in* the saying of some utterance of hate speech” as opposed to the “downstream results” of hate speech (Anderson & Barnes, 2022: para. 1.1), for example, relies on this very speech act theoretical concept of illocutionary force. That is to say, harm might be caused, simply on the account that one intends to cause harm with their utterance. The utterance “only Christians in Finland”, for example, does not lexically target anyone outside of Christians, nor insult any specific group, yet it seems evident that the illocutionary force of the utterance is to other the non-Christian population of Finland.

### 2.1.3. Perlocutionary force

Finally, the perlocutionary force of an utterance can be understood as the reaction of the interlocutor, or effect of the speaker's utterance. Perlocutionary acts are the "consequential effects" (Austin, 1975: 102) on the interlocutor, by the speaker's utterance. If you were to close the window your coworker asked you to close, the perlocutionary force of his utterance would have been something akin to persuasion. Perhaps out of all three notions, Austin's concept of perlocutionary force is most directly linked with previous research on hate speech and the conceptual paradigms atop which many legal frameworks of hate speech are built upon. This is because within the perlocutionary act is captured the directly detectable harm of hate speech. That is to say, the responses and reactions of the affected community of hate speech, are within the realm of the perlocutionary force of the language. The notion of "consequential harm" (Anderson & Barnes, 2022: para. 1.1) or the harm caused by hate speech, for example, fully rests on this notion of perlocutionary force of language and is therefore integral in understanding the phenomenon.

Now, it should be mentioned that even Austin recognises that these distinctions between locutionary, illocutionary and perlocutionary acts are not always self-evident (Austin, 1975: 102–103). For example, when questions of *meaning* arise, as in, what is meant by an utterance, there is no end to the debate about whether the centre of attention should be on the locutionary force or the illocutionary force of language. As evidenced in the window example above, the locutionary act does not always act in harmony with the illocutionary act. Furthermore, there is an intricate discussion to be had here about Austin's notion of illocutionary force, because it taps into a philosophically rich, albeit contested discussion around intention and human action, but I will explore this in greater detail in subsection 4.2.3.

### 2.1.4. A speech act theoretical framework of hate speech

By applying this speech act theoretical framework, one can begin to outline the contours of what I understand to constitute hate speech and the dynamics at play with the phenomenon. At the very core of the phenomenon are utterances that on all three (locutionary, illocutionary and perlocutionary) levels, reveal a dynamic of harm or hostility. Such utterances that fulfil all three conditions could be called *explicit* or *overt hate speech* and much of previous research in the field "has focused on [...] overt hate speech" while "failing to address a more pervasive form based on coded or indirect

language” (ElSherief et al., 2021: 345). Prototypical death threats and calls to violence, for example, easily fulfil these three criteria, by lexically reflecting a violent intention on the speaker’s part, while also producing a negative response on the interlocutor’s part. Previous research shows us however, that overt hate speech (hereafter referred to as OHS) is increasingly avoided and that the “majority of hate language is communicated covertly” (Parvaresh, 2023: 75).

The borders of hate speech therefore become fuzzy when one or more of these speech act theoretical conditions begin to shift from harm and violence to something else. For example, one can understand *covert hate speech* (hereafter referred to as CHS), that is, hate speech that is somehow disguised or otherwise linguistically hidden, as hate speech where both the illocutionary and perlocutionary forces of the utterance are hostile, yet where the locutionary force of that same speech act is not. To illustrate this, consider the following example: PragerU, a popular account on X (formerly known as Twitter), replied to a comment criticising PragerU’s prior tweet stating that “Men should not wear women’s clothing” (PragerU, 2023a), with the cigarette emoji: “🚬” (PragerU, 2023b). Here, the cigarette emoji insinuates or, rather, *intends*, something that is not revealed by the locutionary force of the utterance itself, namely, that the commenter criticising PragerU is a homosexual, by appropriating the British slang word *fag*, meaning cigarette. Such coded language is particularly problematic to the study of hate speech, because it is successful in inciting hatred, causing harm and because it manifests as “acts of aggression [...] and domestic terrorism [...] while also maintaining *plausible deniability*” for these actions (ElSherief et al., 2021: 345).

## 2.2. The perceptual approach and critical perspectives

In qualitative hate speech research, a particularly popular approach taken by many contemporary researchers to address some of the complications speech act theory reveals to us about hate speech, is one that I call the *perceptual approach*. In this approach, the research centres the targeted community of hate speech in an effort to define and identify hate speech by, for instance, measuring the targeted community’s willingness to interact with hostile language or asking participants to rate the hostility of said language. In a benchmark study, Leets (2002), for example, has investigated hate speech discrimination faced by religious and sexual minorities by reviewing the target communities’ reception of numerous examples of targeted hate speech. The findings of the study conclude that the “short- and long-term consequences” of hate

speech “had striking commonality with other types of crises (e.g., rape, burglary, domestic violence, assault, robbery)” (Leets, 2002: 354) and that the targeted communities understand the primary motives of hate speech to be either ignorance or suppressed hostility. Subsequently, in speech act theoretical terms, the *perceptual approach* centralises the *perlocutionary force* of hate speech: the reactions, responses, et cetera, of the targeted community. Therefore, by logical extension, the *perceptual approach* understands hate speech as speech acts that, to some extent, cause measurable harm or discomfort to their target audience as evidenced by the perlocutionary force of the speaker’s utterance.

Although the *perceptual approach* represents a noble shift of perspective from hostile language to the affected communities of that language, I think this approach nevertheless struggles from some key insufficiencies. For one:

by focusing only on the harms *caused* by hate speech, we are inevitably drawn into a debate about balancing the costs and benefits of permitting or regulating speech, which often leads to an impasse. Alternatively, turning our attention to the acts hate speech constitutes can reveal features that help us avoid [the] question of balancing harms, and opens the door to regulation (Anderson & Barnes, 2022: para. 1.1)

That is to say, when arguing for hate speech regulation on the basis of harm done, one is always faced with the counter-argument of the harm caused by limiting free speech with hate speech regulation. Considering this, let me now explore these conceptual insufficiencies in greater detail, in order to synthesise a speech act theoretical view that could more holistically approach hate speech as a phenomenon, while also considering the targeted communities of hate.

### 2.2.1. Covert hate speech

First, in an effort to outsource the responsibility of defining hate speech to the targeted communities, the *perceptual approach* may make the assumption that the targeted communities possess a particular, holistic awareness of language use that provides them with the ability to accurately identify hate in language. As targets of hate speech, it is evident that members of specific social groups are more aware of the hostile language directed their way, than people outside of these communities who do not face similar hostility. However, as explored above, CHS often employs linguistic strategies such as dog whistling, where the hateful content is disguised or coded as to not be immediately evident (see for example Åkerlund, 2022; Bhat & Klein, 2020). These

strategies are specifically designed to obfuscate or hide in plain sight the harmful and violent language beneath and can therefore be overlooked by people not part of the discourse community in-the-know of the genre-specific lexicon. Consider for example, the much-studied use of the milk emoji that has been utilised by white supremacists and neo-Nazis like Richard Spencer to index the supremacy of white people and valorise the troubling notion of white purity (Stănescu, 2018: 104). The illocutionary force in this instance is of course immeasurably violent, yet the locutionary force of the same utterance, the milk emoji, is so inconspicuous that it would not surprise me, should even the most ardent anti-racist miss it entirely.

Additionally, there seem to exist “community-building aspects” to online hate speech, whereby “in alt-right fringe communities, hate is often “memeified” and mixed with humo[u]r and domain-specific slang, creating a situation in which the use of hate speech can play a crucial role in strengthening bonds among members of the community” (Rieger et al., 2021: 2). This, in turn, further muddies the legibility and makes opaque the hateful language for outsiders, by disguising hate with humour and genre-specific memes and fads. Taking these covert and community-building dimensions of hate speech into account, I do not think it is fair to assume that targeted communities necessarily possess the capacity to detect hate speech, even when presented with incredibly violent language.

### 2.2.2. Unsuccessful hate speech

Secondly, another issue that the *perceptual approach* struggles from, is that by centralising the perlocutionary force of hate speech, the illocutionary force of that same utterance (and by extension the locutionary act itself) can essentially be ignored. After all, if the targeted community is measurably harmed by some instance of language, what do the intentions of the speaker, or the content of the utterance matter? I do, however, think that such a unidirectional approach does pose some problems. Even as there is growing evidence that one may cause harm without intending it (see for example Gabriel, 1998: 1333) and, as Butler notes, “a speech act can go wrong” (1997: 19), I will refrain from the argument that language perceived as violent that does not intend to be harmful is *not* hate speech, to avoid more troublesome discussions about what constitutes harm. Instead, in the interest of questioning the supreme onus afforded to the perlocutionary force of language in the *perceptual approach*, I will argue for the easier, yet often overlooked case of *unsuccessful hate speech*. That is to say, one may intend harm yet be unsuccessful in this attempt when perceived by the

targeted community. For example, on the social media website Reddit, a subforum by the name of r/onejoke (Reddit, n.d.a) functions as a platform for transgender people and invested allies to satirise iterations of an antiquated transphobic comment known as *the attack helicopter joke*. As stated by the r/onejoke description, the subforum features both satirical and sincere examples of the helicopter joke (Reddit, n.d.a). In one example of a sincere iteration of the joke, a photo of a censored interaction on X shows a person questioning the meaning of a “non-binary” option in the menu screen of a game. Another X user then explains the meaning of “non-binary”, to which the original poster simply replies:

Nah bro my gender is a attack helicopter (Biker\_Hackerman, 2020).

Here, the X user’s intention is clearly to diminish what it means to be non-binary, yet the locutionary force once again does not reveal it and instead simply makes a statement about the X user’s gender. One might therefore say that this speech act is an instance of CHS, where the locutionary force of the utterance is non-hostile, while the illocutionary and perlocutionary forces are hostile. However, reading the comments under the Reddit post provides a different perspective. One Reddit user describes the tweet with the following words:

It gets funnier and funnier every time they repeat it. Every time they recycle the same joke that’s been around since, like, 2014 in 2020 with no variation. What comedy gold. (Karjo2000, 2020)

It seems that the perlocutionary force of the speaker’s speech act did not, despite the X user’s ostensible intentions, amount to hate but humour. As I see it, in such cases, the *perceptual approach* would have to determine such an utterance as non-hate speech, since it is perceived humorously by its target audience. Yet it seems clear however, that language that intends harm is, to some degree, harmful, and can therefore be understood as hate speech. After all, is it not a grave oversight to argue that a person intending to harm another, is not harmful, simply because his target does not perceive his intention as such? Considering the X user’s intention to equate what it means to exist in a gender nonconforming body to a personal delusion about identifying as an attack helicopter, I do not think that the humorous response the tweet received online erases its hostile nature. Instead, I think the aforementioned tweet is an example of another fuzzy border of hate speech, namely, *unsuccessful hate speech*, hereafter referred to as UHS. The speech act theoretical dynamic of UHS is as follows;

the illocutionary force of an utterance is hostile, but because of a linguistic quirk (such as a humorously overlaid cliché as is the case in the example above) rendering the locutionary force of the utterance non-hostile, the perlocutionary force becomes non-hostile as well.

### 2.2.3. In-group hate speech

Finally, and most acutely to this paper, I wish to shed light to a particularly problematic form of fringe online discourse, namely, *in-group hate speech*, hereafter referred to as IGHS. In IGHS, hate speech is produced and consumed by its target community and even functions as a community-building motif, reflecting the findings of Rieger et al. (2021). In section 4 below, I will explore examples of such language in greater detail, but suffice it to say, the plausibility of such circular or, to be crass, cannibalistic forms of hate speech, poses a severe challenge to the broad applicability of the *perceptual approach* and its underlying theory of what constitutes hate speech. Much of previous hate speech research has understood the phenomenon, primarily, as a social dynamic between distinct in-groups and out-groups and as “favo[u]ritism toward the in[-]group and discrimination against the out[-]group in order to maintain a positive social identity” (Leets, 2002: 342). Yet, in the case of IGHS, group identity is maintained, among other things, through hostility towards the in-group. This begs the question; if the target community is also the community responsible for producing the hostile language, how do we measure the responses of the target community? What do the reactions of the target community reveal to us about hate speech, if they are part and parcel in producing the hateful language? In section 4, I will illustrate how calls to violence and encouragements to suicide specifically because of one’s transness, a protected characteristic, for example, are not uncommon on /lgbt/, the subforum jokingly referred to as “/tttt/”, because of its community consisting mostly of transgender people (see section 3 for further discussion on /lgbt/). It would seem peculiar to claim that such language is not hate speech, simply on the basis that its target audience perceives it as acceptable.

This dynamic of othering members of the in-group reflects previous research in the field of impoliteness, particularly the notion of *group face*. A study by Culpeper (2021) has compared the research fields of hate speech and politeness theory and found great utility in appropriating terminology most often seen in politeness studies, like the notion of *face*, to the study of hate speech. Previous studies concerned with the notion of face have typically understood the notion as a duality of positive and negative

face desires, whereby positive face is concerned with “the desire to be approved and appreciated” and negative face is concerned with “the need for freedom and autonomy” (Zhu & Bresnahan, 2018: 142). However, this dualistic view of face has since “experienced a backlash from scholars studying non-Western cultures, who pointed out the importance of “group face”” (Culpeper, 2021: 5) and a related concept of *collective face*. Here, group face refers to “an individual’s tendency to conform to culturally-endorsed expected behavior[s] to be able to fit into a group”, while collective face refers to a practice of “intercultural communication [...] focusing on how to demonstrate a group’s competence to impress the out[-]group” (Zhu & Bresnahan, 2018: 143). On /lgbt/, both dynamics seem to be present as reasons for self-hate and intragroup hate speech. One may understand the notion of group face to motivate some members of the DC of /lgbt/ to conform to a form of socially accepted femininity.

Furthermore, the notion of collective face, where “individuals have a strong need to maintain, protect, and promote the collective positive image of their salient community perceived by the out[-]group” (Zhu & Bresnahan, 2018: 143), seems particularly explanatory when examining the types of hate speech explored in section 4 below. That is to say, one motivation for some members of /lgbt/ to engage with and produce hate speech, might be this desire to be viewed positively by the out-group, or the cisgender population. However, because this collective face desire to be perceived in a good light is coupled with the acute realisation that this is often not the case, an enemy or a quasi-out-group of the in-group, is established so that at least a portion of the in-group may be protected and hopefully viewed positively by the out-group. Thus, in some cases, the logic for IGHS can be understood as a positive collective face desire, where the positive image projected towards the out-group of cisgender people, is maintained by suppressing certain sections of the in-group by the means of hate speech. In speech act theoretical terms, communities that form around such collective face desires, produce hateful language or *locutions*, with the *intention* to suppress a section of the in-group, yet because this is viewed as a form of protection, it is *perceived* to be non-hostile.

For accessibility, in Table 1 below, I have situated the various subcategories of hate speech dynamics explored in previous paragraphs on the basis of locutionary, illocutionary and perlocutionary (non-)hostility. Notice that both UHS and IGHS are perceived as non-hostile by the target audience and therefore made inaccessible to the *perceptual approach*. The perlocutionary force of UHS is non-hostile, since the

locutionary force of the utterance is perceived as non-hostile, while IGHS is perceived non-hostile because locutionary and illocutionary hostility is in accordance with the genre norms of a specific discourse community which also acts as the target audience of the hate speech.

**Table 1.** A speech act theoretical framework of four hate speech dynamics

Form	Locutionary force	Illocutionary force	Perlocutionary force
OHS	Hostile	Hostile	Hostile
CHS	Non-hostile	Hostile	Hostile
UHS	Non-hostile	Hostile	Non-hostile
IGHS	Hostile	Hostile	Non-hostile

### 2.3. Genre and discourse communities

As hinted at above, the complications introduced by IGHS are best understood through examining genre. Genres, as I understand Tardy and Swales (2014), are socially based practices of obtaining shared, semiotic goals within given discourse communities (hereafter referred to as DC). Genres can be understood as the communicative events where the interplay of “form, context, rhetorical action, and function or purpose” are pivotal and which, at their most basic sense, are means through which one “[carries] out actions and purposes” (Tardy & Swales, 2014: 166) in a *community of shared goals*, or a DC. Furthermore, genres “both shape and are shaped by the communities and contexts in which they exist” (Tardy & Swales, 2014: 166), meaning that genres themselves influence the language of the DC. As explored above in sections 2.3.1, genres tend to produce genre-specific jargon and according to Liimatta’s (2019) study on Reddit, the smaller the membership of the DC and narrower the niche of that community is, the more prevalent the production of this genre-specific jargon is as well. This is because communities with wider appeal draw more diverse audiences which in turn also diversifies the language of the genre, while genres with narrower niches focus on increasingly specific topics for which the creation of newfangled language may be necessary. Additionally, it should be noted, that the emergence of user-generated content online facilitated by Web 2.0 architecture has resulted in a “lack of control over hate speech” (George & Scerri, 2007: 6). This liberation of access to worldwide discussion forums has meant that increasingly fringe interests are afforded the space to propagate and form communities around these interests. Due to the emergence of online forums such as 4chan and 8chan, previously silent and private

in-group resentments, now have public platforms, where voicing these resentments are tolerated if not encouraged.

In the broad genre of online messaging forums, websites like 4chan and 8chan have been particularly popular platforms for hate speech research and 4chan subforums such as /pol have been studied extensively from various sociological, political, and linguistic research perspectives. For example, Rieger et al. (2021) have explored the alt-right (abbreviation of alternative right wing) movement online by applying a corpus linguistic design to analyse hate speech on 8chan, Reddit and 4chan. Their findings conclude that communities on the fringes of online discourse tend to be more hateful (Rieger et al., 2021: 11). Furthermore, a common target of 4chan hate speech are members of the LGBT community (Rieger et al., 2021: 7). These findings suggest that /lgbt/, as a genre ostensibly consisting mostly of queer and transgender members, on a platform determined to be hostile towards the LGBT community, is, not only a forum on the extreme fringes of online discourse, but therefore also likely to be plentiful in the types of hateful language this study is concerned with. Considering this, it is my understanding, that hate speech can best be understood as genre-dependent hostile speech acts, since within the analysis of genre are captured the speech act theoretical dimensions necessary to understand the dynamics of hate speech. By examining the communicative goals of a given genre and the genre norms permeating it, one may approach how fringe forms of hate speech like IGHS operate and are perceived by the DC. A genre-conscious, corpus-assisted discourse analysis of /lgbt/ should therefore prove productive in shedding light to the fuzzy borders of hate speech.

### 3. Materials and methods

To understand the dynamics of IGHS as I have outlined it above, a central element is the environment within which this language appears and the norms of the genre that govern and make possible this form of hate speech. In this section, I will first explore 4chan as a platform, particularly as it pertains to hate speech production, and then move on to the subforum /lgbt/ to illustrate its position within online discourse, the broader genre to which it belongs and hostile language online. In the second half of this section, I will outline the methodological design applied to approach the topic of hate speech, while remaining cognisant of its limitations.

### 3.1. Materials and 4chan

The data for this study was gathered from the social media website 4chan; a free, open access discussion forum, self-described as a “simple image-based bulletin board where anyone can post comments and share images” (4chan, n.d.a). According to previous research, 4chan “attracts around 22 million users a month worldwide”, of whom 46 percent are from the United States (Zelenkauskaitė et al., 2020: 3). Additionally, the website is heavily dominated by men, with 70 percent of users being “males aged 18–34” and features “around 900,000–1,000,000 messages per day” (Zelenkauskaitė et al., 2020: 3). Furthermore, 4chan has a considerably shorter reach when compared to other, related text and image-based forums, like Reddit, for example (Rieger et al., 2021: 4), which situates 4chan within the fringes of online discourse. The website is semi-anonymous, meaning that identification of users is possible through optional usernames, but few people on the forum choose to use these usernames, with a vast majority of the users choosing to remain anonymous. Currently (as of April, 2024), 4chan hosts 76 different discussion forums, sorted to seven different categories: *Japanese culture*, *video games*, *interests*, *creative*, *misc.*, *adult* and *other*. The board chosen for this study to inspect the dynamics of in-group directed hate speech is one that is found on the category *other*: */lgbt/*.

#### 3.1.1. */lgbt/*

The 4chan subforum */lgbt/*, as stated in the page rules of 4chan, “is for the respectful discussion of LGBT lifestyle and the LGBT community” (4chan, n.d.b). While absolute claims about the gender identity of */lgbt/*’s userbase is impossible due to the anonymity afforded to its users by 4chan, a genre-conscious approach provides us some evidence on this account. For example, it can be assumed that a central function of the forum, instead of discussing LGBT issues more broadly, is to focus on issues pertaining particularly to transgender people. By the DC of */lgbt/*, the forum is often jokingly referred to as “/tttt/” due to the topics of the forum mostly centring transgender people. For example, in the following exchange, two members of */lgbt/*, discussing the harmfulness of “political transness” or “theyfabism” (more on this in section 4.2.2), the latter speaker refers to */lgbt/* as */tttt/*:

(1a) transexual is biological transgender is political We should recognize political transness (theyfabism) as harmful to us actually transitioning people.

(1b) lets do it somehow /*tttt*/<sup>2</sup> is actually still quite influential when it comes to online trans culture. we could probably make it happen

Notice how, in (1a), the speaker refers to themselves as an “actually transitioning” person, and the second speaker, in (1b), couples this goal presented by 1a with /*tttt*/ and its influence on “online trans culture”. This exchange perhaps best illustrates, not only the significance of genre on hate speech by recognising how “influential” genres like /*lgbt*/ are on culture and discourse, but also the acute awareness of this influence by the members of the DC of /*lgbt*/. Despite this awareness however, it should be noted, that outside of the genres of 4chan and /*lgbt*/, the subforum has a resoundingly negative image within the transgender community (see section 4.2.3 for further evidence on this). Therefore, /*lgbt*/ as a subforum of 4chan, a website recognised by previous research as hostile towards the LGBT community (Rieger et al., 2021) should prove productive in the analysis of in-group hate speech. Furthermore, while other subforums like /*pol*/, for example, have been studied extensively with various research interests (see for example Zelenkauskaitė et al., 2020; Rieger et al., 2021), to my knowledge, no study has specifically explored /*lgbt*/.

### 3.1.2. Overview of corpus data

The corpus data for this study was manually gathered between November 2023 to January 2024. Due to 4chan rules (4chan, n.d.b) banning the use of scrapers and automatic data-retrieval tools, the data for the corpus was collected by hand without additional third-party software. The collection process occurred indiscriminately, meaning that posts included in the corpus were selected randomly, without preference for, for example, posts containing certain terminology or comments exploring given topics. This was done to assure that a representative corpus of organic /*lgbt*/ discourse would be compiled and to avoid cherry-picking for specifically hateful content to suit a given hypothesis. To establish statistical grounds for diachronic variation between datasets, two datasets from different time periods were collected. Therefore, the corpus was divided into two subcorpora of approximately equal size, one consisting of posts from 2015 to 2016 and the other from 2023 to 2024. Due to 4chan’s automatic deletion of old posts from the forum, the archival website 4ChanArchives (4chanarchives, n.d.) was accessed to gather the deleted material for the 2015–2016 subsection of the corpus.

---

<sup>2</sup> Italics in this and all subsequent examples are my additions.

**Table 2.** Overview of datasets

/lgbt/	2015–2016	2023–2024	Total
Words	102739	109264	212003
Comments	5419	5174	10593
Avg. wordcount	18.96	21.12	20.01

As displayed on Table 2 above, the corpus collected for this study contains more than 200,000 words and over 10,000 comments in total. For each subcorpus, the wordcount is greater than 100,000 and the number of comments  $\geq 5000$ . The average wordcount per comment across datasets is approximately 20 words, with the 2023–2024 subcorpus having slightly longer comments on average. All things considered, a corpus of 200,000 words is still relatively small when compared to some of the largest corpora that reach up to billions of words. Therefore, more intimate, pragmatic approaches ought to be considered where quantitative data is not sufficient.

### 3.2. Methodology

Since both the target audience, as well as the members producing the language on /lgbt/ are, ostensibly, transgender, to ascertain whether a phenomenon of in-group directed hate speech exists on the forum, one must first examine how transgender people are discussed on the forum. One way to go about this, is by establishing a list of terminology used by the DC to refer to transgender people. To do this, throughout the data gathering, parsing and cleaning process, a simultaneous process of close reading was applied, whereby the terminology used to refer to transgender people was documented on a separate file each time new terminology emerged. This list of /lgbt/’s transgender lexicon and the frequencies of the individual lemmas can thereafter be examined in greater detail with a corpus linguistic design. After compiling the corpus and the list of terminology used to refer to transgender people, AntConc (Laurence, 2023) was used to explore the frequencies and semantic preferences of the transgender lexicon of /lgbt/. Simultaneously, this lexicon was contrasted with itself, to see which lemmas obtain the most amount of semantic utility on the forum and which lemmas remain relatively obscure. By doing this, a statistically informed hierarchy of the linguistic items used to refer to transgender people can be established and thereafter analysed.

After RQs 1 and 2 were sufficiently answered using a corpus linguistic design, a more qualitative approach, following Parvaresh’s (2023) framework, was necessary to answer RQ3 and RQ4. By identifying particularly salient terminology present in the corpus, a deeper, pragmatic exploration of the lemmas in-context (and as parts of

longer discussions) is made possible. These excerpts containing salient terminology were then examined in greater detail through an additional process of close reading to gain insight into the discourse prevalent on /lgbt/ beyond quantitative data. A discourse analytical approach was applied particularly where investigation into specific exchanges between speakers was necessary, to ascertain what speech act theoretical forces are present and in what way. To understand, for example, the intentions of a speaker in uttering something, the information provided by a single comment might not yet be explanatory enough. For example, the word “tranny” may sound hostile out of context because of its association of being a transphobic slur, but in context it may not communicate any hostility at all towards transgender people. Therefore, to determine whether a term carries hostile communicative value, one must familiarise themselves with the textual environment within which an utterance is occurring, to understand what the speaker might mean with said utterance and, also, how its audience might perceive it. To do this, one may examine the comments preceding and following an utterance and the responses that utterance has received. From the comments preceding an utterance, one can determine what a speaker's intentions may be founded on when posting, and from the responses of an utterance, one can measure the perlocutionary force of a speaker's speech act. Here, the speech act theoretical framework applied for this study is particularly important, because within speech act theory is verbalised these three dimensions of language: intention, lexical content and audience perception, through which hate speech may be understood. Therefore, to grasp, not only the locutionary, but the illocutionary and perlocutionary forces of an utterance, as is key for the study of hate speech, pragmatic discourse analysis of comments containing salient lemmas to the DC of /lgbt/ was applied where quantitative approaches failed to reach deep enough.

## 4. Results and analysis

In this section, I will explore the transgender lexicon of /lgbt/ in closer detail by examining the frequencies and semantic fields of the various general and genre-specific lemmas referring to transgender people. I will primarily focus on the genre-specific terminology of /lgbt/, in order to better gauge the attitudes particular to the DC of /lgbt/ since, as I claim, a community's attitudes are best reflected in the language created specifically for and by the members of that community. To address RQs 1 and 2, regarding the variation and frequency of the transgender lexicon of /lgbt/, the

corpus data and results will be explored in greater detail in subsection 4.1 below. To answer RQs 3 and 4, regarding the attitudes of /lgbt/ towards other transgender people and whether the language produced by these attitudes could reasonably be understood as intragroup directed hate speech, the results of the pragmatic discourse analysis of the genre-specific lexicon of /lgbt/ will be explored in subsection 4.2.

#### 4.1. Transgender lexicon of /lgbt/

The process of close reading the data gathered for the corpus produced twenty distinct nouns used to refer to transgender people. In Table 3 below, the twenty general and genre-specific transgender terminology and jargon are listed in order of frequency. For clarity, the lemmas I claim to be specific to /lgbt/ as a genre are presented in italics.

**Table 3.** Distribution of the transgender lexicon of /lgbt/ in order of frequency

Lemma	2015–2016	2023–2024	Total	Freq/10000 words
trans	254	277	531	23.90
tranny	78	176	254	11.43
<i>AGP</i>	47	189	236	10.62
<i>hon</i>	41	96	137	6.17
FTM	59	45	104	4.68
MTF	48	47	95	4.28
transgender	33	28	61	2.75
<i>poon(er)</i>	0	29	29	1.31
transsexual	8	20	28	1.26
<i>tranner</i>	0	25	25	1.13
<i>passoid</i>	0	24	24	1.08
<i>transbian</i>	10	14	24	1.08
trap	7	16	23	1.04
<i>troon(er)</i>	0	14	14	0.63
<i>theyfab</i>	0	12	12	0.54
tgirl	1	9	10	0.45
AMAB	0	6	6	0.27
AFAB	0	5	5	0.23
<i>transOCD</i>	0	3	3	0.14
<i>trender</i>	1	1	2	0.09
TOTAL	587	1036	1623	73.03

It should be noted here, that there is room for debate about my classifications of general and genre-specific jargon, since the borders between genres are always fuzzy at best and lexical bleeding from genre to genre occur constantly. One could argue that the word “transbian”, for example, has wider usage outside of 4chan and /lgbt/, and is therefore not specific to /lgbt/ as a genre. My process of eliminating terms from the general category was based on the following three questions:

(i) is the term in common parlance, so much so, that e.g., a newspaper would be able to use it in an article without providing additional explanation?

(ii) Is the term documented in reputable dictionaries, such as Oxford English Dictionary et cetera?

(iii) And finally, is the term more commonly associated with other genres or DCs instead of /lgbt/?

If not, then I would see it that, for the purposes of this study, they are sufficiently genre-specific. Therefore, out of the twenty words, half are determined genre-specific: “AGP”, “hon”, “poon(er)”, “tranner”, “passoid”, “transbian”, “troon(er)”, “theyfab”, “transOCD” and “trender” and the rest are general: “trans”, “tranny”, “FTM”, “MTF”, “transgender”, “transsexual”, “trap”, “tgirl”, “AMAB” and “AFAB”. Furthermore, the genre-specific jargon of /lgbt/ represents around a third, or  $\approx 31\%$ , of all lexical items identified in the corpus, and therefore constitutes a substantial amount of the discourse pertaining to transgender people and transgender issues. Additionally, the general noun “trans” is by far the most frequently utilised variant and, by itself, also represents around a third, or  $\approx 33\%$ , of all the lexical items referring to transgender people. The diachronic data displayed on the second and third columns of Table 3 also illustrates that, while neutral and general terminology, such as “FTM” (acronym for “female to male”), “MTF” (acronym for “male to female”) and “transgender” lessen in frequency over time ( $\approx 83\%$  in 2015–2016 and  $\approx 60\%$  in 2023–2024), genre-specific terminology, such as “AGP”, “hon” and “poon(er)” experience exponential growth (from  $\approx 17\%$  to  $\approx 40\%$ ), providing further evidence of the importance of examining such language in closer detail. Notice also, that the total amount of transgender terminology used, nearly doubles from 2015–2016 to 2023–2024, which, in large part, is fuelled by the increased adoption of this genre-specific jargon in place of and in addition to more general lingo.

As presented on Table 3 above, while the most frequently used noun on /lgbt/ is the tonally neutral “trans”, the three following lemmas: “tranny”, “AGP” and “hon”, are notably more dubious in tone. For example, “tranny” is, of course, a common slur often weaponised against transgender people (Kurrek et al., 2020: 139) and, as a slur, can therefore be understood to possess a hostile locutionary force. However, in the interest of charitability, and as “concepts do not straightforwardly align with the meanings of words” (Culpeper, 2021: 6), one should not immediately rush to

conclusions about the genre-specific semantic content of any word sans contextual cues. Reclamation of slurs and pejoratives “[striving] for a reversal of social norms by breaking established linguistic conventions” (Jeshion, 2020: 109) by minority groups is well documented in previous research, and closer examinations of the concordance lines of “tranny” does indeed reveal a dynamic more akin to reclamation than hostility. Consider, for example, the following excerpt of a user valorising the beauty of another transgender woman:

(2) i'm a big fan of certain clocky *tranny* phenotypes, and the brunette checks off a buncha boxes... hhhhhhhng

Here, instead of ridicule, the t-slur is used to celebrate the “clocky tranny phenotypes”, meaning features easily identifiable as belonging to a transgender person. So, while the locutionary force of the slur itself may be hostile, in genre-conscious context, the complete speech act reveals a positive appraisal of a person’s physical attributes. To be clear, this is not to say that violent iterations of the slur are absent from the data, but in the case of “tranny”, all hostile instances represented in the corpus are self-referential, meaning that they target the speaker, and not another person or a group. In other instances where the word “tranny” is not used in a self-referential or positive manner, the lemma is attached to other, more hostile jargon to which I will get to in the next section.

## 4.2. Genre-specific lexicon of /lgbt/

Let me now move on to some of the genre-specific jargon of /lgbt/, the semantic variation of this lexicon and explore in closer detail two lemmas in particular: “AGP” and “hon”, as they are substantially more in-line with the dynamic of IGHS as I have outlined it above in subsection 2.2.3, than the aforementioned “tranny”, for example. As evidenced by the use of the milk emoji in subsection 2.2.1 above, the genre-specific jargon, as opposed to general terminology, is a matter of interest when it comes to the study of online hate speech. From the pragmatic examination of this genre-specific lexicon, one may establish certain functions particularly important for the genre and its DC. This is because the creation of new, genre-specific language is only necessary in such situations where appropriate language does not already exist to describe the same phenomenon. And, while new language is created constantly in genres and by the DCs permeating them, the language that *sticks*, or remains quantitatively relevant over time, suggests that it has obtained a degree of utility for the genre worth inspecting. So,

by analysing this lexicon, one may understand what the genre deems important enough to produce new language for, and what the DC thinks is not sufficiently expressed by general terminology.

#### 4.2.1 Semantic variation of genre-specific lexicon

In Table 4 below, I have situated the ten genre-specific transgender lemmas in six, broad semantic categories: *sexual*, *non-passing*, *uncivil*, *passing*, *not-trans-enough* and *mental illness*. The most frequently occurring category represented in the corpus

**Table 4.** Semantic categories of the genre-specific transgender lexicon in order of frequency

Semantic category	Lemma	Combined frequency
Sexual	<i>AGP, transbian</i>	260
Non-passing	<i>hon, troon(er), poon(er)</i>	183
Uncivil	<i>tranner</i>	25
Passing	<i>passoid</i>	24
Not-trans-enough	<i>theyfab, trender</i>	14
Mental illness	<i>transOCD</i>	3

are lemmas related to the semantic category of sexuality, meaning lemmas associated with sexual tendencies, paraphilias or sexual identity. The second most popular, and lexically most rich category is the semantic category of poorly passing transgender individuals. *Passing*, in this context, means the ability to either appear cisgender or one's desired gender that is different from the gender one was assigned at birth. This notion of *passing* is often contrasted with the act of "clocking", meaning the ability to distinguish a transgender person as transgender or, in the case of looking "clockable" (explored above in section 4.1), looking like a transgender person, which often carries a negative insinuation. The following four categories occur significantly less frequently in the corpus and cover topics relating to transgender identities associated with mental illness ("transOCD"), diminishing transgender and non-binary identities to social trends ("theyfab" and "trender"), well passing transgender people ("passoid") and general, uncivil iterations of the words "transgender" ("tranner").<sup>3</sup> In this text, I will primarily focus around the two most popular categories.

#### 4.2.2. "AGP"

As presented in Table 3 above in subsection 4.1, "AGP" and "hon" represent the fourth and fifth most popular lemmas respectively and both specifically target transgender

<sup>3</sup> "Tranner", in typical 4chan lingo fashion, is a portmanteau of the words "trans" or "tranny" and "4chaner", i.e., a 4chan user.

women instead of transgender people of all genders more broadly. Altogether, the two lemmas constitute nearly a fourth, or  $\approx 23\%$ , of the whole transgender lexicon represented in the corpus gathered for this study. Let me first explore the lemma “AGP”, its collocates and then explore some examples of the lemma where “AGP” is used in such a way that I would see it to constitute as hate speech, before moving onto “hon”.

AGP”, an abbreviation of *autogynephile* or *autogynephilia*, is derived from Greek for “love of oneself as a woman” (Moser, 2010: 791). The concept has its roots in Blanchard’s (1991) typology of transvestism whereby autogynephilia is understood as a “male’s propensity to be sexually aroused by the thought or image of himself as a female” (Blanchard, 1991: 235). In his work, Blanchard routinely equates phenomena most often associated with transgender people (such as gender dysphoria, i.e., the distress one feels when one’s gender identity differs from the sex one was assigned at birth) with this autogynephilic paraphilia (Blanchard, 1991: 246). It perhaps goes without saying that this theory is in contrast with the way most transgender women perceive themselves. Because Blanchard’s typology seemingly reduces transgender identities to a matter of sexuality and, in the case of autogynephilic transsexuals (to use Blanchard’s outdated terminology) to a taboo sexual paraphilia, it is vehemently rejected by many transgender people, because it reinvigorates a well-established, transphobic narrative that transgender women are merely sexually perverted (and therefore dangerous) men. Evidently then, it is of course immediately concerning to see a word like “AGP” in such popular use in a community of transgender people. However, as illustrated by the genre-specific use of the word “tranny” explored in the section above, one must examine the word in-context before coming to any conclusions about its potential hostile communicative value. Therefore, to understand what the DC of /lgbt/ means with “AGP”, I ought first to refer to the community’s own descriptions of the term. In the following exchange, two members of /lgbt/ discuss the meaning of the noun:

(3a) Is *AGP* real? /lgbt/ seems split on the issue. If it is real then what's the difference between it and a normal transgender person?

(3b) A transgender person has the inner gender of the sex opposite they were assigned. A guy with *AGP* will get off to the idea of being a woman, doesn't say anything about their gender.

Notice in (3a), how “AGP” is distinguished from “normal transgender” people and how in (3b), transgender people are wholly separated from “AGPs”, who are described as men with a paraphilic interest of idealising themselves as women. One could therefore conclude that, to the DC of /lgbt/, “AGP” has nothing to do with transgender people, and that any hostility directed at “AGPs” by members of /lgbt/ is not, in fact, in-group directed hate speech, but a more general form of out-group directed hate speech, because “AGPs” are distinctly *not* transgender. Here, however, I come to the very core of what I understand to constitute IGHS, namely, a form of self-othering and resentment that lashes out at other members of the in-group. A key dynamic of IGHS is precisely this project, present in both examples (3a) and (3b), of separating oneself from a traumatic or socially unacceptable identity. To illustrate this, in contrast to (3b), another user describes “AGP” with the following, illuminating words:

(4) People just kind of use *AGP* as a shorthand for stereotypically trannyish, clocky, or bad

Here, “AGP” is not detached from transgender people like in (3b), or even “normal transgender” people, like in (3a). Instead, in (4), “AGP” is used to describe transgender people who are easily distinguished, or “clocked”, as transgender and “trannyish”, and who are therefore “bad”. This, I think, is a more honest and less psychologically defensive rendition of the term but reading through the concordance lines of “AGP”, I am inclined to believe that a synthesis of the two aforementioned descriptions to represent the DC’s understanding of “AGPs” most accurately. That is to say, “AGPs” are understood as sexually perverse, poorly passing transgender women, who are therefore referred to as men by other transgender women as a form of self-validation through othering, mockery and, I claim, hate speech.

Furthermore, the collocates of “AGP” presented on Table 5 below seem to support such a reading, whereby an intention to distinguish “AGPs” from “real” or otherwise admirable transgender women, the “HSTS” (acronym for “heterosexual transsexual”) is reflected in at least the collocates “hobbies”, “voice”, “voices”, “HSTS” and “flamer”. In “hobbies”, “AGPs” are claimed to have specific, autogynephilic hobbies distinct from other transgender women. Furthermore, “AGPs” are associated with “reppers”, meaning people who repress their autogynephilic desires or desires to realise their sexual fantasies, perhaps because to be an “AGP” is also associated with

disfigurement and therefore ought to be avoided, as displayed on the ninth item on Table 5.

**Table 5.** Collocates of “AGP” with a window span from 5L to 5R in order of likelihood

Collocate	FreqLR	Likelihood	Example
hobbies	13	98.389	“what makes them AGP? the voices, the <i>hobbies</i> and the specific people you're talking about”
voice	26	96.716	“An agp <i>voice</i> is a clocky <i>voice</i> that doesn't sound like a flamer gay voice”
HSTS	16	94.899	“AGP is anime, <i>HSTS</i> is makeup tutorials”
voices	13	79.783	“WHAT ABOUT AGP WITH FLAMER VOICES ER AND WIDE PITCH RANGES OR RISING INFLECTIONS JESUS FUCKING CHRIST YOU ARE SO FUCKING STUPID”
tgirls	5	36.073	“The fact that I like non-passing <i>tgirls</i> with AGP voices”
repper	5	36.073	“literally half of those are just signs of being an agp <i>repper</i> ”
flamer	6	34.182	“If you aren't a <i>flamer</i> , you're classified as agp”
disfigured	5	34.783	“As the resident <i>disfigured</i> AGP, I'm trooning I guess, but yeah it kind of sucks.”
transexuals	4	28.857	“AGP <i>transexuals</i> and Cis AGPs”
valley	4	27.008	“NO what MAKES a VALLEY GIRL voice AGP?”

Finally, “AGPs” are distinguished from the “flamer”, which, according to the Oxford English Dictionary, is a derogatory term for a “gay man whose sexuality is regarded as conspicuous or apparent, esp. because of his flamboyant, affected, or exaggerated mannerisms or clothing” (OED, 2023). “AGPs” are argued to have a particular “clocky voice” that is distinct from the “flamer voice”, which, in the context of /lgbt/, seems to denote the voice of a male-attracted transgender woman. Yet, a close reading of the concordance lines of “AGP” reveals that, while “AGPs” may be separated from the “flamers”, they are not necessarily separated from “faggots”, interestingly enough. Consider the following example:

(5a) Kill yourself *AGP* faggot, im talking about the real trannies. not the ones who date other effortless andro trannies who are still wearing boy clothes.

In the above example, “AGP” is paradoxically coupled both with the words “faggot” and “trannies”. This then poses the question; if “AGP” is supposed to address transgender women, how can it be bundled with “faggot” (which primarily describes

gay men), and yet, if “AGP” is supposed to describe men, why is it coupled with “trannies” (which, in this instance, refers to transgender women)? As I see it, the speaker in (5a) is perfectly aware of the mutual impossibility of their utterance, yet proceeds with it anyway, because they know it to be maximally hurtful, by targeting two key insecurities of many transgender women:

(i) being accused of sexual perversion or degeneracy (by being called an “AGP”) and

(ii) being called a man or not-a-real-woman (by proxy of being called a “faggot”).

On the latter point, denial of one’s self-identified gender is a common trope on /lgbt/, and this vitriol against transgender people perceived as not-transgender-enough, or not-appropriately-transgender, is present in some of the DC’s other genre-specific jargon like “theyfab” and “trender” as well. “Theyfab”, for example, a portmanteau of the third-person pronoun “they” (used by many non-binary people) and “AFAB”, (acronym for “assigned female at birth”), follows the misogynistic and transphobic trope that non-binary identities are superficial social trends, which is why mostly women, wishing to be seen as trendy, identify as non-binary and are therefore also called “trenders”. This is, of course, demonstrably not an accurate assessment of non-binary people, but nevertheless sheds light to the way some of the members of /lgbt/ who choose to use this language perceive transgender and non-binary people.

To now examine the speech act theoretical dimensions of (5a) with hate speech in mind, I think it is evident that the locutionary force of the utterance, namely, an imperative demanding the interlocutor commit suicide, is undoubtedly *hostile*. The speaker is also specifically targeting the interlocutor’s gender identity in their utterance, if not by “AGP”, then at the very least by the referent “andro trannies”. Furthermore, it seems that the motivation to utter such a sentence on the speaker’s part stems from the distaste they feel towards “effortless andro trannies who are still wearing boy clothes”, and an effort to juxtapose “real trannies” from mannish, sexually perverse, or otherwise bad transgender women. Therefore, on the illocutionary level, a *hostile* intention to separate, or other, sections of the in-group into a peculiar quasi-out-group, seems most explanatory here. To then take a look at the perlocutionary

force of the speaker's utterance, one may inspect the responses the comment received. One 4chan user responds with the following words:

(5b) lesbian transgender women don't exist? gtfo

In (5b), the interlocutor notices the paradoxicality of the original speaker's utterance bundling "faggot" with "AGPs" and transgender women and aims to reason against this juxtaposition by remarking the existence of "lesbian transgender women". The commenter then finishes their response with a curt "gtfo" (acronym for "get the fuck out"). According to previous perceptual research, the most common strategy utilised by victims of hate speech is passivity and withdrawal, while strategies centring constructive activity like reasoning as well as aggressive reactions are least common or totally absent in victims' behaviour (Leets, 2002: 352). In (5b) however, both constructive activity in the form of reasoning with the original speaker, and aggression in the form of "gtfo", are present in the interlocutor's response. If the person in (5b) was directly harmed by the original speaker's comment, their response does not reveal it. In fact, considering that, mere moments ago, they had been ordered to commit suicide, the interlocutor's response assumes a strikingly nonchalant stance, suggesting that the violent language displayed in (5a) is common enough of an occurrence on the forum not to warrant a more passionate response.

#### 4.2.3. "Hon"

Let me now move on to what I perceive to be the most hostile term out of the transgender lexicon of /lgbt/: "hon". "Hon" has its roots in the Baltimorean term of endearment "hon", short for "honey" (Puglia, 2015: 139), but on /lgbt/, the term assumes a notably less benign tone. Namely, "hons" are the "ugly", "cringe", "disgusting" and "freakish" transgender women, who, among other things, are called "fat ogres"<sup>4</sup> and are routinely contrasted with the well passing transgender woman, the "passoid". In "hon", are combined the disgust for poorly passing transgender women rampant on /lgbt/, and the envy for the successfully passing transgender women. On the subreddit r/asktransgender (Reddit, n.d.b), in a discussion thread debating the terminology pertaining to transgender people, a commenter describes the two lemmas, "hon" and "passoid", with the following words:

---

<sup>4</sup> All the aforementioned words in quotation marks are represented in the corpus as direct premodifiers of "hon", meaning that they immediately precede the noun.

I think "passoid" just means a trans person who passes ridiculously well. The opposite of a passoid is a "hon" who is a non-passing transfem but maybe gets told she looks great regardless. "You look great, hon!" And yeah, these are very toxic 4chan trans things... Avoid if you want to maintain your sanity and self-respect. (mouse9001, 2023)

While the commenter establishes the two lemmas, "hon" and "passoid", as conceptual opposites of one another, they also attribute the terminology to "toxic 4chan" lingo and urges others to avoid the website if they wish to maintain their "sanity and self-respect". Let me now inspect where such a warning might stem from. In Table 5 below, the ten most frequently associated collocates of "hon" are presented in order of likelihood of appearance alongside the noun. From these ten collocates, some patterns seem to emerge:

- (i) "hon" being a difficult or negative stage of one's gender transition ("mode", "moder", "twink" and "permaboymoding"),
- (ii) "hons" associated with sexual abuse ("rape" and "twink"),
- (iii) "hon" as bestial, perverse or otherwise disgusting ("beast", "passoid" and "prime").

In (i), "hon" is used to describe a difficult or uncomfortable stage of one's transition from a male-presenting to female-presenting, and jargon like "hon moder" centralise this particular experience. "Hon moder", as presented in Table 6 below, is simultaneously contrasted with "man moder", or a person who represses their urge to transition, while also being described as something unbearable with the words "I can't be a hon moder either". In (ii), a similar trope present in "AGP" is carried over to "hon", whereby "hons" are associated with sexual perversion and misconduct. In particular to "hon", an association with sexual abuse and rape seem to be especially strong. As presented in Table 6, transgender women perceived as "hons" are called "malebrained rape hons" who have to delude themselves with "copium" (a portmanteau of "cope" and "opium") to "not feel as bad" about themselves, their gender and sexual desires. However, for the purposes of this study, the latter category, (iii) "hon" as bestial, proves most revealing. Here, a dynamic of quasi-out-grouping sections of the in-group is particularly strong and is accomplished by juxtaposing well passing and beautiful transgender women with "hons". This othering is often formatted in such a way that it may not manifest in similarly violent utterances as (5a), for example, yet nevertheless,

**Table 6.** Collocates of “hon” with a window span from 5L to 5R in order of likelihood

Collocate	FreqLR	Likelihood	Example
mode	11	69.479	“you don't have to <i>hon mode</i> either”
rape	12	59.800	“they did it as well and are on copium overtime to not feel as bad about being malebrained <i>rape hons</i> ”
moder	4	33.180	“I can't be a man moder anymore. But I can't be a <i>hon moder</i> either.”
twink	6	26.614	“Can I <i>twink hon mode</i> and not look like a predator?”
beast	3	23.090	“ <i>hon beast prime</i> are always a transbian.”
dosed	3	23.090	“And i originally was referencing just knowing you're not being <i>hon dosed</i> .”
stap	2	19.570	“ <i>HON STOP IT STAP IT STOP IT STOPIT</i> ”
permaboymoding	2	19.570	“i know being a hon sucks, but if given the option between severely depressed male and slightly depressed <i>permaboymoding hon</i> ”
passoids	3	18.156	“ <i>Passoids</i> in <i>hon</i> spaces are made to feel priviledged and like their struggles are fake.”
prime	3	16.958	“Sometimes I wonder what the weirdos posting this stuff actually look like because it sounds funny if it's some smol tranner, but I'm 99% it's just <i>hon beast prime</i> ”

it portrays a deep resentment and a sense of disgust towards what the DC calls “hons”. To illustrate this, consider the following three examples:

(6) based, small trannies need to speak up more! we let too many *heighthons* dominate the discussions and board

(7) ive fucked boys without saying a thing and they didnt clock me, we just dont post here cuz mostly *hons* post here. we lurk and laugh.

(8) Still, I don't pass and can no longer bring myself to dress up in pubic, especially after seeing all the cringe *hons* on this site (sorry for being transphobic I still have love in my heart for cringe *hons*)

In (6), “small trannies” (petiteness being associated with femininity) are contrasted with “heighthons” (height signifying masculinity or deviance from traditional femininity), with a revealing discursive narrative dispersed throughout of “small trannies” simply submitting and “[letting] [...] heighthons dominate the discussions and board”. This reflects a common psychosexual dynamic where men are described as dominant and proactive, while women are submissive and passive. So, not only does (6) aim to other the “heighthon” but does this by alluding to her masculinity. In (7), “hons” are pictured as targets of laughter and ridicule, or, what the DC calls

“lolcows”; cows to be milked for laughter (Bishop, 2014: 157). Additionally, an attempt to separate transgender women who cannot be “clocked” even during sexual intercourse from “hons” is also evident here. Out of the three excerpts, the most interesting to the study of IGHS is example (8), which openly displays both a deep sense of self-hatred and shame, expressed with the words “I don't pass and can no longer bring myself to dress up in pub[l]ic”, as well as a deep in-group resentment in: “after seeing all the cringe *hons* on this site”. Furthermore, the commenter also illustrates a genuine recognition of and a regret for this resentment with the words “sorry for being transphobic I still have love in my heart for cringe *hons*”. Here, we exist on the very fringes of what I would conceive to still constitute IGHS, since, as the commenter themselves demonstrably recognises, the utterance both *is* and therefore also ostensibly *intends* to be “transphobic”, yet the comment itself is primarily self-directed, with the exception of its reference to “cringe *hons*”. Additionally confounding, the commenter simultaneously goes on to display the “love in [their] heart” for said “cringe *hons*”, making its hostile communicative value even more ambiguous.

Now, it should be noted that the speech act theoretical framework employed in this text, particularly Austin's (1975) notion of illocutionary force, runs into some conceptual problems here. Some researchers, like Allwood (1977), might note that the mere *awareness* of a potential outcome, like causing harm with a transphobic statement, does not yet mean that one *intends* to do it, nor that it is enough evidence to establish what a speaker's illocutionary force with an utterance might be. After all, there are many things which we end up doing knowingly, which we do not necessarily intend to do, in the very same way we can intend to do something and fail at doing it. One can, as Bratman (1984: 400) notes, intend to go for a run and while running, break the soles of one's running shoes without ever intending it. Here, the argument is that breaking the soles of one's shoes is a probable outcome of running (and a potential outcome an experienced runner must be aware of), but that this awareness does not yet amount to intention. Therefore, it could plausibly be argued, that the speaker in (8), only reveals an awareness of the potential for causing harm with their transphobic statement, without ever intending it, as in the case of Bratman's runner who acknowledges the possibility of breaking his shoes without actually intending to do so.

In the case of (8), however, I think it is difficult to conceive how an utterance like “sorry for being transphobic” does not insinuate an intention to harm, since the speaker expressly recognises that their utterance is harmful enough to warrant an

apology, and still goes on to post it on the forum regardless of this awareness. It would seem to me that, if the speaker did *not* intend to be transphobic, they either:

- (i) would not have included an apology for transphobia in the first place (i.e., the speaker is unaware of the harm and therefore not intending it), or,
- (ii) while recognising that the utterance necessitates an apology, modified the content of the comment *before* posting, so that utterances apologising for transphobia would not be necessary (i.e., the speaker is aware of the harm but not intending and committing to it).

By posting a comment the speaker themselves call “transphobic”, they, at the very least, commit to an utterance they know will possibly hurt people. Whether this utterance is intentionally harmful, I think, is suspect only to the same degree that sentences like “I am not racist, but...” are, because both utterances lexically recognise the necessity to pre-empt the harm that their following bigoted words may cause. It may be, that the speaker in (8) does not *want* or *desire* to cause harm and therefore not intend it, in the same way that the person who says “I am not racist, but...” might not *want* to hurt ethnic minorities. However, since we cannot hope to determine what a person’s true desires and wants actually are, to me, what seems paramount to intention in hate speech, is whether or not the speaker commits to those following transphobic or racist words which they know to be hurtful, and not whether they do it with a will to be hurtful. In this sense, I take to Walter’s remark that, instead of wilful intention, what matters is that “the offender intends to express the insult and is aware that to do so will likely demean the victim's identity” (Walters, 2014: 70).

#### 4.2.4. Genre as a mediator of hate

Above, I have outlined some examples of exchanges on /lgbt/ that I think sufficiently look like hate speech, not only because of their hostile lexical content, or the locutionary force of these utterances, but also because of the speakers’ intentions, or the illocutionary force displayed behind these utterances. However, still the question of harm, and the perlocutionary force of these utterances remains only vaguely addressed. Let me now examine this issue and attempt to outline why it is I think hate speech that does not directly harm its audience, may yet be hate speech, by exploring the use of “hon” in genre-conscious context.

Although many of the hostile iterations of “hon”, explored above in section 4.2.3, rely on a strategy of othering that does not overtly betray violence, that is not to

say that overtly violent iterations of the noun are non-existent. On the contrary, with “hon”, overtly hostile language akin to example (5a) is plentiful, like in the following two examples:

(9) Fuck you delusional *hon*

(10a) kill yourself you ugly *hon*

In (9), the interlocutor is cursed out, with “hon” being modified with “delusional”; delusion being a common trope of transphobic language reflected in, for example, *the attack helicopter joke* (explored above in section 2.3.2). In (10a), the structure of the utterance is almost identical to the first half of (5a), namely, an imperative to commit suicide and, like (5a), the locutionary force of the utterance is obviously very violent. However, what interests me here, is the response received by example (10a). Unlike in (5b), where the response displayed a degree of nonchalance to the command to commit suicide, the responses to (10a) are considerably more drastic. Consider the following two examples:

(10b) you suipushers are the worst fuck

(10c) this is not a nice site

In (10b), the original speaker in (10a) is accused of being a “suipusher”, which is 4chan lingo for “suicide pusher”.<sup>5</sup> This suggests that, not only was (10a) received poorly, but that it was accurately perceived as “suicide pushing”, which is undoubtedly harmful and hostile. In (10c), a commenter responding to (10a), states that /lgbt/ “is not a nice site”, and thus recognising, not only the harmfulness of the original speaker’s utterance, but the genre and DC of /lgbt/ as a whole. Therefore, it would seem that the perlocutionary force of the original speaker’s utterance is hostile, meaning that (10a) would *not* be an instance of IGHS, and instead something more general, like OHS, for example, since it displays a dynamic of hostility on at least the locutionary and perlocutionary levels. Here, however, one must consider the complexities introduced by genre to any concept of hate speech. If “suicide pushing” is common enough of an occurrence in a given genre to warrant the creation of newfangled terminology, is there

---

<sup>5</sup> “Suipusher” is related to other 4chan neologisms centring suicide, like “suifuel” (short for “suicide fuel”), meaning audiovisual content specifically designed to inspire interlocutors and members of the DC to consider self-harm and suicide.

not a degree of complicity towards this type of behaviour from the DC? What I mean by this, is that:

(i) if one is not involuntarily bound to participating in a DC that routinely engages in violent language and “suipushing”, and

(ii) if one is aware of this phenomenon of “suipushing”, as the speaker in example 10b shows themselves to be by appropriating the language of this phenomenon,

∴ to continue participating in such a DC to which one is not involuntarily bound, one must also, to a degree, know to expect violent language and “suipushing”.

Furthermore, a DC that actively and genuinely disapproves of hostile speech acts, such as imperatives to suicide, would be unlikely to host genre-specific jargon such as “suipushing” or “suifuel”, because language perceived as “suipushing” and content perceived as “suifuel” would be banned and members posting such content would be removed, making the creation of newfangled language for such speech acts unnecessary. However, this is clearly not the communicative goal of /lgbt/ as a genre. In fact, considering the examples explored above, one could claim the exact opposite is instead. This is, of course, not to say that hate speech portrayed in such examples as (5a) and (10a) may not be *perceived* as harmful (the reactions portrayed in (10b) and (10c) show otherwise), but that they are nevertheless tolerated and anticipated by their target audience, making the interlocutor of hate speech an *active partner*, rather than a *passive receiver*. This genre-informed anticipatory relationship towards IGHS is perhaps better represented by the following two examples responding to the subsequent discourse awakened by (10a):

(10d) We're gonna have to do a cleansing of 90% of this website to get it to be "nice"

(10e) it's almost like behind a curtain of anonymity where you can say whatever you want with no repercussions! someone should make a site where that's the whole point

In (10d), the speaker, responding to (10c), notes how the vast majority of /lgbt/ would have to be “cleansed” for the site to become “nice”, displaying an acute awareness of the harmful language produced by the DC. In (10e), the speaker goes as far as to allude, that “the whole point” of anonymous forums like 4chan is to “say

whatever you want [without] repercussions”. Both examples reveal a degree of anticipation to the attitudes and hostility expressed in (10a), by recognising that the majority of the discourse on /lgbt/ is akin to (10a), and that such behaviour is perfectly in accordance with the norms of the genre. It seems that, while harm may have been caused by (10a), one must ask what kind of harm? Much like victims of domestic violence who, due to financial or religious reasons, for example, seek no separation from their partner knowing that this decision will only lead to further violence, members of /lgbt/ face hate speech in full awareness of this hostility. And as it would be strange to claim that domestic abuse is not violence, on the grounds that the victim remains in a relationship where violence can be anticipated, claiming that the violent utterances explored above are not hate speech, simply because they are anticipated, would also seem counterintuitive. Particularly, since the examples explored above target the interlocutors’ protected characteristics, namely gender identity. Yet, if utterances like (10a) are hate speech, as I would claim them to be, then one must review what we mean by the term. One would think that the harm caused by hate speech is something its targets would be likely to avoid, if given the opportunity to do so. Yet, as expressed in (10d) and (10e), and to a lesser extent in (10b) and (10c), a genre-conscious approach reveals to us that while harm may be done, this is, in fact, an expected outcome and a genre norm. Furthermore, the fact that a vast majority of the genre-specific jargon produced by /lgbt/, such as “AGP”, “hon”, “theyfab” and “trender” are resoundingly negative in tone and content, suggests that a key function of the genre is to produce hostile language specifically about transgender people. This then puts a question mark to why we should singularly privilege the perlocutionary force in our attempt to define hate speech, when one may simultaneously anticipate and be harmed by the same instance of language. Is it not, at least in the case of IGHS, the interplay of genre and the locutionary, illocutionary and perlocutionary forces of our utterances, that determine what constitutes hate speech, and not merely the harm caused by disparate, disconnected utterances?

## 5. Discussion

In previous sections of this study, I have aimed to establish some of the contours of what I understand to constitute in-group directed hate speech. By centralising /lgbt/ and the genre-specific jargon pertaining to transgender people produced by the discourse community, my aim has been to provide evidence for a phenomenon I call

*in-group hate speech*. In this section, I want to return to the vast field of research on hate speech and violent online discourse, to compare and contrast how the findings of section 4 above relate to previous research.

Much of previous research examining online hostility has centred on the manosphere and far-right online discourse, or, what Massanari (2017) calls “toxic technocultures”, whose toxicity is, in part, enabled by the platforms and genres within which these communities are situated. In particular, 4chan is a website often associated with the manosphere and hosts multiple forums deemed to be a part of the community (Ribeiro et al., 2020: 2). The study by Ribeiro et al. further finds that out of several websites and subforums, a forum on 4chan, /pol, “is the community with the highest volume of toxic and hateful posts” and that 4chan hosts a significantly greater amount of hate speech than other, related websites like Reddit, for example (Ribeiro et al., 2020: 7). Massanari notes that the defining features of such hostile communities are “retrograde ideas of gender, sexual identity, sexuality, and race” and that they tend to “push against issues of diversity, multiculturalism, and progressivism” (Massanari, 2017: 333). Therefore, the discourse community of /lgbt/ is interestingly situated within an environment that seems overall very hostile towards LGBT people, which is why the toxicity and hate pervasive to /lgbt/ is hardly surprising. What interests me here, however, is that, as evidenced by its name, manosphere online communities typically centre men, while the audience of /lgbt/, evidenced by the most frequently used genre-specific terminology like “AGP” and “hon”, are seemingly women. The antiquated attitudes about gender, sexual identity, and sexuality noted by Massanari (2017), still persist on /lgbt/, yet they seem to be turned inwards. This phenomenon is perhaps best displayed by the regressive language on gender and sexuality rampant on the forum, like “AGP”, which represents an aversion towards what is widely understood as sexual degeneracy by the members of /lgbt/.

Additionally interesting, Massanari makes the claim that hostile online communities are recognised by their tendency to other “those perceived as outside the culture” (Massanari, 2017: 333). On /lgbt/ however, this dynamic is achieved, not by attacking members outside of the transgender community, but primarily by othering particular subsets of the transgender community to which the DC establishes newfangled language and through time, seems to increasingly focus on. This reflects previous research on genre that suggest that the smaller the membership and narrower the niche of that genre, the greater the volume of genre-specific lexicon (see for

example Liimatta, 2019). In this sense, notions like group face and collective face, discussed in section 2.2.3, are directly linked to genre, because, as noted by Tardy and Swales (2014), genres form around and function to fulfil shared communicative goals within which these collective and group face desires may be communicated and satisfied. Thus, it seems that, on /lgbt/, a key communicative goal of the genre is to realise a collective face desire to be viewed positively by the out-group of cisgender people. This desire is accomplished by suppressing sections of the in-group, like the “AGPs” or the “hons”, who are othered as “bad”, “cringe” and sexually undesirable transgender people and who are thus distinguished from the desirable “passoid”. In this sense, by targeting transgender people, the hostile speech acts of /lgbt/ resemble previous paradigms of hate speech that broadly define the phenomenon as “forms of speech that attack members of groups or classes [of] people identified by certain protected characteristics” (Brown, 2017a: 427) including, for example, race, religion, sexuality, and gender identity. In the case of /lgbt/ however, instead of broadly recognised characteristics or groups, the hate speech of /lgbt/ targets more specific sections of the transgender community.

Considering hate speech as a phenomenon, thus far, apart from some of its fuzzy borders, I have skirted away from the responsibility of defining hate speech at its broadest sense, partly because of the tremendous challenge that such a task implies. I have outlined various subcategories of hate speech, like OHS and IGHS, and attempted to provide necessary conditions with a speech act theoretical framework where possible. As it pertains to hate speech itself, however, I am more hesitant to provide such conditions. Primarily, because I believe the term to obtain most utility when left unshackled by limiting conditions that may exclude instances of language that intuitively seem like hate speech. Here, I take to Brown’s remark about the plurality of forms hate speech may assume, and how it is difficult, if not impossible, to “find only one purpose and a single definition” (Brown, 2017a: 73) to the term hate speech. Necessary conditions that require, for example, the perception of an utterance to be of harm or hostility for that utterance to be identified as hate speech, excludes instances of language and speech acts, like IGHS, that could easily be conceived as hate speech otherwise. On the necessary requirement of harm imposed on the concept of hate speech, Brown notes:

Consider [a] politician in [a] majority white racist society who makes an extremely subtle, veiled speech about inherent differences between racial

groups, one that fails to arouse emotions, feelings, or attitudes of hate or hatred against the speech. We might yet call this ‘hate speech’ even if people do not hate it. The point is that hearer reactions, including hate or hatred, are not always reliable guides to how we should use the term ‘hate speech’ (Brown, 2017a: 448).

The same goes for applying limiting conditions on the lexical content or the intention of the speaker. One may intend harm yet be unsuccessful in causing it, as is the case with UHS explored above in section 2.3.2, or one may produce language that has the locutionary force of hate, like the slur “tranny”, for example, but with it intend and cause no harm at all, as is ostensibly the case on /lgbt/. Considering this, I think a “family resemblances” view of hate speech, as argued by Brown (2017a; 2017b), to best represent what I mean by the phenomenon. That is to say, “there are particular qualities, features, or descriptive properties that can be found in some or even in many examples of hate speech” but “these are not always present, and are not necessary conditions” for the concept of hate speech (Brown, 2017b: 594). As pointed out by Wittgenstein (1953: 35), much like the semantic concept of a “game” can be taken to mean an array of things with no common characteristics, the *hate* in hate speech is a category which we recognise as a category without necessary common characteristics. To find necessary and jointly sufficient conditions of a *game*, and therefore understand how it operates and how to engage with it, one must further define it by additional modifiers, such as *board* as in *board game*, where a necessary condition is that it is a game played on a playing surface or a board. And like the word *love*, for example, can be understood to denote filial, erotic, platonic or sado-masochistic love and even the love of one’s favourite snack, *hate* in hate speech can denote out-group directed or in-group directed, overtly or covertly, passionately or dispassionately, et cetera, expressed speech acts. Thus, in this study, I have argued for a subcategory of hate speech that both intends harm and lexically reflects this intention, yet which is perceived as harmless or genre-appropriate by its target audience, and where an anticipatory relationship to this hostility is integral, in an effort to further illustrate the “stunning heterogeneity of the phenomenon” (Brown, 2017b: 565).

## 6. Conclusion

In this study I have aimed to situate the 4chan subforum /lgbt/ within the broader frame of hostile online discourse, by examining a representative corpus of over 10,000

comments. The first and second research questions regarding the variation and frequency of the transgender lexicon of /lgbt/ is answered by the corpus data gathered for the study. The corpus data reveals that, while general words like “trans” and “tranny” are the most frequently used lemmas used to refer to transgender people, genre-specific terminology, like “AGP” and “hon”, are, not only very hostile in their communicative value, but also increasingly salient as time goes on. Indeed, genre-specific terminology with hostile origins and uses are increasingly preferred over general and neutral terminology. The collocational discourse analysis applied in this study provided further answers to research questions 3 and 4 regarding the attitudes of /lgbt/ towards transgender people, and the conceptual question of whether the language of /lgbt/ could be considered hate speech. From the collocates of the genre-specific term “hon”, and the etymology of the term “AGP”, one can determine that a function of the genre of /lgbt/ is to produce and propagate language that associates sexual deviancy, degeneracy and perversion with transgender people, particularly transgender women. A speech act theoretical pragmatic examination of the hostile exchanges between members of the DC, not only revealed a degree of nonchalance to such violent speech acts as imperatives to suicide, but also an acute awareness and an anticipatory relationship towards this type of language. This poses a question mark to previous conceptual paradigms of hate speech, outlined in section 2.2, that centralise harm causation as a necessary condition of hate speech. If violent language, like the examples presented in section 4, is anticipated, or even tolerated by its target audience, it cannot be harmful insofar as it would count as hate speech. Yet, if violent language like imperatives to suicide on the basis of one’s gender identity, a protected characteristic, is not hate speech, then what is?

It should be noted that nothing I have said thus far is intended to be taken to mean that approaches centring the target audience of hate and their responses to violent language are bad approaches to hate speech research. On the contrary! In this study, I have aimed to shed light to a subcategory of hate speech that is particularly problematic to the perceptual approach to demonstrate the complexity of the phenomenon and the tremendous task of defining hate speech. However, this is not to say that the perceptual approach cannot meaningfully study the phenomenon. In fact, for future research, the questions left unanswered by this study would be best answered, precisely, by perceptual approaches measuring the responses of the targeted audience. Questions such as; how do transgender people outside of the DC of /lgbt/

respond to IGHS and terms like “hon” and “AGP”? Or, how does genre and membership to a discourse community influence our responses to hate speech? Or even, how does it feel to hear such language as displayed on /lgbt/, as a transgender person? These questions, not only benefit from, but require input from the targeted community and no doubt would further illuminate what we mean with the term “hate speech”. No amount of discourse analysis, corpus linguistics or conceptual analysis of hate speech can answer these questions. My interest here has been to contribute to the vast field of research on hate speech and explore some of the speech act theoretical dynamics of the fuzzy borders of hate speech, add nuance where necessary and simplify categories of hate speech where possible.

## References

- 4chan (n.d.a). *What is 4chan?* <https://4chan.org/>.
- 4chan (n.d.b). *Rules*. <https://www.4chan.org/rules#lgbt>.
- 4chanarchives (n.d.) *4chanarchives*. <https://4chanarchives.com/>.
- Allwood, Jens. (1977). A critical look at speech act theory. *Logic, pragmatics and grammar*, 53–99.
- Anderson, Luvell, & Barnes, Michael. (2022). Hate speech.
- Anthony, Laurence. (2023). AntConc (Version 4.2.4) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>
- Austin, John. (1975). *How to do things with words* (Vol. 88). Oxford university press.
- Baider, Fabienne, Assimakopoulos, Stavros, & Millar, Sharon. (2017). Hate speech in the EU and the CONTACT project. In *Online hate speech in the European Union: A discourse-analytic perspective* (1–16). Springer.
- Bhat, Prashanth, & Klein, Ofra. (2020). Covert hate speech: White nationalists and dog whistle communication on twitter. *Twitter, the public sphere, and the chaos of online deliberation*, 151–172.
- Biker\_Hackerman [@Biker\_Hackerman]. (2020, November 13). they actually said the helicopter joke [Online forum post]. [https://www.reddit.com/r/onejoke/comments/jtnjjz/they\\_actually\\_said\\_the\\_helicopter\\_joke/](https://www.reddit.com/r/onejoke/comments/jtnjjz/they_actually_said_the_helicopter_joke/)
- Bishop, Jonathan. (2014). Trolling for the lulz?: using media theory to understand transgressive humour and other internet trolling in online communities. In *Transforming politics and policy in the digital age* (155–172). IGI Global.
- Bratman, Michael. (1984). Two faces of intention. *The Philosophical Review*, 93(3), 375–405.
- Brown, Alexander. (2017a). What is hate speech? Part 1: The myth of hate. *Law and Philosophy*, 36, 419–468.

- Brown, Alexander. (2017b). What is hate speech? Part 2: Family resemblances. *Law and Philosophy*, 36, 561–613.
- Burnap, Pete, & Williams, Matthew. (2016). Us and them: identifying cyber hate on Twitter across multiple protected characteristics. *EPJ Data science*, 5, 1–15.
- Butler, Judith. (1997). *Excitable speech: a politics of the performative / Judith Butler*. Routledge.
- Culpeper, Jonathan. (2021). Impoliteness and hate speech: Compare and contrast. *Journal of pragmatics*, 179, 4–11.
- ElSherief, Mai, Ziems, Caleb, Muchlinski, David, Anupindi, Vaishnavi, Seybolt, Jordyn, De Choudhury, Munmun, & Yang, Diyi. (2021). Latent hatred: A benchmark for understanding implicit hate speech.
- Gabriel, Yiannis. (1998). An introduction to the social psychology of insults in organizations. *Human relations*, 51(11), 1329–1354.
- George, Carlisle, & Scerri, Jackie. (2007). Web 2.0 and User-Generated Content: legal challenges in the new frontier. *Journal of Information, Law and Technology*, 2.
- Herz, Michael, & Molnar, Peter. (Eds.). (2012). Interview with Robert Post. In *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (11–36). chapter, Cambridge: Cambridge University Press.
- Langton, Rae. (2018a). Blocking as counter-speech. *New work on speech acts*, 144–156.
- Langton, Rae. (2018b). The authority of hate speech. *Oxford studies in philosophy of law*, 3(1997), 123–152.
- Leets, Laura. (2002). Experiencing hate speech: Perceptions and responses to anti-semitism and antigay speech. *Journal of social issues*, 58(2), 341–361.
- Liimatta, Aatu. (2019). Exploring register variation on Reddit: A multi-dimensional study of language use on a social media website. *Register Studies*, 1(2), 269–295.
- Jeshion, Robin. (2020). Pride and prejudiced: On the reclamation of slurs. *Grazer Philosophische Studien*, 97(1), 106–137.

- Karjo2000 [@Karjo2000]. (2020, November 14). It gets funnier and funnier every time they repeat it. Every time [Online forum post]. [https://www.reddit.com/r/onejoke/comments/jtnjjz/comment/gc7pw7i/?utm\\_source=share&utm\\_medium=web3x&utm\\_name=web3xcss&utm\\_term=1&utm\\_content=share\\_button](https://www.reddit.com/r/onejoke/comments/jtnjjz/comment/gc7pw7i/?utm_source=share&utm_medium=web3x&utm_name=web3xcss&utm_term=1&utm_content=share_button)
- Kurrek, Jana, Saleem, Haji, & Ruths, Derek. (2020). Towards a comprehensive taxonomy and large-scale annotated corpus for online slur usage. In *Proceedings of the Fourth Workshop on Online Abuse and Harms* (138–149).
- Massanari, Adrienne. (2017). #Gamergate and The Fappening: How Reddit's algorithm, governance, and culture support toxic technocultures. *New media & society*, 19(3), 329–346.
- Matsuda, Marie. (1989). Public Response to Racist Speech: Considering the Victim's Story. *Michigan Law Review*, 87(8), 2320–2381. <https://doi.org/10.2307/1289306>
- Moser, Charles. (2010). Blanchard's autogynephilia theory: A critique. *Journal of homosexuality*, 57(6), 790–809.
- Mouse9001 [@mouse9001]. (2023, February 10). I think "passoid" just means a trans person who passes ridiculously well. [Online forum post]. [https://www.reddit.com/r/asktransgender/comments/10y8h88/comment/j7xsdtv/?utm\\_source=share&utm\\_medium=web3x&utm\\_name=web3xcss&utm\\_term=1&utm\\_content=share\\_button](https://www.reddit.com/r/asktransgender/comments/10y8h88/comment/j7xsdtv/?utm_source=share&utm_medium=web3x&utm_name=web3xcss&utm_term=1&utm_content=share_button)
- Oxford University Press. (2023). “flamer (n.), additional sense,” In *Oxford English Dictionary*. Retrieved April 2024. <https://doi.org/10.1093/OED/8665893740>.
- Parvaresh, Vahid. (2023). Covertly communicated hate speech: A corpus-assisted pragmatic study. *Journal of Pragmatics*, 205, 63–77.
- PragerU [@prageru]. (2023a, November 3). Men should not wear women's clothing. [Tweet]. Twitter. <https://twitter.com/prageru/status/1720493135082745969>
- PragerU [@prageru]. (2023b, November 3). 🗑️ [Tweet]. Twitter. <https://twitter.com/prageru/status/1720521294985490862>

- Puglia, David. (2015). *Hon Culture Wars: Revering and Reviling the Vernacular in Baltimore and Beyond*. The Pennsylvania State University.
- Reddit (n.d.a). r/onejoke. <https://www.reddit.com/r/onejoke/>.
- Reddit (n.d.b). r/asktransgender. <https://www.reddit.com/r/asktransgender/>.
- Ribeiro, Manoel, Blackburn, Jeremy, Bradlyn, Barry, De Cristofaro, Emiliano, Stringhini, Gianluca, Long, Summer, Greenberg, Stephanie, & Zannettou, Savvas. (2020). From Pick-Up Artists to Incels: A data-driven sketch of the manosphere.
- Rieger, Diana, Kümpel, Anna, Wich, Maximilian, Kiening, Toni, & Groh, Georg. (2021). Assessing the extent and types of hate speech in fringe communities: A case study of alt-right communities on 8chan, 4chan, and Reddit. *Social Media+ Society*, 7(4).
- Sellars, Andrew. (2016). Defining hate speech. *Berkman Klein Center Research Publication*, (2016-20), 16–48.
- Simpson, Robert. (2013). Dignity, harm, and hate speech. *Law and Philosophy*, 32(6), 701–728.
- Stănescu, Vasile. (2018). ‘White power milk’: milk, dietary racism, and the ‘alt-right’. *Animal Studies Journal*, 7(2), 103–128.
- Tardy, Christine, & Swales, John. (2014). 6. Genre analysis. In K. Schneider & A. Barron (Ed.), *Pragmatics of Discourse* (165–188). Berlin, München, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783110214406-007>
- Teh, Phoey, Cheng, Chibin, & Chee, Weng. (2018). Identifying and categorising profane words in hate speech. In *Proceedings of the 2nd International Conference on Compute and Data Analysis* (65–69).
- Wittgenstein, Ludwig (1953). *Philosophical Investigations*. New York, USA: Wiley-Blackwell. Edited by Gertrude Anscombe.
- Zelenkauskaitė, Asta, Toivanen, Pihla, Huhtamäki, Jukka, & Valaskivi, Katja. (2020). Shades of hatred online: 4chan duplicate circulation surge during hybrid media events. *First Monday*, 26(1). <https://doi.org/10.5210/fm.v26i1.11075>

- Zhu, Yi, & Bresnahan, Mary. (2018). Collective face, politeness strategies, and discomfort: Communication of American domestic students and Chinese international students. *Journal of Intercultural Communication Research*, 47(2), 141–159.
- Åkerlund, Mathilda. (2022). Dog whistling far-right code words: the case of ‘culture enricher’ on the Swedish web. *Information, Communication & Society*, 25(12), 1808–1825.
- Walters, Mark. (2014). Conceptualizing ‘hostility’ for hate crime law: Minding ‘the Minutiae’ when interpreting Section 28 (1)(a) of the Crime and Disorder Act 1998. *Oxford Journal of Legal Studies*, 34(1), 47–74.