

UC Santa Barbara

UC Santa Barbara Electronic Theses and Dissertations

Title

The politics of community language change: a computational analysis of language norms in an online trans community

Permalink

<https://escholarship.org/uc/item/45d035nh>

Author

Brown, Cedar

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Santa Barbara

The politics of community language change: a computational analysis of language norms in
an online trans community

A Thesis submitted in partial satisfaction of the
requirements for the degree Master of Arts
in Linguistics

by

Cedar E. M. S. Brown

Committee in charge:

Professor Lal Zimman, Chair

Professor Simon Todd

Professor Lina Hou

June 2022

The thesis of Cedar E. M. S. Brown is approved.

Simon Todd

Lina Hou

Lal Zimman, Committee Chair

June 2022

The politics of community language change: a computational analysis of language norms in
an online trans community

Copyright © 2022

by

Cedar E. M. S. Brown

Acknowledgements

As this work centers on English in a US context, it is important to acknowledge that English in this country has a colonial history and that the colonial state of the United States is located on the lands of many indigenous peoples. I acknowledge the traditional owners of the land on which I did this work, the Barbareño Chumash people and the Wurundjeri Woi Wurrung people. Additionally, when talking about the internet, it's important to keep in mind that the internet is not an abstract entity, but rather is made of physical components such as servers housed on indigenous land.

I want to thank my committee for their help and support throughout this project: Lal Zimman for his qualitative and conceptual support and for his work on our LSA conference presentation that presenting some of these ideas; Simon Todd for his investment and support in all aspects from the very inception of this project; and Lina Hou for her feedback and ideas throughout.

I also want to thank Will Hayworth for their assistance with accessing the corpus, as well as their feedback on my ideas. Additionally, the feedback from members of the TRILL lab, Simon Todd's lab, CEILing, and Sociocult was immensely helpful throughout. Particularly, extensive feedback from Chloe Willis, Marina Zhukova, and Chadi Ben Youssef helped me to flesh out my ideas and expand my thinking.

Additionally, I want to thank my friends and family. Particularly Kristy Ali, Julia Pelosi Thorpe, Meri Leeworthy, and Jalen Lyle-Holmes for allowing me to talk through ideas with

them and for emotionally supporting me throughout this process; and my gran, Vilma Schokman, for sharing space with me throughout the pandemic, during which a lot of this work took place.

I lastly want to thank the ftm LiveJournal community members for the use of their words in this work.

ABSTRACT

The politics of community language change: a computational analysis of language norms in
an online trans community

by

Cedar E. M. S. Brown

A perception that LGBTQ+ terminology is undergoing constant and rapid shift is often mobilized ideologically to justify resistance to transgender and/or non-binary people's language activism; however, there is little empirical work that explores this purported language change. How much is language use in queer and trans communities really changing, and how do changes occur? Additionally, there is criticism from within queer and trans communities that mainstream trans and queer spaces uphold and reinforce other hegemonic norms relating to whiteness, ableism, and classism. When language use in online queer and trans communities shifts, how might those with most power in the community influence the way that change unfolds? In this thesis, I explore these dynamics by examining shifting language patterns in an online trans LiveJournal community (TransLiveCorpus – Zimman & Hayworth 2020). Using bigram snapshot language models, I analyze community homogeneity over time, as well as identifying how top users pattern with respect to the overall community. This computational analysis is paired with qualitative analysis of how interactions of creating and enforcing language norms play out. Based on this, I argue that when approaching language change, we must consider how community language use exists within broader socio-political landscapes.

Table of Contents

Acknowledgements.....	iv
Contents	vii
1. Intro/background.....	1
2. Language change in online communities.....	5
3. Data.....	10
4. Lexical change	15
Lexical change: methods	15
‘Transgendered’ vs ‘transgender’: analysis of cohort change	16
‘Bio’ vs ‘cis’: analysis of cohort change	23
Lexical change discussion	29
5. Community change	30
Community change: methods	31
Community change results.....	34
6. Patterning of individuals.....	38
Top users: method.....	40
Top user patterns.....	41
Top user identities and behaviors	45
7. Discussion & conclusion	60
Future work.....	65
References.....	66
Appendix 1.....	74

1. Intro/background

The first time my gender made sense, I was reading a comment thread in a Facebook group. Sitting on the edge of my bed at 2am with my family in the next room, I fumbled through comments, eyes damp, feeling the lightness and ache of being seen. There are many parts of transition that are confusing, alienating, and uncomfortable. When someone was trying to get an appointment with the doctor I had just had a bad encounter with, I was able to detail to them what had happened. When I was thinking about a new medication, I was able to read others' histories with it. I have floated between different internet spaces over the years, sometimes reaching out for advice, sometimes making sense of my own experience through sharing it with others. I have seen some of my favorite groups collapse because a shift in membership meant that the moderators felt unable to handle the increasing white supremacy in the comments. I have followed threads about the way people are deploying language that end with everyone involved getting blocked by the moderators. I have noticed differences in what language is being used in nonbinary vs transmasc vs general trans spaces, in those that are location-based vs those that focus on a particular topic. These experiences have led me to this work.

Discussions on the internet wherein ideas and depictions of trans people are circulated play a large role in constructing an “imagined” trans community (Anderson 1983, Valentine 2007, Zimman 2017). There are a number of discourses that circulate surrounding the language of such a community. On one hand, there is a societal perception that LGBTQ+ terminology is undergoing rapid shift. This is often mobilized to justify resistance to transgender, gender nonnormative, and/or non-binary people's language activism – as people claim such

terminology is shifting unnaturally fast or in a way that is not reasonable to keep up with. On the other hand, mainstream LGBTQ+ rights activism champions inclusive language as a key site of resistance and activism against homophobia and transphobia – often campaigning for language reform within oppressive and violent institutions. This includes advocating for the avoidance of gendered language in police training (Ryan, 2022), and pushing for nonbinary options on state documents, such as the US government newly allowing *X* as a third gender option on passports (Hernandez, 2022). Both of these discourses are premised on the idea that linguistic change and social change go hand-in-hand. In a way, they also agree that current language is insufficient to reflect the complexities of how gender is understood by many today – in the push for the expansion of language by mainstream LGBTQ+ rights activism and in the invalidation of gender complexity on linguistic grounds by the resistant.

Where such discussions draw upon an idea of a LGBTQ+, queer, trans, or other such community, people orient both to “imagined” LGBTQ+/queer/trans communities and to more local LGBTQ+/queer/trans communities of practice such as those that exist in physical space (Jones, 2014) as well as in online space (Leuckert, 2020). Given this, it is important to think about how people are orienting to the imagined community, even while looking at specific communities of practice (Jones, 2014).

Online trans communities of practice can be life saving for marginalized or isolated queer people. They can provide “counterpublics and care structures” to support trans people in navigating a transphobic society (Cavalcante, 2016). This can take the form of finding social support in connecting around mundane trans experiences (Jackson et al., 2018), being a sounding board about medical advice received from gatekeeping or clueless doctors, as well as being a space to navigate and understand the self (Byron et al., 2019).

Within such communities, however, there is often much internal language politics (Dame, 2016; Jacobsen et al., 2022; Miller, 2019). One aspect of this is a criticism that mainstream trans and queer spaces uphold and reinforce other hegemonic norms such as whiteness, ableism, and classism (Cohen, 1997; Jones, 2014). Another aspect is disagreement about the utility of mobilizing and organizing under different identity terms (Bassichis et al., 2011). There are also disagreements about ‘policing’ when it comes to language (Dame, 2016).

In thinking about the dynamics and language use in online trans communities, and how this is in broader conversation with ideas of trans language change, this thesis asks:

1. How do changes in trans terminology play out in online communities?
2. How much is language use changing in general in online trans communities, above and beyond cases of specific terminology?
3. What role do individuals play in community language and what might this mean for how change unfolds?

In Section 2, I discuss previous work on language use and change in online communities.

Section 3 introduces the TransLive corpus that will form the basis of my analysis. In Section 4, I explore changes in trans terminology by looking at two pairs of terms identified as changing in previous research: *transgender/transgendered* (as terms for people who do not identify with the gender assigned to them at birth) and *bio/cis* (as ways of referring to people who do identify with their assigned gender at birth), both of which are potentially contentious forms of variation in this community. Previous work on non-politicized terms in online communities have found that cohorts of users (those who join a community within the same time period) generally maintain the variable use of lexical alternants from that period, even

as the use of alternants changes in the broader population over time. In the TransLive corpus, I find that users do not maintain the variable use of lexical alternants that was present in the community at the time they joined it, but rather participate in ongoing change from using one alternant to another (for a taxonomy of change types, see Sankoff & Blondeau, 2007). That is, users who join the community during different periods all make the terminological changes seen in the community overall, which I argue is in part due to the prevalence of metalinguistic commentary on terminology. I then move to looking at language change beyond the lexical level in Section 5, investigating whether this penchant for community-wide change on key terminology plays out in the development of wider language norms. I take a quantitative approach to this question, training predictive language models to calculate community language variability over time. I find that in the initial period under study, while the membership of the community was increasing, language usage remained highly variable, with no clear norms forming. However, as users began to leave the community, more convergence around norms arose. In order to understand these dynamics further, in Section 6, I investigate potential effects of more dominant users. In looking at user distribution, I find that a heavy majority of users make below 500 entries, while six users emerge as the most prolific, each making over 1500 contributions (i.e., posts or comments on posts). Informed by usage-based theories of language (e.g., Bybee & Hopper, 2001), I take prolificness as a proxy for influence, and explore the quantitative patterning and the interactional strategies of these six users. In doing so, I find that as community norms develop, the language of these six users increasingly becomes more similar to future language use, suggesting that their language increasingly forms a model for future usage. Looking qualitatively at the interactions of these users in top threads and in metalinguistic discussion, I find patterns

regarding the deployment of localized power, the holding of privileged identities, and adherence to a transnormative view of transness. Thus, in its innovative integration of both computational and qualitative methods, this thesis presents a complex and multifaceted image of language change in an online trans community of practice, where self-aware politically-grounded metalinguistic commentary, dynamics of membership change, and the interactional effect of prolific posters all contribute to the way that change unfolds and is perceived.

2. Language change in online communities

In this section, I turn from a discussion of trans language to a discussion of how previous work has approached language change online through examining lexical change, community size and dynamics, interaction, and user lifecycle. In doing so, I ground the approach I take to looking at online trans community language use in the approaches from previous studies.

Several studies identify and predict the diffusion of lexical innovations across a host of online communities on platforms such as Twitter and Reddit (Cole et al. 2017, Kershaw et al 2015, Kershaw et al. 2017, Tredici and Fernandez 2018). For example, Tredici and Fernandez (2018) explore the effect of social networks within online communities on the uptake of lexical innovations. They look across 20 subreddits that span a range of topics such as r/cars and r/apple. They build on Milroy (1987) in quantifying the *tie strength* between users and assessing how the introduction and uptake of an innovative term relies on whether those who use it are strong- or weak-tie users. Across the communities, they consistently find that weak-tie users who occupy a central place in the community are the ones to introduce new terms, but that strong-tie users who are arranged in smaller subgroups or cliques are key

in spreading these innovations. While studies such as this analyze the use of lexical innovations generally – as new words that enter the community’s lexicon – they do not specifically look at politicized or high profile lexical change, such as that which is the subject of discourse surrounding and within queer and trans communities.

Another aspect of online communities that has received attention is the dynamics of community size and membership. Lin et al. (2017) find that, contrary to what is assumed in many community discourses, Reddit communities did not suffer substantial changes in the topics of posts or a decrease in user satisfaction (quantified in terms of complaint words and upvotes) after massive growth and that strong moderation (quantified in terms of number of deleted comments) helps the community maintain its linguistic identity. They did find, however, a change in interaction pattern after massive growth, where more contributions are clustered around a smaller number of posts. Mensah et al. (2020) look at interaction styles in growing and failing communities. They operationalize “interaction styles” as both the duration of a conversation and patterns in member activity during the first, middle, and last third of a conversation. They find similar interactional patterns regardless of whether a community is increasing or decreasing in size: that most members participate in the early part of a conversation; that the most active members over time participate in the middle part of a conversation; that conversations are shorter when there are more active users; and that positive conversations or those about leisure have fewer participants. This analysis of interaction and community size dynamics is interesting, but, by seeing “interaction” as only about duration and activity, it crucially fails to attend to how identity and power are present within interaction, shaping a community as it increases or decreases in size.

Several studies do attend to the way interaction unfolds on social media beyond the timing of user contributions. For example, Danescu-Niculescu-Mizil et al. (2011) find that speakers accommodate to their interlocutors' language on Twitter, and Pavalanathan & Eisenstein (2015) find that Twitter users style-shift depending on the size of their audience: using more non-standardized features when tweeting to a smaller audience. However, neither of these studies look at how interactional variation plays out within a specific community of practice.

While the above studies allow us to see interesting patterns within social media data and online communities broadly, by aggregating data from multiple varying communities, they do not pay close attention to the specificity and political dimensions of the communities under study.

The analysis in Danescu-Niculescu-Mizil et al. (2013) differs from computational analyses of multiple communities by focusing on a specific kind of community: online beer rating forums. Within two distinct communities of this type, they analyze user-level and community-level change. Within both communities, they find that the community develops norms over time, both in terms of the emergence of specific lexical conventions and in terms of decreases in the overall variability of language within the community. These changes occur even though community membership is constantly changing. They find that users adopt the lexical conventions (e.g., choosing to describe beer scent with either the term *smell* or *aroma*) from when they enter the community, with those who come into the community later adopting the more recent variant at higher rates. This is in line with the adult stability assumption or apparent-time hypothesis, where the community changes overall but where people retain the variants that were most common in their youth (Cukor-Avila & Bailey 2013, cf. Sankoff & Blondeau 2007). Danescu-Niculescu-Mizil et al. (2013) also look at user

lifecycle, examining the rate of lexical innovation adoption and how ahead or behind users are relative to the overall community across their time in the group. They find that users are initially ahead of the community's language change and adopt lexical innovations at a high rate, then fall behind it and adopt fewer lexical innovations as new cohorts join, until they leave the community. While this framework allows for the analysis of specific norms within a community, Danescu-Niculescu-Mizil et al. (2013) still do not attend to political or interactional elements of language change – even though changes they describe, such as that from *aroma* to *smell* are not without potentially politically charged indexical meanings by virtue of their class-based associations. It thus remains to be seen whether the same patterns apply in online trans communities, whose membership is more specifically political and identity focused.

Qualitative studies have also explored community language use and change online. Leuckert (2020)'s qualitative analysis of community in a *RuPaul's Drag Race* subreddit sheds light on how community members construct community membership through positive discussions of the community, through identifying with the broader *Drag Race* fan base by employing similar linguistic strategies to those used on the show, and through coining new words. Dayter & Rüdiger (2016) discuss how “pick up artists” (i.e., men who use specific seduction tactics to get women to sleep with them) use community-specific terminology and discourse frames on online forums to situate their narratives about hitting on women as understandable and successful within the community, orienting the reader to view their posts in a favorable light. Community-specific language in this case includes employing terminology from stereotypically masculine domains, such as military, science, and sports, as well as using semantic and grammatical shift to customize word meaning and creating new words. By

qualitatively looking at a single community, such work can attend deeply to the culture, interactions, and specific uses of language within. Attending even more specifically to interaction, Gordon & İkiizoğlu (2017) examine one thread on a food and nutrition discussion board, where a poster asks for advice on behalf of her boyfriend. Through a close analysis of a single thread, they are able to explore the way that the poster positions herself and how others take stances in relation to her and each other. Through this, they demonstrate the complex interactional potential of online forums. However, even such works do not necessarily attend to the dynamics of power within the communities they are looking at.

Qualitative studies looking at trans communities on Tumblr (Dame, 2016; Jacobsen et al., 2022) and YouTube (Miller, 2019) have explored the way that trans terminology is debated online. In looking at “tags” on Tumblr, Dame (2016) finds that trans identity tags formed a folksonomy, where users utilized specific tags as a form of self-identification and to create a sense of community. However, debate about who should use what language was prevalent, with posters claiming and contesting legitimacy. Jacobsen et al. (2022) further explore the concept of legitimacy in discussing the hostility between two groups of trans Tumblr posters, *transmedicalists* and *tucutes*. They find that users may uphold hegemonic norms through a range of discursive strategies, revealing trans Tumblr to be as much a site of contestation, as of community-building. Such qualitative work is able to explore the way meaning is negotiated through interaction, attending to the actual stakes of language use and its dynamic potential in constructing community membership. However, such qualitative analyses miss larger, more aggregate trends and dynamics within the community of study.

Combining qualitative and computational methods thus allows us a more holistic analyses of language use in online spaces. A notable paper that has begun such work is Dame-Griff

(2019)’s analysis the term *cisgender* on “Usenet”, an online discussion forum first established in the late 70’s. Through data visualizations and some qualitative analysis, Dame-Griff (2019) suggests that high frequency users had an impact on the way *cisgender* was being used, with their metalinguistic commentary being quoted by future users.

The current thesis builds on this work, bridging the gap between quantitative and qualitative analysis. It applies methods from Danescu-Niculescu-Mizil et al. (2013) for large-scale analysis of language norms, alongside closer interactional analysis, as in Leuckert (2020) and Gordon and İközöğlü (2017), within an online trans community.

3. Data

The data for the paper comes from the TransLiveCorpus (Zimman & Hayworth 2020a,b, Zimman 2014), which is a 23.6mil word corpus of interactions on public posts and comments from four trans communities on LiveJournal.com from 2000-2017.¹ Such identity-themed communities were common on this blogging site wherein users could make posts and then comment on the posts of others in threads. This forum-style interaction allows us to study language norms and how they are built through interaction. As described by Zimman and Hayworth (2020a), while LiveJournal is no longer much in use, it “was particularly popular among trans people in the 2000s, and many of the discourses and terminological norms that have risen to some prominence...can be seen in early form here” (p. 144). This makes it an interesting place to study the development of trans language norms. Further, analyzing it gives us a chance to look at the rise and then fall of an online community and to tease out

¹ The corpus is formatted for cloud storage in Google’s document database Cloud Firestore. For the purposes of this analysis, I reformatted it as a CSV file by executing SQL code through BigQuery.

some of the complexity around how platform use interacts with community-building online. The time period I am looking at is also interesting in that it was around the time of the birth of social media and in the transition from web 1.0 to web 2.0 (Fuchs et al., 2010). This transition from static pages to more community and social spaces meant a growth in platforms that were hosting trans-based groups, allowing trans people a new way to find community and connection outside local circles.

This paper focuses on the largest community in the corpus: “ftm”. *Ftm* (also *FtM* or *FTM*) is an abbreviated form of the term “female-to-male”, a term used by or for people who don’t identify with the female sex/gender they were assigned at birth, but rather identify more strongly or exclusively with maleness or masculinity. While it has previously been in common usage, it is now dispreferred by many individuals and communities in favor of “transmasculine” or “transmasc” for the perceived binarity and bioessentialism it encodes. The ftm LiveJournal community consists of around 17 million words made up by 20,000 posts, with between 0-290 comments each (11,382 total). It was characterized by users as the “largest and most visible” transmasc community on LiveJournal. This paper focuses on periods with more than 1,000 *entries* (combined posts and comments) per 3-month period: those posted between 2002 and 2011. In order to look at how these entries are distributed over the community’s life, Figure 1 shows the number of entries over time. The x-axis spans the 18 years of the community, within which each dot represents one month. The y-axis shows the number of entries on a logarithmic scale for ease of comparison between comments and posts.

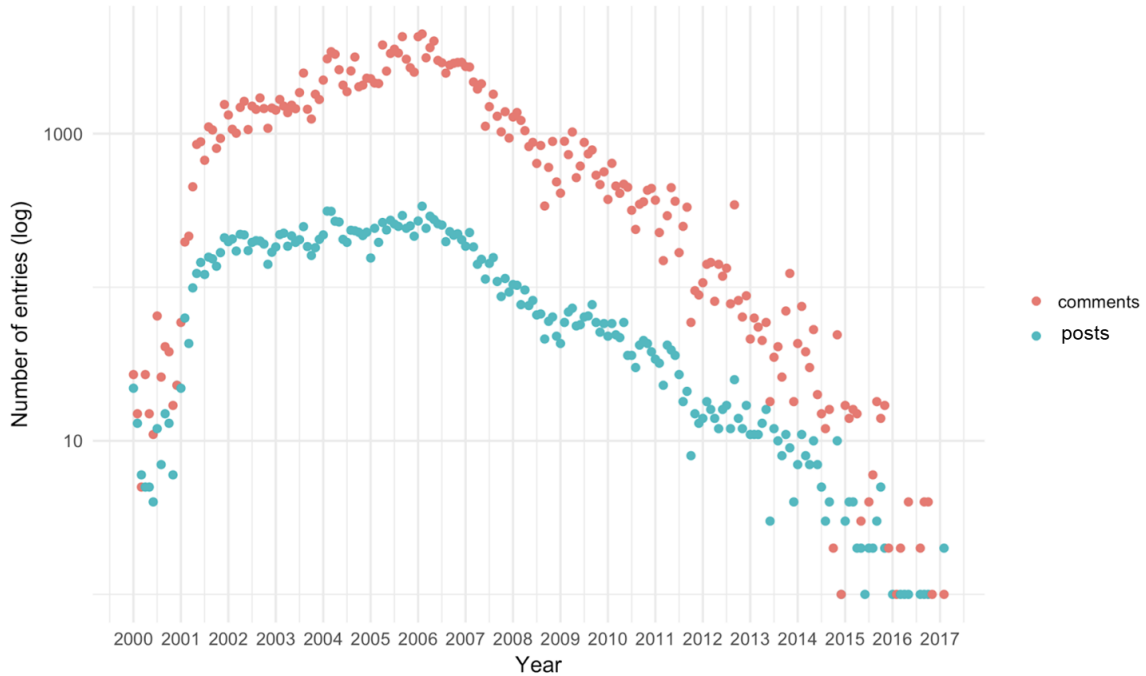


Figure 1. The number of comments (red dots) and posts (blue dots) across the life of the ftm LiveJournal community (fig from Zimman and Hayworth 2020a, p. 144). The number of entries increases to a peak at around 2006 until declining gradually through 2017.

As Figure 1 shows, the number of entries rose in the community, from its inception in 2000 until its peak in 2006. After this crucial point of change, where other social media companies such as Facebook and Twitter were emerging, the number of entries began to decline, dropping to almost zero by 2017.

While Figure 1 shows the overall traffic of the community, it does not indicate whether changes in activity were due to changes in the number of members or just in the frequency with which members were posting. Figure 2 adds to this picture by displaying the number of users active in each period during the period analyzed in the present study (i.e., 2002-2011). Here, the x-axis represents 3-month periods between 2002 and 2011, while the y-axis shows the total number of users who made entries in a given period. The number of active users gradually increased to a peak of around 1400 users in early 2007 and then decreased.

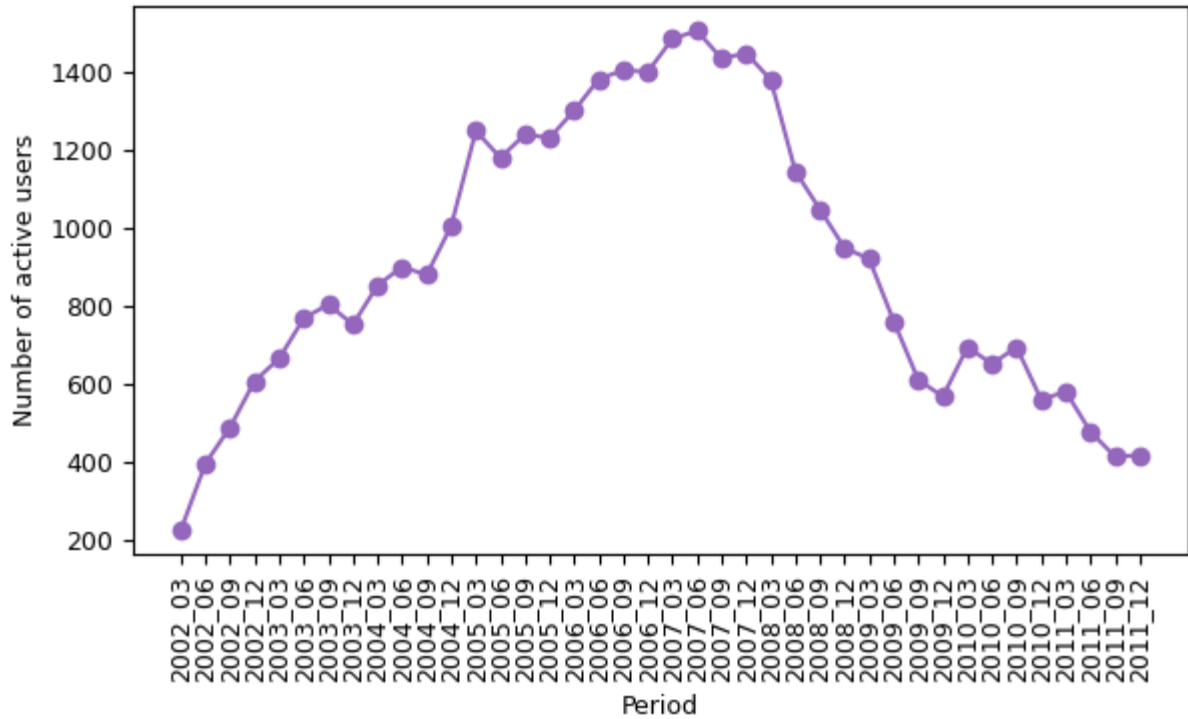


Figure 2. Number of active users over time. The number increased to a peak in early 2007, then decreased.

The patterning in the number of active users mirrors the patterning in the number of entries, where the number rose gradually to a peak in 2006/2007 before declining. This suggests that the gradual drop off in entries was due to a gradual reduction in the number of active users rather than being driven primarily by how frequently existing members are posting.

To understand more about the kinds of active users who were posting across the community’s lifespan, Figure 3 examines the percentage of total entries made by *new* users, or users who first posted within the 3-month period in focus. The number of entries by new users declined over time from 50% in 2002 down to under 10% in 2009 before becoming a bit more variable after 2010 when there were much fewer posts in general.

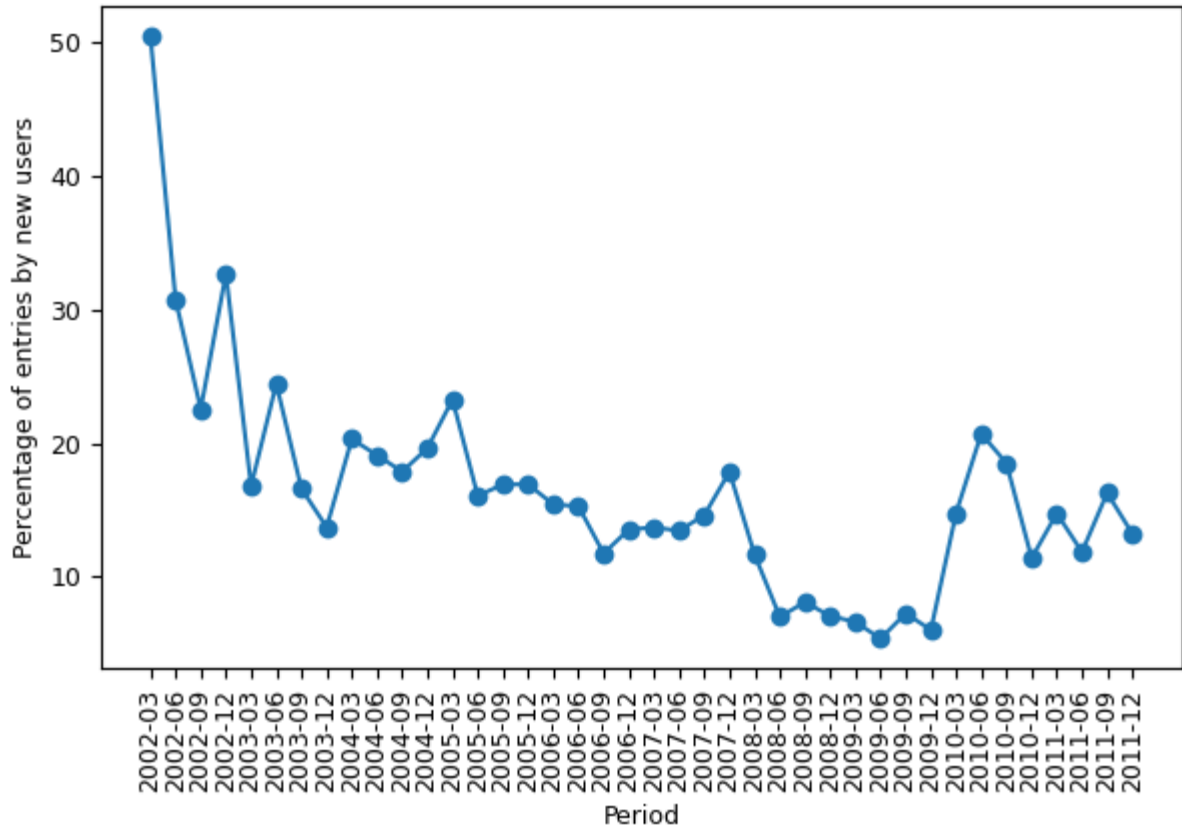


Figure 3. Percentage of total entries made by new users. The number of entries by new users declined over time.

In 2002, there was a high percentage of entries being made by new users, but this declined over time. This suggests that most of the community’s content was being created by those who had been in the community for more than three months, even as the total number of entries and active users increased and then declined.

Given that we see the population changing in size, and the number of interactions increasing and decreasing accordingly, does this mean that *language* in the community was also changing?

4. Lexical change

Given the highly politicized nature of the lexicon in discourses around trans language, I begin my analysis by focusing on lexical change. I seek to shed light on my first research question: how do changes in LGBTQ+ terminology play out in online communities?

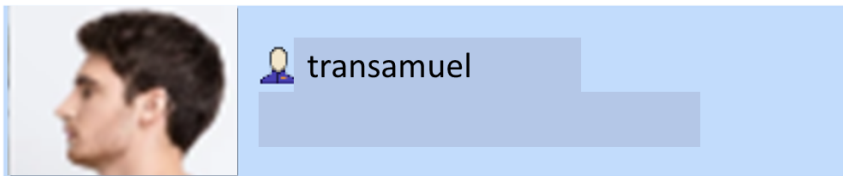
Zooming in to look at a couple of key terms allows us to investigate in detail the way that such terms were dispersed across the community as well as negotiated in discourse.

Lexical change: methods

As discussed earlier, Danescu-Niculescu-Mizil et al. (2013) found that when there was a community-wide lexical change, users retained the popular variant from when they arrived in the community – connecting to the adult language stability assumption, but operationalizing “lifetime” as the length of time a user is in the community rather than a person’s actual life. In order to investigate whether this version of the adult stability assumption occurred in this community, I look at how two key community-wide lexical changes identified by Zimman and Hayworth (2020a,b) unfold in different cohorts of users. I compare the percentage of uses of the terms under analysis for users who first posted in three cohorts: 2002, 2004, and 2006. This mirrors the two-year-apart cohorts used in Danescu-Niculescu-Mizil et al. (2013). While choosing cohorts is in one sense arbitrary, these cohorts were strategically chosen to have a large enough gap between them that one might expect differences to emerge, but one small enough that multiple cohorts can be observed over the years in question.

Transgendered vs transgender: analysis of cohort change

Trans identity terms were subject to a great deal of metalinguistic commentary over the life of the community. This took a range of forms – from longer explanations about their use, to shorter, more direct corrections. During the period under study, the *-ed* ending, as in *transgendered*, was being problematized for its implication that being transgender is something that has been done to someone, rather than something they are born as – and is now seen as offensive by many English-speaking trans people. In Example 1, when replying to a comment suggesting that the coinage of the term *transgender* was borne of anti-transsexual sentiment, the user, transamuel,² gives an explanation of its origin.



Actually, Virginia Prince coined the term ‘transgender’ to describe herself as she had no interest in bottom surgery and did not consider herself to have ‘transed the sex barrier.’ It caught on with other people who felt similarly. It wasn’t about separating herself from ‘those transsexual freaks.’

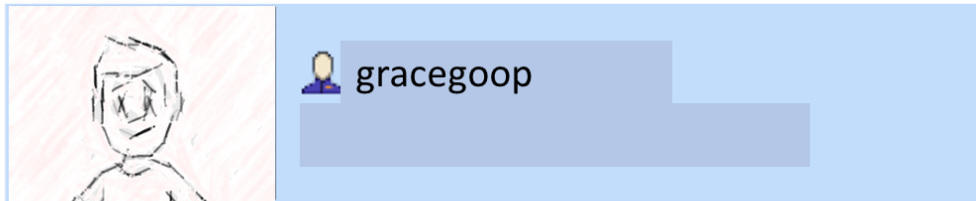
[Reply](#) [Parent](#) [Thread](#) [Collapse](#)  [Like](#)

Example 1.

transamuel aligns negatively with the previous position, using the fragments “actually” and “it wasn’t” to differentiate his perspective from that of the previous poster. This kind of discussion about who uses what terms and why was common throughout the community,

² Pseudonyms are given for users and their photos are replaced with stock photos that capture the essence of their profile

where users presented arguments and counterarguments to each other's linguistic choices and identifications. The interactional dynamism of this is suggested by Example 2, wherein gracegoop disaligns with the previous commenter and aligns with one further back, evaluating the term *transgender* as “horrible” and evoking standard language ideologies of grammatical correctness (Flores & Rosa, 2015).



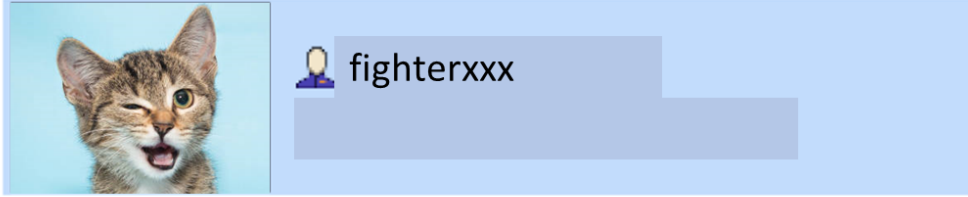
No, I am saying “transgender is horrible. I was agreeing with the commenter above who I thought I was replying to.

Transgendered is grammatically correct and sounds correct.

[Reply](#) [Parent](#) [Thread](#) [Collapse](#)  [Like](#)

Example 2.

In a later thread, fighterxxx corrects a previous poster’s use of the term *transgendered* in Example 3 by saying “Please don’t say transgendered,” and explaining their stance in terms of *-ed*’s potential resultative meaning.



Please don't say transgendered. No one was turned into anything, no one was "transgendered" by the transgender fairy. Transgender is appropriate. Thanks!

[Reply](#) [Parent](#) [Thread](#) [Collapse](#)  [Like](#)

Example 3.

fighterxxx rejects the use of *transgendered* through implicit analogy with words like *transformed*. They also mobilize ideologies of correctness and appropriateness.

Previous work with this corpus (Zimman & Hayworth 2020a,b) looks at the patterning of the lexical items *transgender* and *transgendered* over time, finding that the "plain form" (*transgender*) increases in frequency while the form with the *-ed* ending decreases. Figure 4 looks at the frequency of these two terms relative to their total combined use.

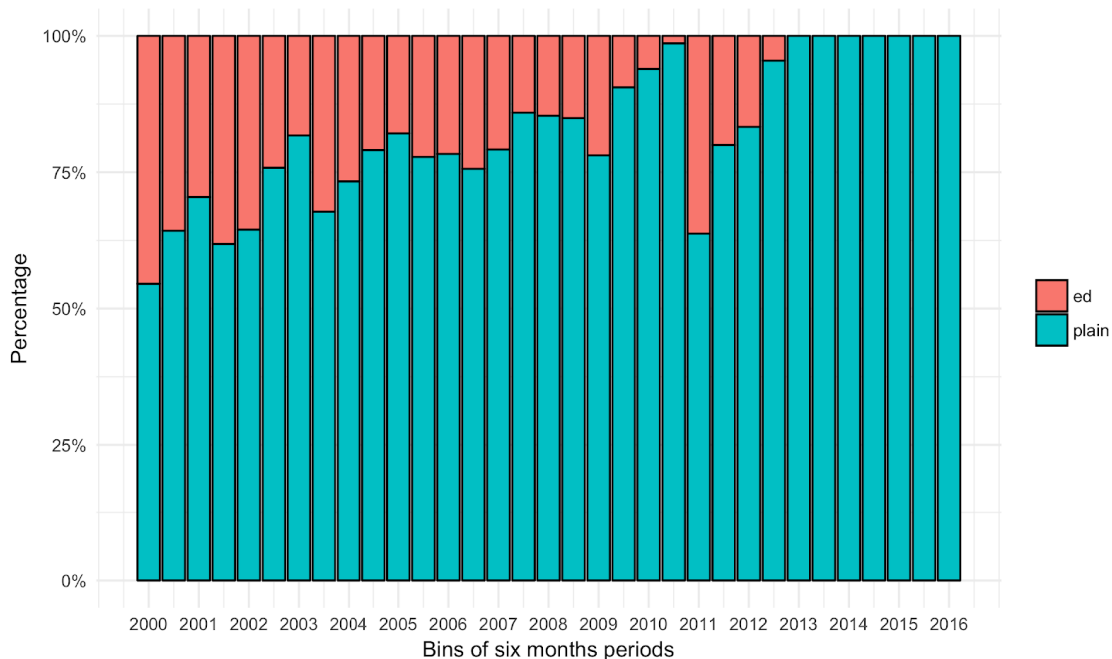
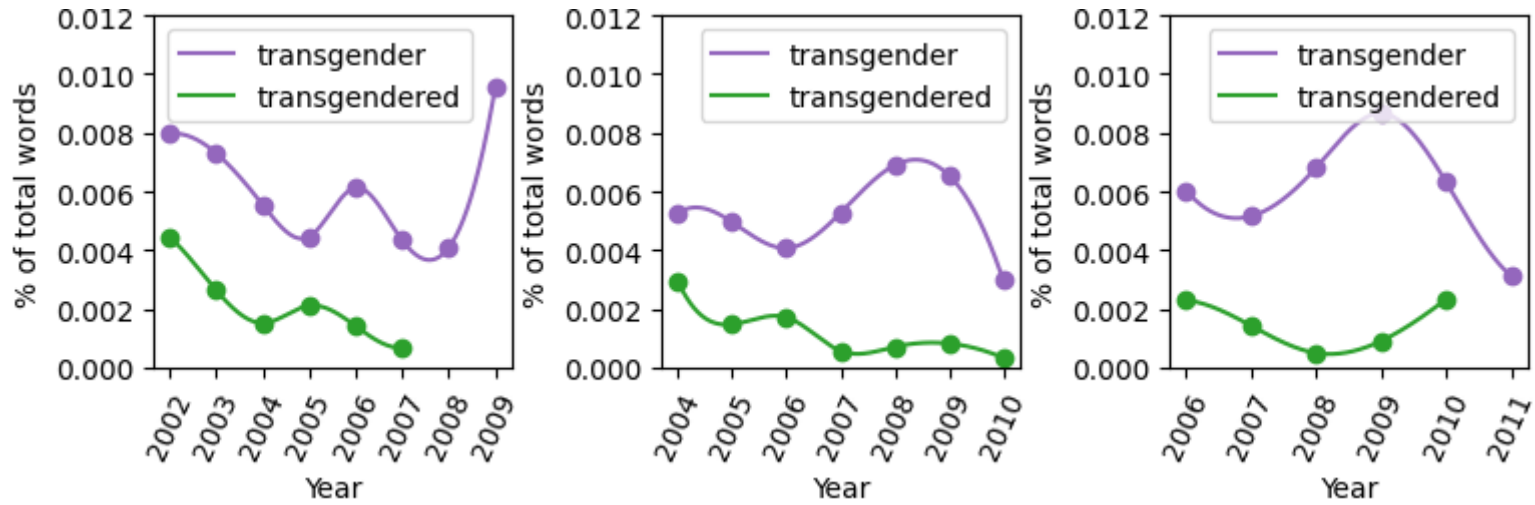


Figure 4. Relative frequency of terms *transgender* (blue – denoted as ‘plain’ in legend) and *transgendered* (red – denoted as ‘ed’ in legend) over time (fig from Zimman and Hayworth 2020b, p. 507). Over time, *transgender* replaces *transgendered* as the variant in use.

While at the beginning of the community *transgendered* was used just under 50% of the time, by the end of the community, *transgendered* was not used at all and *transgender* was used 100% of the time (though there were a lot fewer tokens in total). Thus, we see *transgender* replaces *transgendered* as the variant in use across the life of the community. Given the pattern identified by Danescu-Niculescu-Mizil et al. (2013), this finding raises a question about whether users in earlier cohorts maintain use of *transgendered* while later cohorts move toward *transgender*, or whether it was a community-wide change. The first hypothesis would suggest that this change was similar to a change in beer terminology such as *aroma* to *scent* and position this trans LiveJournal community as similar to the beer communities. The second hypothesis would suggest that something else is going on with this

terminology such that the whole community is shifting together – perhaps due to internal or external discourses.

In order to explore this, the set of figures below (Figures 5(a)-(c)) show the patterning of *transgender* with respect to *transgendered* over time. The x-axes span the time from when the cohort joined until the end of 2011. The y-axis gives the percentage of total cohort user words that the term in question makes up in the given time period (on a scale of 0.002%).



Figures 5(a)-(c). Uses of *transgender* (purple/higher line) vs *transgendered* (green/lower line) for users who first posted in 2002 (left), 2004 (middle) and 2006 (right). Uses of *transgendered* decrease over time in all cohorts, with *transgender* being the more common variant.

As can be seen from Figures 5(a)-(c), for all cohorts, *transgendered* is almost completely replaced by *transgender*. The 2002 cohort stop using the word *transgendered* the earliest, in 2007. The 2004 cohort start off using *transgendered* at half the rate of the 2002 cohort, and gradually decrease in use of that form (though they retain some use of it alongside *transgender* until that cohort stops using both words in 2010). The 2006 cohort also start off with a similar usage of *transgendered*, but their usage of it drops off while their use of *transgender* continues in 2011. In all the cohorts, we see the overall use of these terms fluctuating, depending on the posts and topics under discussion.

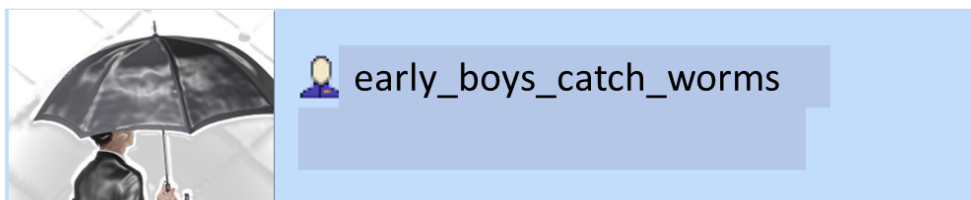
While there is some difference in the years that the change from *transgendered* to *transgender* happens between the cohorts, this difference does not line up with the hypothesis that people continue to pattern in terms of primarily employing the variant they used when they entered the community. To the contrary, those who enter the community in the earliest cohort make the change earliest. This could suggest that those in the earlier cohorts could be the ones driving the change, with the later cohorts taking time to acquire that norm. Those who are in the community longer have correspondingly longer exposure to and engagement with metalinguistic commentary on this terminology, encouraging them to shift.³

Given this surprising finding with respect to previous research, I will now turn to look at another change identified in Zimman and Hayworth (2020a,b) in order to see whether this result is limited to this one change or whether it could be a broader pattern.

³ I have not excluded the metalinguistic commentary from my quantitative analysis of the use of terms, as its use, even in a metalinguistic sense, still illustrates that this terminology was being circulated and discussed in the community at this time. Future analyses may remove all metalinguistic uses of such terminology.

Bio vs cis: analysis of cohort change

I now move to the community shift from *bio* to *cis*. These terms were used in the community to refer to people who are not trans, with *bio* being a shortened form of “biological” and *cis* being the antonym of *trans* and meaning “on the same side of.” They were used in a range of formulations, such as “cis-men”, “biomale”, and “cis guys”. Similar to the change from *transgendered* to *transgender*, there was a lot of metalinguistic discussion about the uses of these terms, with users getting into sometimes heated debates about them. Many users objected to the naturalization assumed in the term “bio” as well as its (problematic) assumption that the maleness of trans men is not biological. On the other hand, many users objected to the academicness of the indexical meanings, and thus the class elitism, of the term *cis*, as well as its lack of transparency to those who were unfamiliar with it. These positions can be seen in Examples 4 and 5, which are in response to a post that states “does the term cisgender bother anyone else as much as it annoys me?”



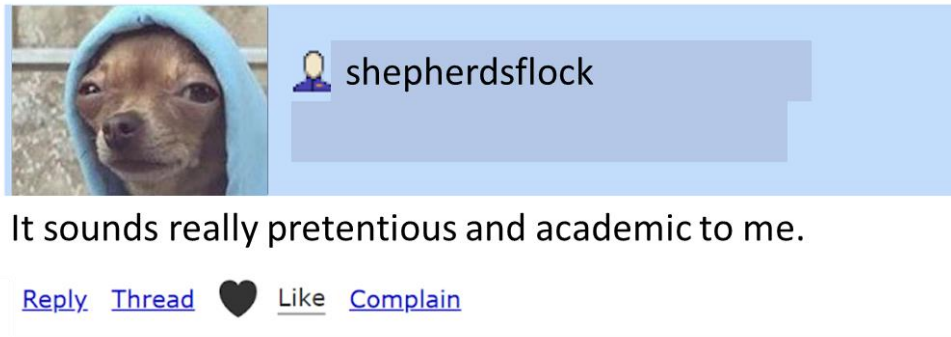
I like it. It doesn't set up the implications that “bio” guy does (like saying a bio guy is “normal” and a trans guy isn't)

[Reply](#) [Thread](#)  [Like](#) [Complain](#)

Example 4.

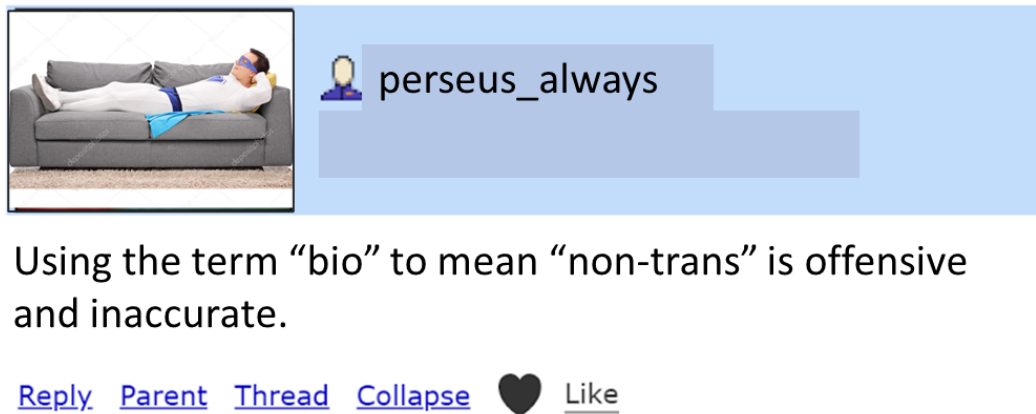
Example 4 expresses positive affect towards the term *cis* on the grounds that it does not create implications that a trans guy isn't “normal,” and thus disaligns with the original

poster's stance against *cis*. On the other hand, Example 5 below aligns with the original poster in evaluating the term *cis* as “really pretentious and academic.”



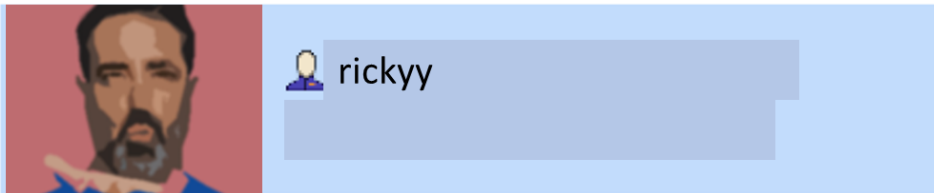
Example 5.

Metalinguistic discussion about terms for non-trans people also took the form of corrections, as can be seen in the Example 6. *perseus_always*, who was a moderator at the time, corrects a comment on a thread about standing to pee that uses the term *bio*, by evaluating the use of the term as “offensive and inaccurate.”



Example 6.

In Example 7 below, ricky characterizes the discussion of both *transgendered/transgender* and *bio/cis* as angry and divisive.



[...]

When someone posts something saying ‘blah blah blah bio-male, blah blah, transgendered bah blah,’ and then someone barges in and comments with ‘THOSE TERMS ARE OFFENSIVE TO EVERYONE SO SHUT UP YOU IGNORANT TWIT!’ then it becomes a Big Issue.

[...]

[Reply](#) [Parent](#) [Thread](#) [Collapse](#)  [Like](#)

Example 7.

rickyy portrays those making corrections as “barging in” and reports their supposed words using capitalization and insults (“IGNORANT TWIT!”) to frame them as unwarrantedly acrimonious. This kind of name calling and strong language is present in threads from people on both sides of these linguistic debates.

The shift from the common usage of *bio*, to metalinguistic discussion about *cis* vs *bio*, to corrections when someone uses the term *bio* parallels a shift in relative frequency from predominantly *bio* terminology to predominantly *cis* terminology across the community’s lifespan. Figure 6, from Zimman and Hayworth (2020b), shows the relative frequency of various terms for non-transgender people over the life of the community.

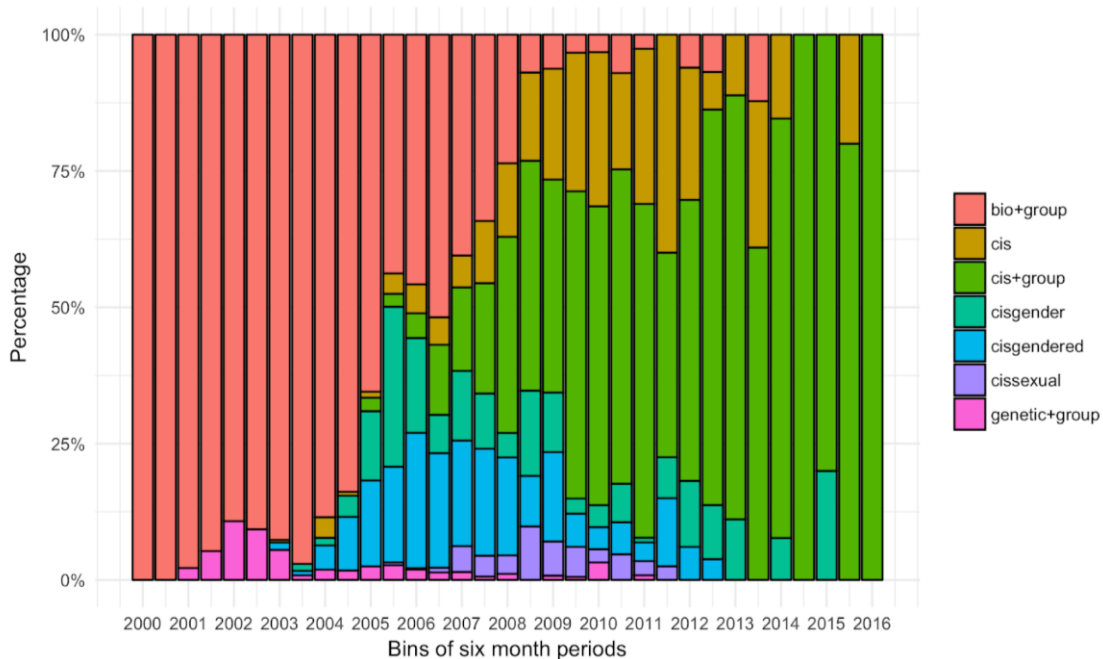


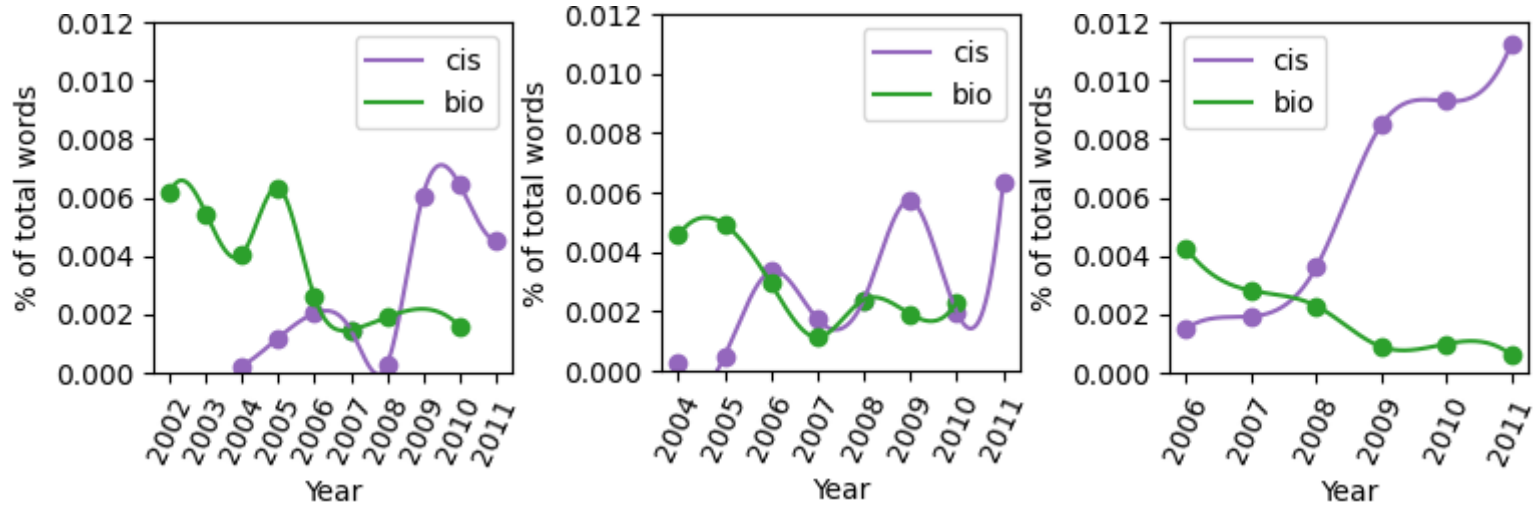
Figure 6. Relative frequency of terms for non-trans people over time (fig from Zimman and Hayworth 2020b, p. 509). *Cis* replaces *bio* as the variant of use across the community’s lifespan.

Zimman and Hayworth (2020b) find that, while originally the construction ‘*bio+group*’ (as in *bio men*) for non-transgender people was used around 90% of the time in 2002, this terminology shifted to be almost 100% either ‘*cis*’ or ‘*cis+group*’ by the end of 2011. Will this change follow the pattern of *transgender* to *transgendered* whereby all cohorts changed their use of terminology, or will these words follow the pattern seen in Danescu-Niculescu-Mizil et al. (2013), where cohorts stick with the variant most popular when they arrive?

As with *transgender/transgendered*, I examine the cohorts of 2002, 2004, and 2006, looking for their uses of the terms *cis* and *bio* across time.⁴ Figures 7(a)-(c) show both these terms as a percentage of total words. The x-axis spans the time from when the cohort joined until the end of 2011, divided into 6-month periods. The y-axis gives the percentage of total cohort

⁴ I used the regular expression pattern `(([^\a-z]^\a-z)$|((wo)?m(a|e)n|girl|boy|guy|(fe)?male))` to match variations of ‘*bio+group*’ as well as a similar one for ‘*cis+group*’.

user words in that period that the term in question makes up (on a scale of 0.002%). In each case, the users make a clear switch to using *cis* instead of *bio*, though the time and patterning of that change differs between groups.



Figures 7(a)-(c). Uses of *cis* (purple) vs *bio* (green) to refer to non-trans people for users who first posted in 2002 (left), 2004 (middle) and 2006 (right). All cohorts move from using *bio* to using *cis*.

Similar to the results for *transgender(ed)*, we see a deviation from the pattern found in the beer communities. The 2002 cohort start off exclusively using *bio* and begin using *cis* in 2004, which really takes off in 2008, completely replacing *bio* in 2010. Similarly, the 2004 cohort begin by using much more *bio* than *cis* in 2004, which shifts in 2006, when the two terms pattern similarly until 2008 where *cis* becomes more popular, replacing *bio* by 2010. The crossover period for the 2006 cohort is also in 2008, whereafter *cis* becomes much more popular as use of *bio* dwindles. Thus, for all cohorts, the term *cis* completely replaces *bio* as the variant of use at a similar time. Similar to the previous result, the previously attested tendency to stay with the original term is not present.

Lexical change discussion

As can be seen, in this data, the pattern of cohort-based lexical change in previous studies does not hold when looking at *transgendered* vs *transgender* or *bio* vs *cis* within this community. Instead, the way language change happens in regard to these politically-charged identity terms is through community-wide change, rather than the change being located within specific cohorts of users, meaning that the adult language stability assumption validated in Danescu-Niculescu-Mizil et al. (2013) is not seen. Rather, change is likely happening both in response to broader norms of the imagined community and to the plentiful metalinguistic discussion within the community of practice. Across the community, we see a gradual change from use of *transgendered* to *transgender* and from *bio* to *cis*; however, it is also one heralded by strong stances taken by individuals. These vehement arguments for language change could externally be perceived as rapid community-internal shift.

While the use of certain politicized terms is undergoing shift within the community, it is not yet clear whether the same is true of language use in general, beyond such terms. Is this surprising pattern of community-wide language change only applicable to language use that garners metalinguistic commentary, or does it show up in the community more broadly? One way of examining this could be to look at other pairs of words that are less politicized. However, conceptualizing which words are actually contrastive and comparing them with the more overtly political words is outside the scope of this thesis.⁵ Further, such an approach limits our analysis to lexical items. Instead, another way to move away from such self-aware terms and look at language change more broadly is to look beyond specific lexical variables and ask how the community's language as a whole is patterning, outside of only focusing on specific terminology.

5. Community change

In this section, I move to address my second research question: how much is language use changing in general in online trans communities, above and beyond cases of specific terminology? While political discussions often revolve around specific lexical items, as discussed in the previous section, an approach that only focuses on these terms fails to see the way they are embedded in context. Instead, here I take a more aggregate approach to language change by looking at the development of community norms.

⁵ Robustly identifying candidate pairs of words in a quantitative manner may require more data than is available in this corpus, due to the wide range of contexts in which words can be used. Furthermore, each candidate pair would have to be subjected to intensive manual scrutiny in order to determine whether the words are sufficiently conceptually related to be considered alternants.

Danescu-Niculescu-Mizil et al. (2013) found that within the two beer communities they looked at, community norms developed over time, with later posts using increasingly predictable language, suggesting increased group cohesion. If it is true that trans language is constantly changing, we could expect to see less convergence around norms in the ftm LiveJournal community than with the beer communities. However, if this community is similar to other communities that have a less overt set of political goals, we might expect to see community language becoming more similar over time.

Community change: methods

In order to look at community change over time, I trained bigram language models at different points along the community's development. A bigram language model is a representation of the probability of all one and two word sequences that occur in the data it is trained on.⁶ Such models could be used to generate text – such as how predictive text on a smartphone generates the possible next word as you type. However, they can also calculate a “cross-entropy” value that represents how likely the model would have been to generate a string of text that you feed it. The smaller the cross-entropy, the more likely the model would have been to generate the string of text, which means the more predictable from, and thus similar to, that string is to the data the model was trained on. In order to look at community language norms, I split the entries from the ftm LiveJournal community into training and test sets at different time periods. For each of these periods, I trained bigram language models on the training data and then ran the test sets through the models to get an overall measure of cross-entropy. In this way, *community norms* are operationalized as similar patterns of

⁶ Bigrams are used, as the limited amount of data makes use of larger n-grams untenable.

language use, such that the probability that the test set could be predicted by the training set is higher. When we see smaller cross-entropy, and thus greater self-similarity, this means that there is more convergence in both the words people are using and the way they are putting words together.

In this case, I trained snapshot bigram language models (LMs) for each 3-month period of the community's most active years using the SRILM toolkit (Stolcke, 2002).⁷ The LM for each 3-month period was created to give an approximation of what the community language norms were during this period. I adapted the method of obtaining test and training sets from Danescu-Niculescu-Mizil et al. (2013), where each LM was trained on a training set of 1000 posts and the corresponding test set was composed of all other entries from that time period. Further, each test and training set only included the first 30 words of each entry, as Danescu-Niculescu-Mizil et al. (2013) suggest that normalizing the post length removes length effects, as longer posts have larger cross-entropy.⁸ However, where they trained LMs for each month on training sets of 1000 posts made up from two posts from each of 500 randomly selected users, I trained each LM on a training set of 1000 randomly sampled entries (a combination of posts and comments). I made this alteration as the ftm LiveJournal corpus is smaller and many users only made one post in each period.

An issue with a random-sampling approach is that each time 1000 entries are randomly sampled, there is the chance that they could all be very similar in a way that doesn't represent

⁷ As in Danescu-Niculescu-Mizil et al. (2013), I used Laplace (additive) smoothing with parameter 0.2.

⁸ This approach does have some drawbacks, as innovations in later parts of the post are not part of the models, and posts less than 30 words are discarded. Additionally, the first 30 words sometimes contain quotes from previous posts that a user is responding to. In order to investigate whether the results are affected by this, future work should sample words from different parts of the posts and/or experiment with different numbers of words.

the broader community at that time. For example, they could by chance all be by the same user (which is avoided by the method in Danescu-Niculescu-Mizil et al., 2013) or about a similar topic. Thus, instead of a LM being an approximation of what everyone in the community is doing at that point, a LM trained on 1000 randomly sampled posts might end up being more an approximation of the language of two prominent users or the language people use to talk about, say, dating. In order to circumvent this, for each time period, I used a resampling-based approach by training 1000 LMs, each trained on a different 1000 randomly sampled entries. Results averaged across these LMs therefore give a clearer sense of the whole community, as they are not determined by a particular training set. For each time period, I created 1000 different training sets of 1000 entries each, with a corresponding 1000 test sets that included all the remaining community entries from that period. I fed the corresponding test set through each LM, getting cross-entropy measurements (which show the predictability of an entry's bigrams given the bigrams of the 1000 entries used to train the LM) for each entry contained therein.⁹ I then averaged these cross-entropy measurements to give a single measurement of community self-similarity for that LM. These measurements of community self-similarity were then averaged across all 1000 LMs trained for that time period, to obtain an unbiased estimate of community self-similarity for that time.

Figure 8 demonstrates this process for a single 3-month period with six example entries.

⁹ SRILM outputs a 'perplexity' value, which is converted to cross-entropy by taking its base-10 logarithm.

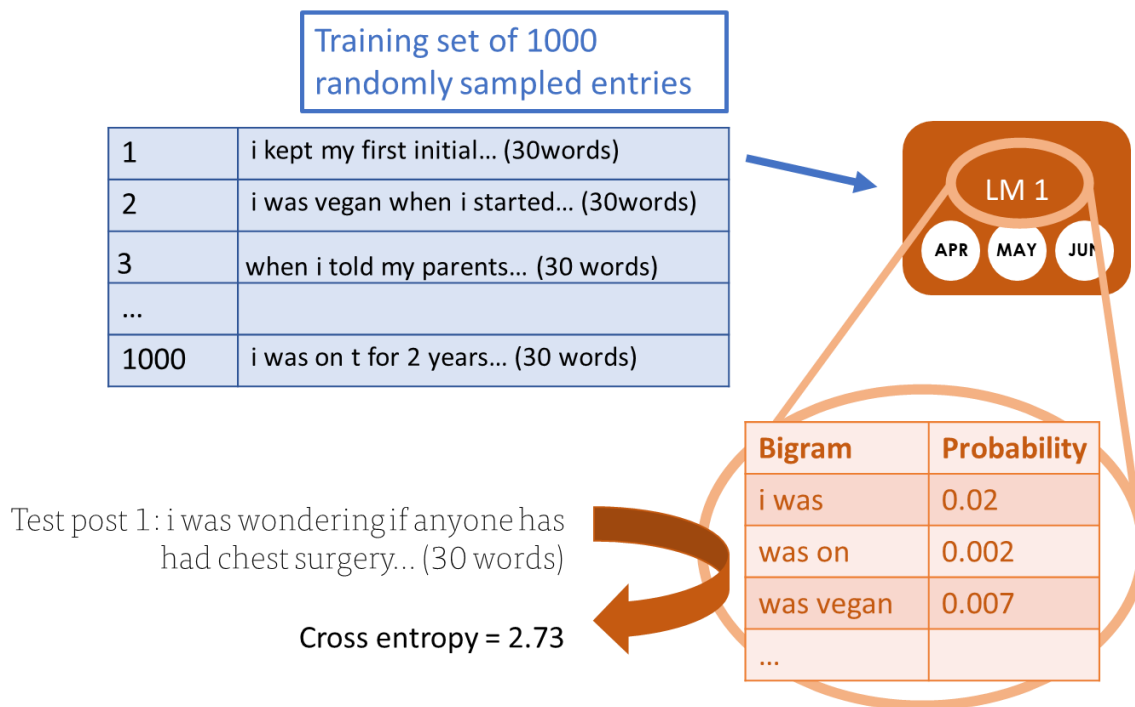


Figure 8. Example of how training entries determine the probabilities in a language model which are then used to get the cross-entropy of a test post, within one 3-month period. A diagram summarizing the full procedure is in Appendix 1.

Community change results

The convergence of community language over time can be visualized as a graph of cross-entropy over time. Given that the results from 1000 LMs for each period were used, Figure 9 presents both the mean (the plotted points and line) and a resampling-based 95% confidence interval for each time period, in order to validate the variability of the mean as a reasonable measure. Lower values mean the posts are easier to predict from the model and thus that the posts for that 3-month period are more cohesive. The x-axis spans the 9 years under study in 3-month intervals while the y-axis represents the cross-entropy for that period. The confidence intervals hug the mean until after late 2009, whereafter there are fewer posts so more variability is expected.

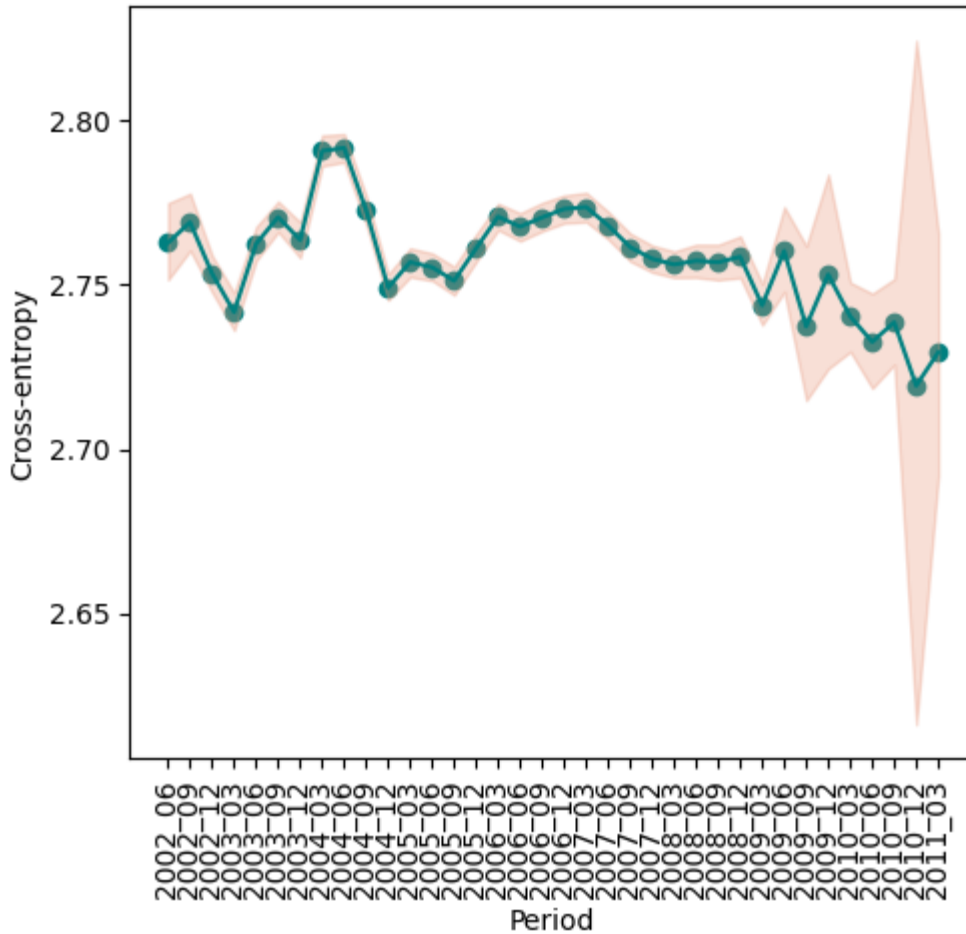


Figure 9. Predictability of language use in each 3-month period, calculated as the cross-entropy of a 3-month period’s entries when run through the language model of that period. The peach confidence intervals track along with the mean (line and points).

As we are looking at community-internal patterns, in these results, the most important thing is the way that cross-entropy changes across time. Given that the confidence intervals track closely along with the mean until the paucity of data after 2009 introduces more variability to the data, we can take the mean as reasonably representing the cross-entropy at each point. In order to more fully see the patterns in the mean, in Figure 10 the confidence intervals are removed to focus on the pattern of the mean over time.

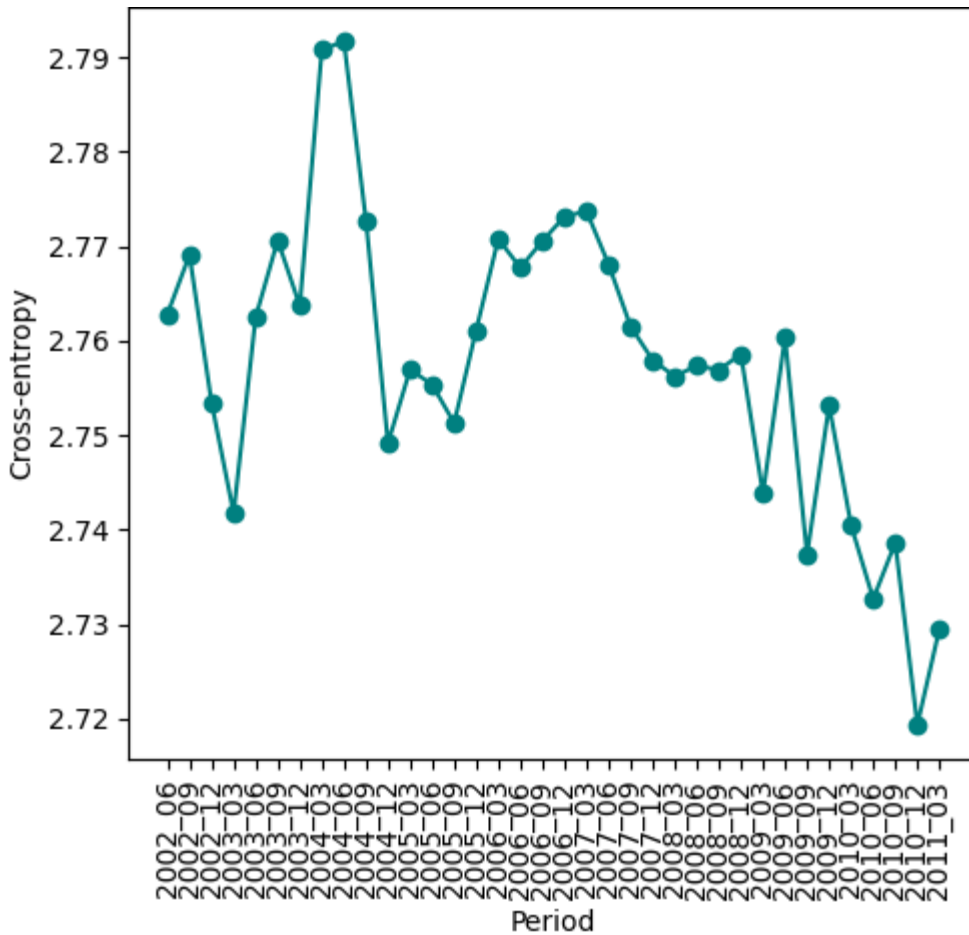
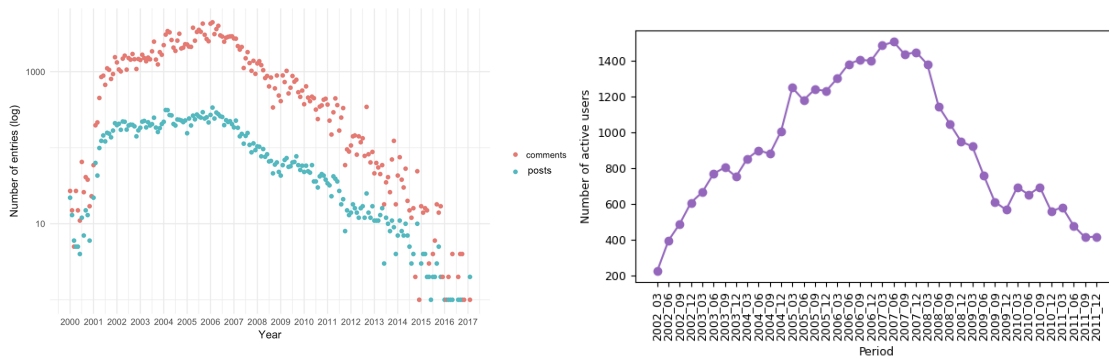


Figure 10. Predictability of language use in each 3-month period, as above, with only the mean. The mean is variable up until the end of 2006, whereafter it decreases. This means that after 2006, the language of the community is becoming more cohesive.

As can be seen from Figure 10, the cross-entropy is variable from 2002-2006: the community has greater self-similarity in 2003, while being more disparate in 2004, and then becoming more cohesive again in 2005, then less similar again in 2006. From this point, the cross-entropy trends downwards, which means that the community is becoming more similar to itself in the way language is being used. This timing maps onto where entries and users declined (as seen in Figures 1 and 2, repeated below).



Figures 1 and 2 (repeated): total number of posts and comments (left) and users (right).

As an increasing number of people joined and more comments and posts were made, the content in the community shifted around in how cohesive it was. However, after 2006, where entries and users began to drop off, the community's language started to become increasingly cohesive. In other words, while clear norms did not emerge in a linear fashion when there were many users, norms do develop as community participation drops off.

This raises a tension in how we can interpret the findings described above. On one hand, this result could be used to answer the research question by simply saying that language norms *do* develop in trans communities, and therefore trans language is not changing incomprehensibly fast – providing a counter argument to transphobic critics. However, we may not want to be so quick to do so; if we are observing increasing homogeneity as community engagement dwindles, the *norms* we are seeing could in themselves represent a problematically limited sample of trans discourse – i.e., one in which certain trans voices are amplified and others are silence. In this period of 2006-2007, platforms such as Twitter and Facebook were rising in popularity, and this was a period with a lot of changes in the way that transness was culturally being conceptualized: as more recognition of non-binariness saw the emergence of *nonbinary* as a term (Zimman & Hayworth 2020a) and binary-identified trans activists and writers were claiming authentic womanhood/manhood more strongly (Serano 2007). In this

context, who is leaving for other spaces and who is staying? This is important to think about because, as stated earlier, ftm was considered by some users as the “largest and most visible” transmasculine community on LiveJournal. In all likelihood, across the lifespan of the community, many non-“users” (i.e., cis and trans people alike who may lurk, or read the forum but not post to it) would have been observing and learning from content, developing an understanding of what transness is. This has implications for how the imagined community of trans people was being (re)constructed more broadly at the time. Where there is a greater diversity of perspectives expressed on the forum, the impressions of non-“users” would reflect this heterogeneity – and likewise where there is less diversity, their impressions would be more homogenous. A decrease in community membership along with increased homogeneity – and the norms around which this homogeneity is centered – thus could have implications wider than just shaping the language within this one community of practice. Being a public online community, and thus, in a way, an online source of information, ftm-internal language could affect, in a broader sense, how cis people conceived of trans people and how trans people understood their identity on personal and political levels.

6. Patterning of individuals

Questions of what is going on within the language of the community as homogeneity increased leads me, in this section, to consider what role individuals play in community language and what might this mean for how change unfolds.

In order to approach this question, I first sought to determine the distribution of entries by user. Figure 11 reveals the overall distribution of user contributions between 2002-2011. The

x-axis shows the number of entries made in bins of 500, while the y-axis shows the number of users who made a number of entries within each bin.

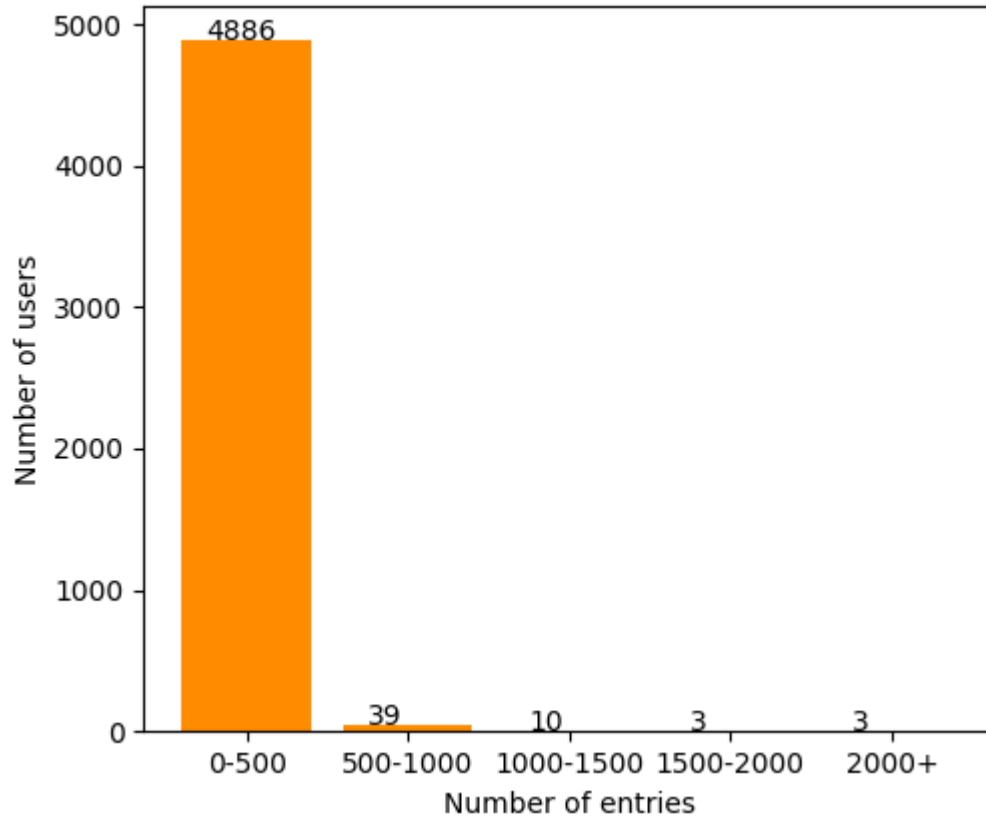


Figure 11. The number of entries made by users within the 2002-2011 timespan under study. Almost 5000 users made under 500 entries, where 39 made between 500-1000, 10 made 1000-1500, 3 made 1500-2000, and 3 made more than 2000.

As can be seen from Figure 11, most users make 0-500 entries. There are only 55 users who make over 500 entries, with six top posters emerging across the most active community years, making over 1500 entries each. These users make up 10-15% of total entries between them during the periods from 2006-2010, where community norms are developing. While dividing the 0-500 bin up further would give us a more detailed sense of posting behavior for low-post users, this analysis of user patterns more generally is outside the scope of this thesis. Instead, due to the clear leaders in terms of number of entries, the top six users will

form the basis of the proceeding analysis, which will explore the dynamics of how these individuals interacted with the rest of the community.

Top users: method

In order to look at how the language of the top users compared to the overall community, I again adapted the method from Danescu-Niculescu-Mizil et al. (2013), looking at whether users were ahead or behind the community across their life in the community. In order to do so, I again made bigram LMs for each 3-month period, but in this case trained on data that excluded the entries of the top six users. In this case, I trained 100 models for each time period (again training each on the first 30 words of 1000 randomly selected entries) instead of 1000 due to time and resource constraints. For each time period, I ran each user's entries from that time period through the LMs of the 12 months beforehand and afterwards. Each entry was then assigned a number between -4 and +4 corresponding to the period within which it had the lowest average cross-entropy value. For example, Figure 12 shows how the first 30 words of a post from the Apr/May/June period (reproduced above the diagram) would get cross-entropy values for LMs from eight time periods.

"i am going on 34 now and am like 2 years on t and like 10 months post op and i started this whole journey like back when i was"

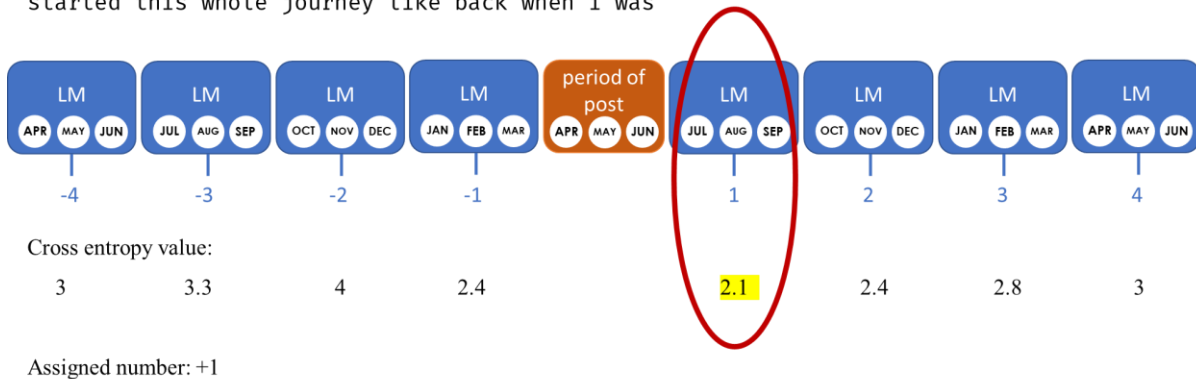


Figure 12. Demonstration of how an entry would be run through the language models one year before and after (receiving eight cross-entropy values and then being assigned a number between -4 and +4 based on these).

The lowest cross-entropy value (in this case 2.1) would be selected, indicating the post is best predicted by, and hence most similar to, the community's language in the three months following the period it appeared in. The entry would then be assigned a number corresponding to the time period of the value (in this case +1). A negative number means the entry is more similar to previous time periods, whereas a positive number means it is more similar to future time periods. These numbers were then averaged across the 100 resampled LMs to get an overall number for each entry. Within each time period, the assigned numbers of each poster's entries were then averaged.

Top user patterns

In order to show how the top six users patterned in terms of the overall community, Figure 13 shows how ahead or behind the community change the top six users were. The y-axis represents which 3-month period the entries in each period are most similar to, where each increment is one 3-month period distant from the period of entries under study (negative for before and positive for after). The red horizontal line shows the point at which someone

would be in line with the community change at that point in time. We would expect to see the lines representing how ahead/behind each user follow the red line if the top users conformed to the community change. As can be seen however, all users start off behind the community and then become more ahead after the 2006 period.

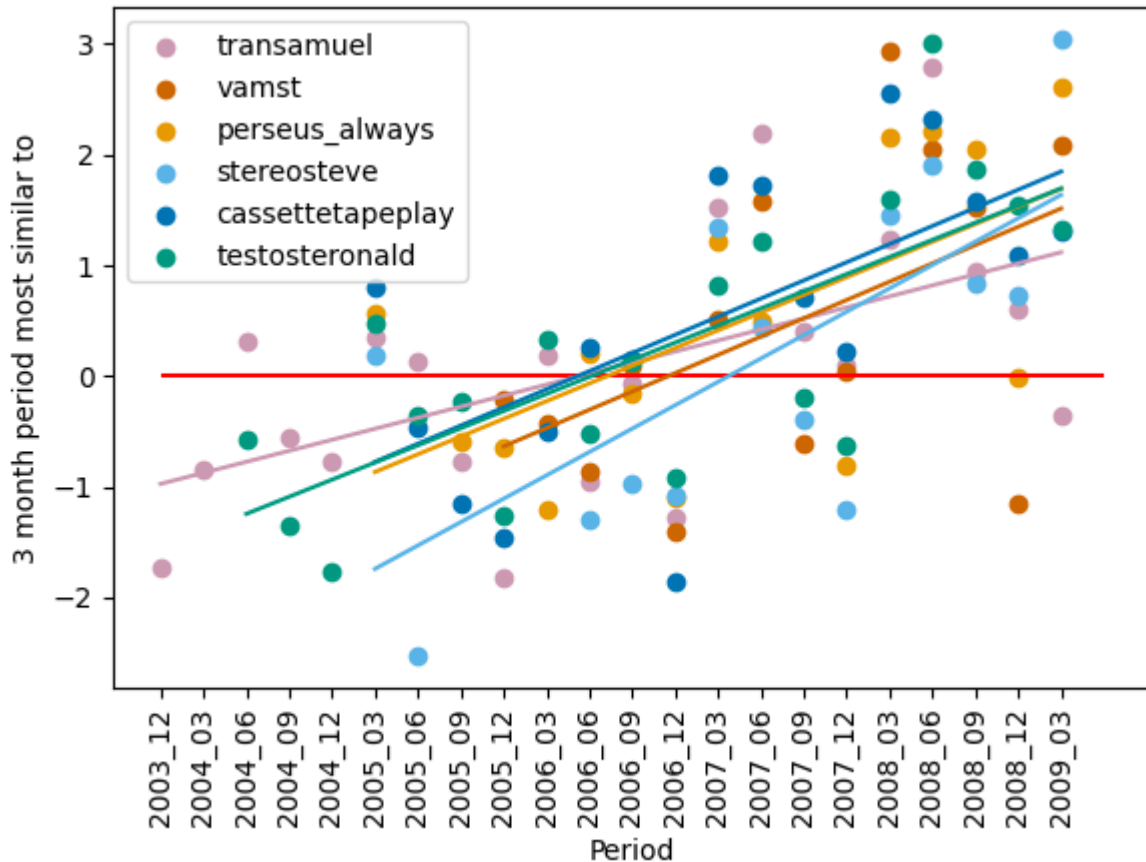


Figure 13. How ahead or behind the community language the top users are. The x-axis represents the 3-month period of the users' entries. All top six users become increasingly ahead of community language use.

The users follow the same pattern in terms of how ahead they become with respect to the community. This pattern is surprising, given the pattern seen in previous communities of users starting off more ahead and then beginning to fall behind as they fall out of step with community change. The primary question of interest then is why do we see the opposite pattern here? One hypothesis is that the top users are changing the way they use language

over time in a way that becomes more similar to community change and then becomes more future-oriented – perhaps because of engagement in external communities of practice or because they are shifting as part of an overall trend. Another possibility is that top users are consistent in the way they use language, but that over time, other users are increasingly modeling their language on the language of the most prolific posters. A related possibility is that, because there is increasingly less content overall, as people mirror the language they see around them, they will increasingly mirror language from the past, meaning that all users' entries will be increasingly similar to future entries.

In order to examine the first hypothesis, I compare top users' language to that of one period of change: 2006 Apr-June. By running a user's posts over time through the LMs from a single period, we can see whether their language is becoming more or less predictable from, and therefore more or less similar to, the language of that period. If we see that their predictability is stable over time, we can take this as a metric that suggests that the way they use language is stable over time (or at least that any changes they are making are within the scope of language use at that time, averaging out to make them equivalently predictable from the community during that period). While running the entries through the LMs of any one time period could give us an indication of such stability, I have selected a period where users cross from being behind the community to ahead of the community. If the users were in fact changing their language (for example, along with or ahead of the community change) after this point, we could expect to see their language becoming less predictable from this period over time.

Figure 14 shows language-use consistency over time for the top six users. The x-axis spans 2002-2010 in 3-month periods and the y-axis represents the cross-entropy of the users. The

points are the mean cross-entropy values from 1000 LMs, accompanied by lines of best fit. If the lines are horizontal, it would mean that the users aren't changing their language. If they are diagonal, it would mean that the user was becoming more similar (downward trend) or different (upward trend) to the given time period.

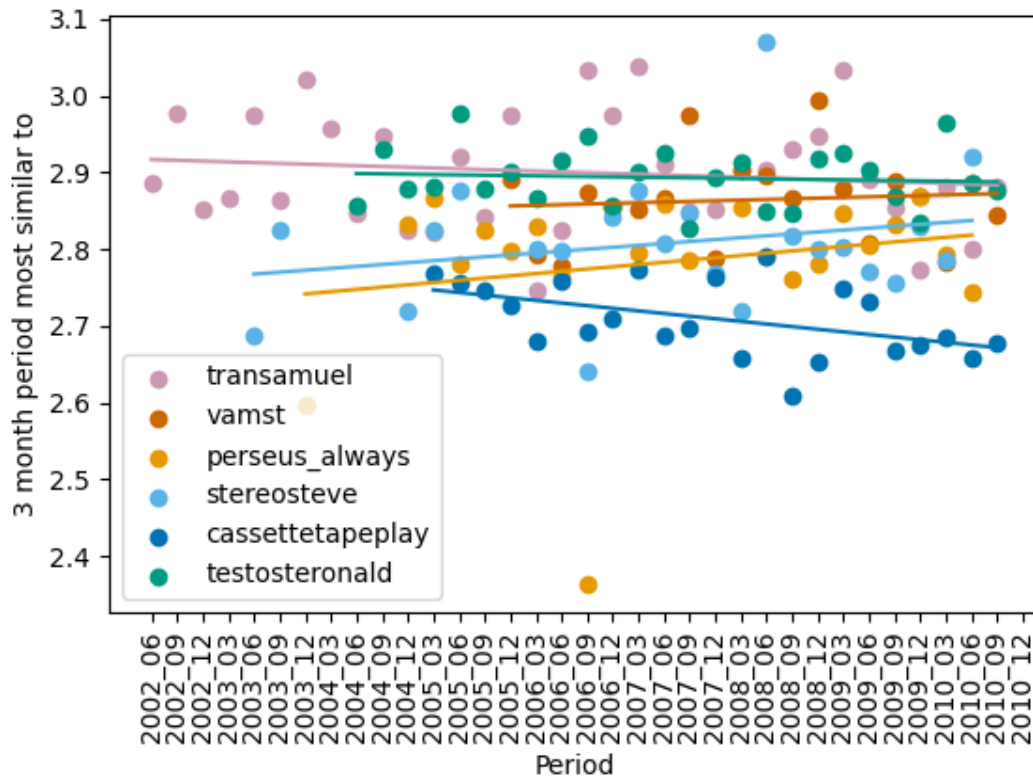


Figure 14. Cross-entropy stability of the top six users with respect to the community's language in April-June 2006. The lines vary in their direction, meaning that while the users differ with how similar they become to the April-June 2006 time period, which does not line up with the clear similar pattern seen earlier.

The lines in Figure 14 vary in whether they become more or less similar to the given time period over time. cassettapeplay and, to a lesser extent, transamuel become increasingly different to the April-June 2006 time period, whereas perseus_always and stereo_steve become more similar to it. vamst and testosteronald stay relatively constant. This suggests that, while the users follow the same pattern in terms of how ahead they become with respect

to the community, this is not explained by similar patterns in the way their language use is (or isn't) consistent.

As there is no clear pattern of the top users' language changing, we can instead look at our next hypotheses. Given that they are becoming increasingly ahead of the community's language after 2006 (Figure 13) and that the community is becoming more cohesive after 2006 (Figure 10), it appears that the community around them is changing to behave linguistically more like those top users. This could either be through people intentionally modeling off their language after influential individuals, or because of a general pattern for content to be more similar to previous content over time. In the latter case, these top users would still have a considerable effect, given that, in the periods from 2006-2010, the top users make up to 10-15% of total entries between them. Given this influence, it is important to ask: who are these highly prolific posters and how are they behaving within the community?

Top user identities and behaviors

The demographic characteristics and interactional styles of the 6 most prolific posters in the community reveal potential patterns around privilege, interactional domination, and transnormativity.

Demographics and interactional styles of the top six users are described in Table 1.

Demographics were taken from users' posts and profile information. These are not comprehensive, as they are based on their online presence. Within the corpus, I searched a variety of race terms, identity terms, sexuality/relationship terms, medical interventions, and location terms for each of the top users. While analyzing specific individuals raises some

questions in terms of ethics, the data from the corpus only contains publicly available information, while the users had the opportunity to make their posts only readable by “friends” (in this case, community members). For this reason, and because the people themselves would have continued to grow and shift in their opinions and self-understanding, I do not analyze these profiles as representative of the humans themselves, but of the constructed online personae they created on this forum.

Table 1: Top six user information

User	Demographics/info	Interactional style	Entries	First year
transamuel	20s, white/Jewish, trans man/guy, pansexual/queer, Massachusetts, moderator from the beginning	Friendly; educates people; wants everyone to get along; mostly doesn't take strong stances; doesn't engage in much metalinguistic commentary	3510	2000
perseus_always	20s-30s, white, male / transsexual, post-transition, female partners, non-disabled, Virginia, moderator from mid-2000s-2010	Confrontational; sarcastic; strong stances; almost always contributes to top posts; engages in a lot of metalinguistic commentary	3129	2003
stereosteve	Female-to-male transsexual, Florida; member of communities for trans Christians and male-identified trans people; possibly MBA student; pre-T/surgery, moderator from mid-	Confrontational; sarcastic; strong stances; agrees with perseus and often mirrors language	2237	2003

	2000s			
testosteronald	40s; Man with a transsexual medical history/ post transition man; dates men; USA	Matter-of-fact; mobilizes expertise – gives detailed earnest advice; strong stances on what he should be allowed to say; sometimes contributes a comment or two to top posts	1650	2004
cassettetapeplay	College-age/late teens; Gay; surgery and T; originally identified as genderqueer but then as trans guy who sees his transness as defect/illness; USA	Shares info and personal experience; some metalinguistic commentary; expresses negative stances towards community conflict; hardly ever contributes to top posts	1630	2005
Vamst	Surgery; ‘gender variant woman’ who is ‘partially transitioning’; grew up Catholic; dates women; ‘trans* but not ftm’; white; USA	Earnest; shares personal experience; asks for advice; often contributes to top posts with questions, suggestions, or hedged stances	1538	2005

Though terminology has changed since the time these posts were made, the contemporaneous self-descriptions of the top users summarized in Table 1 indicate an alignment with binary-based models of trans identity, which was also framed by several of them as a medical issue. Most either had had surgery/were on hormones or positioned themselves as “pre-” such aspects of medical transition. In these ways, the top users mostly oriented to a transnormative

or transmedicalist model of being trans (Jacobsen, 2022; Konnely, 2022). Such a model of conforming to the idea of trans people being “born in the wrong body” and heterosexual echoes medical gatekeeping and the pathologizing of transness (Jones, 2019; Zimman, 2012). The hegemonic status of these normative discourses in the US have historically excluded those whose experience of gender was not binary – limiting access for these people to both community and medical interventions. Another notable aspect from Table 1 is that all users with an indication of race are white and all were US-based. In these aspects, the top users occupy systemically privileged positionalities relative to other transmasc folk. The three top posters were also moderators for part of their time in the community. A “moderator” is someone who is responsible for managing what goes on in an online community – removing content and banning users for violating community terms. Thus, being a moderator comes both with social status and with the ability to wield power within a community.

Given the ways in which the top speakers embodied transnormative and/or otherwise privileged positionalities, looking qualitatively at the data allows us to examine ways that normativity and privilege may have impacted the dynamics of community language within the group – allowing us a more holistic view of the patterning than what can be revealed by a quantitative account alone.

My qualitative analysis, which focuses on stance-taking (Du Bois 2007), is based on the 10 posts with the most comments during the period of change already discussed (2005-2007), as well as other entries containing metalinguistic commentary,¹⁰ demographic information,

¹⁰ Identified through searching for entries containing 'the word', 'word[s]* like', 'language', or 'term'; looking entries with the 'language' tag; and looking at entries discussing 'cis'/'bio' or 'transgender'/'transgendered'.

and/or identity terms. I chose these sites because of their high potential for interactive tension.

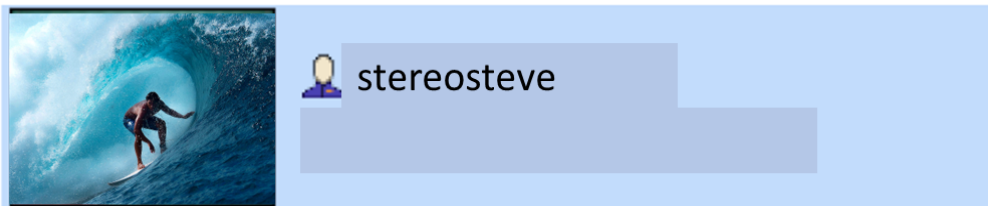
Analyzing top posts shows that conflict in this community frequently centered around oppressive behavior and language. This can be seen in Table 2, where the topic and number of comments from each of the 10 posts with the most comments are detailed.

Table 2. 10 top posts between 2005-2007

	Post topic	Num of Comments in corpus
1	Concerns about a split off group making fun of posts in ftm and concerns about the use of ableist language	291
2	Worry about pressure to transition quickly; terms of discussion; sexism and misogyny	258
3	Someone being banned for using sexist language	244
4	Trans attraction: dating preference for trans men because of 'feminine qualities'	238
5	Asking why it is not okay to use sexist and ableist slurs but okay to use homophobic ones	167
6	Post against sanitized moderation and also about this community being for those who are explicitly having 'sex changes'	166
7	Against transmasc people in women's spaces	166
8	Asking why people don't like educating cis people on the forum	165
9	Fatphobia in the community	164
10	Cis person wanting to research trans sex issues for a workshop	152

Conflict is a driver of engagement (Mensah et al. 2020), which bears out in the conflictual nature of these top posts. Throughout these posts and other metalinguistic commentary, conflict takes three main forms. Firstly, it centers on oppressive behavior and language. Secondly, it is about how people are relating to and mobilizing trans identity terms. Thirdly, it is a negotiation of who and what discussions are allowed in this community.

Regarding oppressive behavior and language, the top posters consistently shut down and call out slurs and other language positioned as oppressive. Through this, they set expectations for linguistic behavior in the community. In Post 3 from Table 2, *perseus_always*, *stereosteve*, and *vamst* all oppose the use of sexist language within the community. In Example 8, *stereosteve* responds to another user, explaining why using the term *bitching* is sexist.



because "bitching" is used to describe complaining like a "bitch" which goes back to the definition and history of the word. which is in and of itself sexist. by virtue of the fact that it degrades women to the status of a dog.

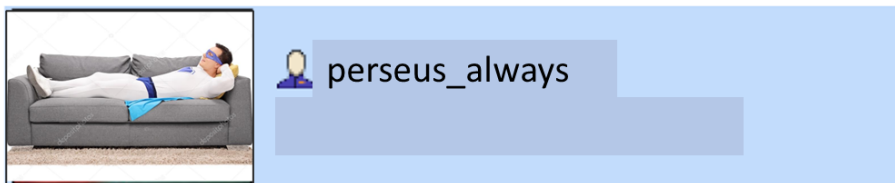
[Reply](#) [Parent](#) [Thread](#) [Collapse](#)  [Like](#)

Example 8.

At this point in time, *stereosteve* was a moderator and spent several comments in this thread defending his moderator status to others. In the comment shown here, he reiterates his stance on different permutations of the word *bitch* and gives a considered explanation. However, he uses full stops between subordinate clauses and their matrix clauses to emphasize the

obviousness of the message, as this conflict around the use of the term *bitch* was prevalent in the community at this time. His invocation of his moderator status gives weight to such statements as the definite stance within the community.

In this same thread, people made analogies between sexist language, slurs for people with intellectual disabilities, and racist language. In response to someone using an anti-Black slur, perseus_always, who was also a moderator at this time, engages in an exchange wherein both parties use a confrontational tone. In Example 9, perseus_always responds to a poster calling him “out of line” and a “jerk”.



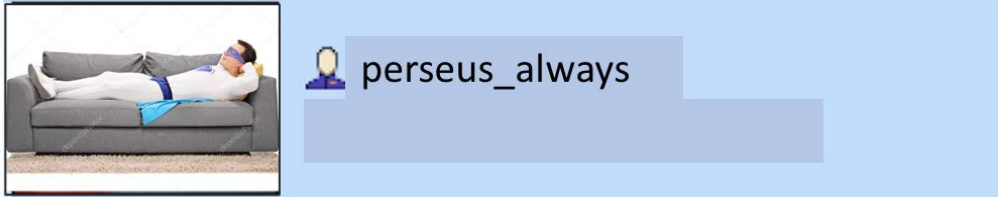
yes. you are racist and unaware of both your privilege and the historical context in which this country deals with race
[...]
if you think for one second that i’m going to be told that i am “out of line” and “a jerk” for pointing out and acknowledging the grossness of what you’ve said here, you are wrong. twice.

[Reply](#) [Parent](#) [Thread](#) [Collapse](#)  [Like](#)

Example 9.

In this comment, perseus uses short fragments (“yes.”; “twice.”) to indicate disalignment with and negative evaluation of the previous user, as they represent a blunt, curt tone and can convey an implicit evaluation of the interlocutor as not being able to comprehend long sentences. He uses predicates to directly label the other user as “racist” and “wrong” for their

comments and assumptions. In response to a further explanation from the poster, in Example 10, we see perseus rejecting the user’s invocation of proximity to Blackness as an index of anti-racism by responding with the sarcastic rhetorical question “this is a joke, right?”



“i have black friends.”

this is a joke, right?

i’m done with this. Unpack your knapsack on your own

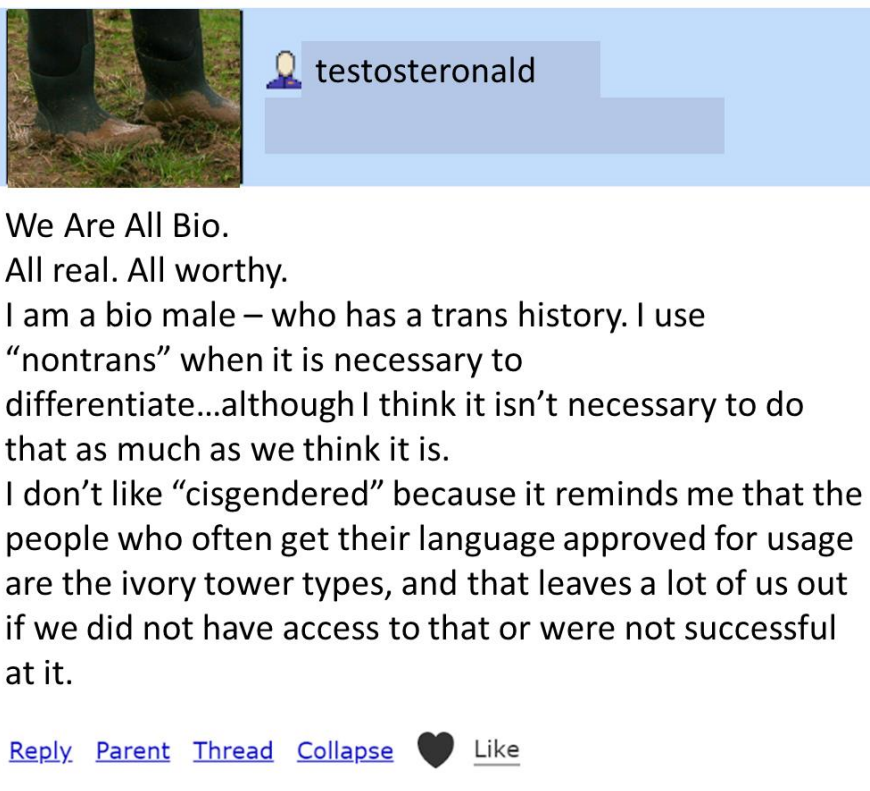
[Reply](#) [Parent](#) [Thread](#) [Collapse](#)  [Like](#)

Example 10.

In saying “Unpack your knapsack on your own,” perseus makes a reference to an essay that uses “a knapsack full of special provisions” as a metaphor for white privilege (McIntosh, 1998[1989], pp 148), indicating that the behavior of the poster had led perseus to abandon him to the task of recognizing his privileges by himself.

In such posts, perseus engages in a confrontational strategy to shut down anti-Blackness. By belittling the previous poster, perseus makes a statement that racism is not welcome in the community. As a white user, perseus aligns with racialized people in the group and constructs an anti-racist persona. However, in response to this strategy, interlocutors frequently did not back down from their positions, but rather, doubled down on their racist language and perspectives.

Confrontational language is elsewhere critiqued from within the community based on the notion that shutting down people’s language can uphold and support classism. Users discussed how the penchant for being highly critical of others’ terminology fails to acknowledge the structural inequity in who has access to information. In one thread about the conflict between *cis* and *bio*, testosteronald, as illustrated by Example 11, discusses the way that academic elitism plays into the way that some words are taken up over others.

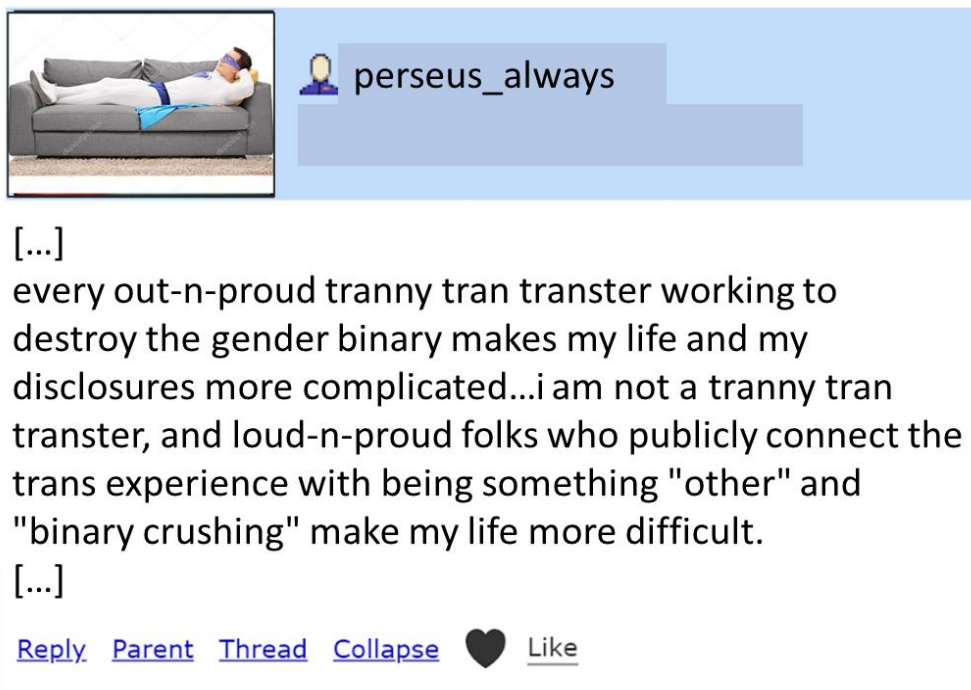


Example 11.

In this, he affirms the collective validity of both trans and cis men as “All Bio” (conflating community members and men with his use of the term “we”), using capitalization for emphasis. He then critiques the use of and correction to the word *cisgendered* for being the kind of language that leaves out those who did not have access to college spaces. This unevenness in access, as well as differences across offline communities of practice is not

recognized by a politic that favors harshly correcting the language of others. Thus, such linguistic expectations are likely to influence who stayed active in the community.

The second main source of conflict is in how people are relating to and mobilizing trans identity terms. The top posters mostly identify as binary and several view their transness as a medical issue. Other conceptions of transness are often ridiculed – which is at odds with the way the very same users were engaging with racism, ableism, sexism, and misogyny. In Example 12, perseus_always engages in a thread from 2006, wherein multiple users identify a “culture war” in the community.



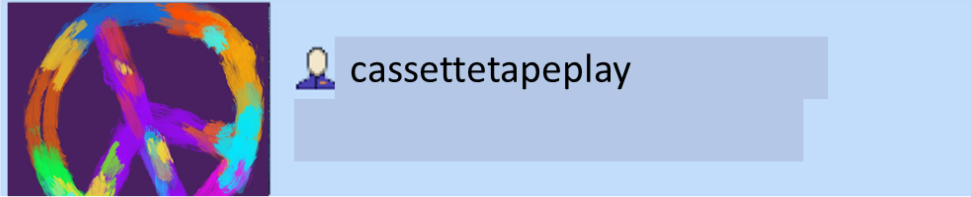
Example 12.

perseus_always produces a binary between two different camps of trans folks: those “working to destroy the gender binary” and those who, like him, conform to normative societal concepts of gender. He uses mocking language to refer to the first group, expanding

the slur for transfeminine people *tranny*¹¹ to make fun of those who identify publicly with their trans status, and using the rhyming compounds “out-n-proud” and “loud-n-proud” to further present these folk as juvenile caricatures. This transnormative discourse – where conforming to hegemonic gender norms and desiring to assume the privilege of cisgender (white) males is constructed as the normal and reasonable path, and a more radical notion of transness that seeks to destabilize oppressive gender norms is characterized as ridiculous – is also produced by other posters.

In a different thread from 2008, cassettapeplay criticizes a posters’ speech for presenting a notion of transness focused on community, pride, and destabilizing binary gender norms. In Example 13, cassettapeplay reacts against an idea that “transsexuals [should be] political pawns for anyone’s movement”, stating that people’s privacy should be as important as “self-expression.”

¹¹ Many transmasculine people at this time had limited awareness regarding its specifically transmisogynist reference.



[...]

Transsexuals are not political pawns for anyone's movement-- they are individuals with medical needs whose privacy ought to be just as well-protected, and just as high a priority, as a gender-non-conforming person's right to self-expression.

[...]

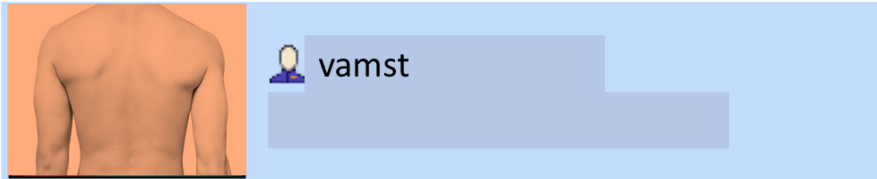
[Reply](#) [Thread](#)  [Like](#) [Complain](#)

Example 13.

In this thread, he positions his own experience and views against a notion of “trans pride” by saying that he doesn’t feel pride in his identity, which he frames as a medical issue. These conflicts about who counts as trans and what the political project of trans people should be, has been pervasive across trans communities (e.g., Jacobsen et al., 2022; Konnelly, 2022; Tao, under review), leading to exclusion and fracturing of communities.

This brings me a third source of conflict regarding who and what discussions were encouraged in the community. Post 2 from Table 2 discusses a pressure perceived by the poster for people to transition quickly, and the sexism inherent in a perception of masculine women as inferior to transmasculine people. In this thread, users share personal experiences about the boundary between butch and ftm people (to read more on this, see Halberstam, 1998). On this subject, vamst, among others, comments on the way that such discussions and perspectives are often not welcomed in the community. In Example 15, Vamst observes that

such posters in discussions are often encouraged to move to less active ‘splinter communities’ (such as ‘femme_ftm’).



What I mean to say is, if people make posts on these subjects and are immediately referred to a splinter community, I think we give the impression that we are trying to get rid of them. Often a day or so later, the same intro post will pop up on the community the poster was referred to, it will get a few 'welcome' replies, and the poster doesn't post here (or there) again. We lose those voices when we ought to be encouraging them--or at least this conversation suggests that many people would appreciate hearing those voices. ftm has created itself as a transitioning community. It can recreate itself as well.

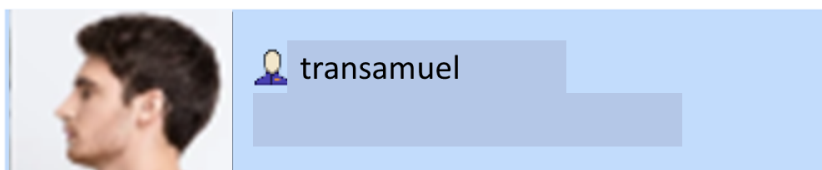
[Reply](#) [Parent](#) [Thread](#) [Collapse](#)  [Like](#)

Example 15.

They produce a narrative of gender non-normative people coming to these online communities, not finding space for them there, and then leaving. They characterize this as “los[ing]...voices” which the community “ought to be encouraging” instead. They characterize the current instantiation of the ftm LiveJournal community as “a transitioning community”, where diverse perspectives and experiences of gender are not welcomed. In leveling this critique and suggesting the possibility of a different community trajectory, they lend authority to their position through invoking an idea of “many people” from previous comments who agree with them. This example shows how linguistically-expressed social

behaviors can directly influence the population dynamics of the community, which in turn affects the community-level linguistic norms.

The top post from this time period (Post 1, Table 2) also reveals several different stances regarding what discussions should be present in the community. In this thread, the users discuss the existence of a spin off community that makes fun of people on ftm, using ableist language to do so. In this thread, perseus_always and stereosteve are among others that defend the spin off community’s right to exist, and the importance of being able to make fun of what they see as a deluge of basic or annoying behavior in ftm. Others defend ftm as a space where people should be able to post novice, trans101 questions and comments. testosteronald and transamuel are among those who criticize this spin off community, and critique the involvement of moderators from ftm in it. In Example 16, transamuel makes one of his few comments in the top threads, aligning with previous posters in condemnation of the other group.



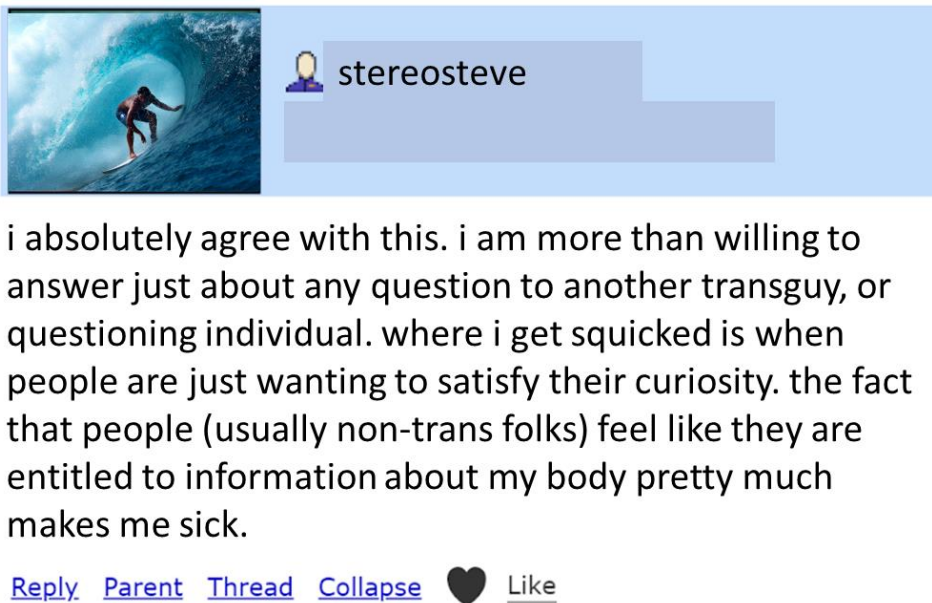
Voicing my opinion as another co-mod, I don't really have anything original to contribute to the discussion at this point; I agree with the comments above.
[...]

[Reply](#) [Parent](#) [Thread](#) [Collapse](#)  [Like](#)

Example 16.

transamuel here deploys his authority as “another co-mod” to back the viewpoints expressed earlier and add weight to his ensuing discussion about moderation in the community.

Another kind of content that was explicitly unwelcome was cis people asking for information about trans peoples' lives. In Post 8 of Table 2, a poster asks why others don't like educating cis people on the forum. A heated discussion ensues, wherein people call each other "asshole" and "jerk" and make snarky and sarcastic comments to each other – either expressing strong alignment or disalignment with previous comments. In Example 17, stereosteve aligns with a previous poster in discussing the key role a questioner's identity plays in how comfortable he is answering questions.



Example 17.

stereosteve evokes the physiological experience of being "squicked" or being made "sick" at cis folks asking questions about his body. The popularity of such posts in the community demonstrate that this was not just a community accessed by trans folks, but also served as a (disavowed by many in the community) source of education for cis people.

The top users, who were undoubtedly influential considering their disproportionate number of entries and percentage of total community content over time, became increasingly ahead of community change as participation in the community decreased and community norms in language use developed. Key conflicts around oppressive and identity-based language were drivers of community engagement, with key positions taken by the top users potentially influencing the overall make-up of the community. In these conflicts, the top users occupied varying positions with respect to each other, the topics, and to the community in general. While they (especially the moderators) frequently called out racism, sexism, and ableism, their manner of doing so often led to hostile exchanges, rather than generative growth and change. Despite the differences in their perspectives on transness and terminology, the top users approached transness from a white US perspective, and mobilized moderator status, the voices of others, and other indications of authorization (Bucholtz & Hall, 2004) in forwarding their perspectives. Gatekeeping created by upholding certain linguistic expectation affects the makeup of the community and what they post about, which provides a potential causal link between the diminishing community size and the consolidation around linguistic norms.

7. Discussion & conclusion

There is a need to more fully incorporate interactional and political dimensions into our theories of language change. As the ubiquity of internet spaces allows us unprecedented analytical access to language use and change on a large scale in real time, it is important that we are also attentive to more local dynamics of power. With this in mind, when thinking about trans community language change, there are four levels to attend to. At the broadest

level, there is the societal perception of trans language change and the way that it is constituted in wider sociopolitical discourse. Linking to this, perhaps most transparently, there is terminology around gender itself and the way that queer and trans language is negotiated, critiqued, and constructed across queer and trans communities. Then there are lexical and discourse norms at the level of communities of practice – and the way such norms are established within them. Finally, there are the actual individuals themselves and how their individual language changes. By employing a mixed methods approach, this work has been able to investigate ways that change at these levels are in conversation with each other.

With respect to trans terminology, in this work I found that lexical change in this ftm LiveJournal community does not unfold in a similar way to lexical change in previously studied beer rating forums. Rather than there being an adult stability effect, users from all cohorts make the community-wide lexical changes together. This suggests that there is something particular about the way that language is being discussed. And in fact, we see in the community that there are whole threads debating specific lexical changes. Such metalinguistic commentary and correction may cause those who use other terminology to either change their language or leave the community. While people vocally advocate for language change, change itself is gradual and distributed – which could mean that broader conceptions of impossibly fast trans language change may be based on vocal individuals or simply the degree of metalinguistic awareness community members display. However, what this picture also disguises is that within these threads there are a plethora of different opinions and perspectives on these linguistic changes that sometimes get ignored or not taken up in favor of the language used by those who are seen to have more structural privilege. This played out in the case of *cis*, where there was strong opposition to it on the grounds that

it was too academic and inaccessible, and yet it ultimately still got taken up and is now widely known even outside of trans communities.

In thinking about whether moves towards homogeneity on a lexical level are part of a broader community pattern, it is interesting to examine community norms in other ways. This thesis has taken a quantitative approach to this by using bigrams to get a sense of the way people are putting words together in their posts. Looking at similarity through a probabilistic method such as this one, where predictive language models are trained on the corpus of entries and then the predictability of posts is calculated, allows us to approximate the experience of someone looking at the community homepage and seeing posts and comments that are either negotiating similar ideas using similar discourse patterns, or alternatively, seeing something much more varied. We see that the level to which someone would observe similarity varies between the years 2002-2006 as the community was growing in content and members. However, as users began leaving with the advent of other social media platforms, the similarity began to increase.

Looking at how the most prolific users pattern with respect to the community's norms adds to this image: after 2006 we see that the top six users' posts were more similar to the community's language in increasingly future time periods. This means that as the community progressed in time, it was increasingly mirroring the kind of language utilized by the top users. As these users are overwhelmingly white and from the US, as well as mostly embodying transnormative/transmedicalist ideas, their language is both framed within and discursively constructs a certain conceptualization of transness. Even though this was an open community, it was very US-centric in its discourse and membership. This has practical implications in that people from other countries had to access more specialized (and perhaps

inactive) groups to find the local information that was available to US users in the general ftm community, if such groups existed at all. But the US-centrism of generic online trans groups also does symbolic work in contributing to a perception of transness or queerness as a US phenomenon, with such discourse able to be co-opted by conservatives in other countries to invalidate local efforts to address transphobia (e.g., see Baran, 2022; Borba, 2022; Tudisco, 2022). Additionally, the overwhelming whiteness and abledness (which is in some cases assumed based on their discussion of these topics without invocation of their own identity, as is common by those who hold unmarked identities) of the top users means that their experience of being trans is not embodied in an intersectional way (see miles-hercules, 2020), also contributing to the valorization and normalization of a particular understanding of transness.

As role models and moderators in the community, the top users have a privileged position in shaping its members and interactions. Previous studies (e.g., Lin et al., 2017) have captured the effects of such moderation through studying which posts and comments get deleted, but such an approach fails to account for the interactional way that moderator power is employed. Even if moderation is done in favor of creating a non-oppressive culture, using confrontational techniques to do so can help produce a culture of insults, name calling, and sarcasm. Such interactional norms were a frequent source of comment and negative evaluation within the community. However, the fact that so much racism, ableism, and sexism needed to be addressed within this community speaks to a common criticism of wider LGBTQ+ communities as being overwhelmingly white and ableist spaces that can feel difficult for racialized and disabled people to feel comfortable and supported in. Another relevant aspect is that such confrontational techniques were not only directed towards

instances of sexism and racism, but also towards transmasculine or nonbinary people who had different conceptualizations of gender. Users discussed how hostility or otherwise lack of welcome towards people who didn't conform to transnormativity or transmedicalism meant that posters may leave the community – and may in fact ultimately leave all LiveJournal communities. This is important because not having access to online trans communities, especially for geographically isolated folk, can mean not having access to crucial peer information about medical or logistical aspects of transition, potentially leading them to accept misleading or noncomprehensive medical assessments and treatment (such as people with high BMI being told they can't have surgeries or will have to accept bad results), or leading them to withhold key medical questions they have about surgeries for fear of being seen as 'not trans enough' (Konnolly 2022). Additionally, given the importance of online spaces for feeling seen and validated in one's identity, being excluded from online spaces can deprive trans youth of 'a crucial lifeline' (Jenzen 2017, p. 1638).

As many people, both trans and cis, look to online groups for education, validation, and inspiration, language and discourse norms within such a publicly accessible community have greater influence than can be seen just by looking at the comments and posts themselves. Online trans spaces play a large – and perhaps increasing – role in constructing and constituting social categories. They can contribute to the way someone relates to the individual and political dimensions of gender identification. It is thus important to investigate the role that community-internal power plays in constructing these norms.

In this thesis, I have explored language change in an online trans community at a time of internet transition from the more static web 1.0 to the more dynamic web 2.0. This transition allowed new ways for trans people to connect, negotiating language change through online

interaction. As we move to web 3.0, where VR provides new opportunities for embodiment, and where community on platforms like TikTok is constructed based on algorithmic patterns of engagement rather than bounded groups, the interplay of language change at different levels and the involvement of corporations becomes more striking. In this context, innovative mixed-method approaches, such as I have taken in this thesis, are crucial to investigate the complexity of language change in an increasingly interconnected world.

Future work

To extend these findings, future research with this community could expand on my analyses of lexical change, community change, and on the influence of users on the community as a whole.

As was discussed earlier, lexical change in key identity terms could be compared with other lexical innovations in the community that are below the level of awareness. Tie-strength and social networks could also be explored in terms of lexical innovations to see how they compare to other studies.

Regarding community change, the community cross-entropy was calculated using the first 30 words of each post. However, sometimes users began their posts with quoting, in italics, a previous commenter. In order to avoid the potential for this to affect the data, the last 30 words, or a middle section of 30 words could be used. In addition, Danescu-Niculescu-Mizil et al. (2013) experimented with different numbers of words and found the results were stable across different numbers. This could also be a way to validate the current results presented here. Additionally, the depth of thread (number of comments per post) could influence patterns of similarity were it to stay consistent, even as posts drop off, as old entries would be

visible in the community for longer. Thus, another avenue in terms of analyzing community change could be to look at such interaction patterns, as in Lin et al. (2017), to see if behavior shifts across the course of the community's life.

In terms of the effect of the users themselves, if we are conceptualizing the top users as influencing or at least representing the interactional styles and opinions of those who are staying in the community post-2006, this leads to the question: what is going on with those who are leaving? I did preliminary work on the language patterns of 'leavers', but this was excluded from the thesis due to a need to further develop an understanding of who should be counted as a 'leaver'. Many users were only active in the community during one 3-month period, so initial results that only looked at entries in a user's final 3-month period often overlapped with the user's initial (and only) 3-month period. These users should be excluded from future analysis of leavers, but standards for how long someone must be in the community and how many posts they must make before being counted as 'leaving' is an open question. Understanding the patterns of those who leave and how they may contrast with those who stay is key to further understanding the community dynamics that have been explored thus far. Interactions with these 'leavers' could also form a key site for qualitative analysis.

References

Anderson, Benedict (1983). *Imagined communities: Reflections on the origin and spread of nationalism*. Verso.

Baran, D. (2022, April 7-9). *The Right-wing Construction of LGBTQ Communities as the*

- National “Enemy” in Recent Nationalist Discourses in Poland.* Society for Linguistic Anthropology Conference 2022, Boulder, CO, United States.
- Bassichis, M., Lee, A., & Spade, D. (2011). Building an Abolitionist Trans and Queer Movement with Everything We’ve Got. In E. A. Stanley & N. Smith (Eds.), *Captive Genders: Trans Embodiment and the Prison Industrial Complex* (pp. 15–40). AK Press.
- Borba, R. (2022, April 7-9). *Reimagining the Nation: Language, Gender, Sexuality and Democratization in Contemporary Brazil.* Society for Linguistic Anthropology Conference 2022, Boulder, CO, United States.
- Bucholtz, M., & Hall, K. (2004). Theorizing Identity in Language and Sexuality Research. *Language in Society*, 33(4), 469–515.
- Bybee, J. & Hopper, P. (2001). Frequency effects and the emergence of linguistic structure. In J. B. Pierrehumbert (Ed.), *Exemplar dynamics: Word frequency, lenition and contrast*, (pp. 137–57). John Benjamins.
- Byron, P., Robards, B., Hanckel, B., Vivienne, S., & Churchill, B. (2019). “Hey, i’m having these experiences”: Tumblr use and young people’s queer (dis)connections. *International Journal of Communication*, 13, 2239–2259.
- Cohen, C. J. (1997). Punks, Bulldaggers, and Welfare Queens: The Radical Potential of Queer Politics? *GLQ: A Journal of Lesbian and Gay Studies*, 3(4), 437–465.
<https://doi.org/10.1215/10642684-3-4-437>
- Cole, J. R., Ghafurian, M., & Reitter, D. (2017). Is word adoption a grassroots process? An

- analysis of Reddit communities. *Social, Cultural, and Behavioral Modeling - 10th International Conference, SBP-BRiMS 2017, Proceedings*, 236–241.
https://doi.org/10.1007/978-3-319-60240-0_28
- Cukor-Avila, P. & Bailey, G. (2013). Real time and apparent time. In N. Schilling, J. K. Chambers (Eds.), *The Handbook of Language Variation and Change*, (pp. 239–262). John Wiley & Sons.
- Cunha, T., Jurgens, D., Tan, C., & Romero, D. (2019). Are All Successful Communities Alike? Characterizing and Predicting the Success of Online Communities. *The World Wide Web Conference*, 318–328. <https://doi.org/10.1145/3308558.3313689>
- Dame, A. (2016). Making a name for yourself: Tagging as transgender ontological practice on Tumblr. *Critical Studies in Media Communication*, 33(1), 23–37.
<https://doi.org/10.1080/15295036.2015.1130846>
- Dame-Griff, A. (2019). Herding the ‘Performing Elephants:’ Using Computational Methods to Study Usenet. *Internet Histories*, 3(3–4), 223–244.
<https://doi.org/10.1080/24701475.2019.1652456>
- Danescu-Niculescu-Mizil, C., West, R., Jurafsky, D., Leskovec, J., & Potts, C. (2013). No Country for Old Members: User Lifecycle and Linguistic Change in Online Communities. *WWW 2013 - Proceedings of the 22nd International Conference on World Wide Web*, 307–318. <https://doi.org/10.1145/2488388.2488416>
- Dayter, D., & Rüdiger, S. (2016). Reporting from the Field: The Narrative Reconstruction of Experience in Pick-up Artist Online Communities. *Open Linguistics*, 2(1), 337-351.
<https://doi.org/10.1515/opli-2016-0016>
- Du Bois, J. W. (2007). The Stance Triangle. In R. Englebretson (Ed.), *Stancetaking in*

- discourse: Subjectivity, evaluation, interaction*, (pp. 139–182). John Benjamins.
- Flores, N., & Rosa, J. (2015). Undoing Appropriateness: Raciolinguistic Ideologies and Language Diversity in Education. *Harvard Educational Review*, 85(2), 149–171. <https://doi.org/10.17763/0017-8055.85.2.149>
- Fuchs, C., Hofkirchner, W., Schafranek, M., Raffl, C., Sandoval, M., & Bichler, R. (2010). Theoretical Foundations of the Web: Cognition, Communication, and Co-Operation. Towards an Understanding of Web 1.0, 2.0, 3.0. *Future Internet*, 2(1), 41–59. <https://doi.org/10.3390/fi2010041>
- Gordon, C., & İközöğlü, D. (2017). ‘Asking for another’ online: Membership categorization and identity construction on a food and nutrition discussion board. *Discourse Studies*, 19(3), 253–271. <https://doi.org/10.1177/1461445617701810>
- Halberstam, J. (1998). Transgender Butch: Butch/FTM Border Wars and the Masculine Continuum. *GLQ: A Journal of Lesbian and Gay Studies*, 4(2), 287–310. <https://doi.org/10.1215/10642684-4-2-287>
- Hernandez, J. (2022, April 11). U.S. citizens can now choose the gender 'X' on their passport applications. *NPR*. <https://www.npr.org/2022/04/11/1092000203/gender-x-us-passport-applications>
- Kershaw, D., Rowe, M., & Stacey, P. (2015). Language innovation and change in on-line social networks. *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, 311–314. <http://doi.acm.org/10.1145/2700171.2804449>
- Kershaw, D., Rowe, M., Noulas, A., & Stacey, P. (2017). Birds of a Feather Talk Together:

- User Influence on Language Adoption. *Proceedings of the Hawaii International Conference on System Sciences*. <https://doi.org/10.24251/HICSS.2017.225>
- Konnolly, L. (2022). Transmedicalism and ‘trans enough’: Linguistic strategies in talk about gender dysphoria. *Gender and Language*, 16(1), 1–25.
<https://doi.org/10.1558/genl.20230>
- Lin, Z., Salehi, N., Yao, B., Chen, Y., & Bernstein, M. (2017). Better When It Was Smaller? Community Content and Behavior After Massive Growth. *Proceedings of the International AAAI Conference on Web and Social Media*, 11(1), 132–141.
- Leuckert, S. (2020). Rethinking Community in Linguistics: Language and Community in the Digital Age. In B. Jansen (Ed.), *Rethinking Community through Transdisciplinary Research*, (pp. 111-125). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-31073-8_7
- Jackson, S. J., Bailey, M., & Foucault Welles, B. (2018). GirlsLikeUs: Trans advocacy and community building online. *New Media & Society*, 20(5), 1868–1888.
<https://doi.org/10.1177/1461444817709276>
- Jacobsen, K., Devor, A., & Hodge, E. (2022). Who Counts as Trans? A Critical Discourse Analysis of Trans Tumblr Posts. *Journal of Communication Inquiry*, 46(1), 60–81.
<https://doi.org/10.1177/01968599211040835>
- Jenzen, O. (2017). Trans youth and social media: Moving between counterpublics and the wider web. *Gender, Place & Culture*, 24(11), 1626–1641.
<https://doi.org/10.1080/0966369X.2017.1396204>

Jones, L. (2014) “Dolls or teddies?” Constructing lesbian identity through community-specific practice. *Journal of Language and Sexuality*, 3(2), 161-90.

<http://dx.doi.org/10.1075/jls.3.2.01jon>

Jones, L. (2019). Discourses of transnormativity in vloggers’ identity construction.

International Journal of the Sociology of Language, 2019(256), 85–101.

<https://doi.org/10.1515/ijsl-2018-2013>

McIntosh, P. (1998). White privilege: Unpacking the invisible knapsack. In M. McGoldrick (Ed.), *Re-visioning family therapy: Race, culture, and gender in clinical practice*, (pp. 147–152). The Guilford Press. (Reprinted from *Peace and Freedom*, pp. 10-12, July/August 1989. Also reprinted in modified form from *White Privilege and Male Privilege: A Personal Account of Coming to See Correspondences through Work in Women’s Studies*, Center Working Paper 189, 1989)

Mensah, H., Xiao, L., & Soundarajan, S. (2020). Characterizing the Evolution of Communities on Reddit. *International Conference on Social Media and Society*, 58–64. <https://doi.org/10.1145/3400806.3400814>

miles-hercules, d. (2020, August 7-8) *The real T: Language at the intersections* [Conference Presentation]. Second Annual Advancing African American Linguist(ic)s Symposium, Santa Barbara, CA, United States. <https://ucsbhbculing.com/aaal2020/>

Miller, J. F. (2019). YouTube as a Site of Counternarratives to Transnormativity. *Journal of Homosexuality*, 66(6), 815–837. <https://doi.org/10.1080/00918369.2018.1484629>

Milroy, L. (1987). *Language and social networks*. Blackwell.

Ryan, J. (2022, April 17). Sir or ma'am are out – but police can still say 'Evening all' (so Dixon Of Dock Green's OK)! Officers are told to use 'gender-neutral' forms of address instead in new guidance. *The Daily Mail*.

<https://www.dailymail.co.uk/news/article-10724959/Police-officers-told-use-gender-neutral-forms-address-new-guidance.html>

Serano, J. (2007). *Whipping Girl: A Transsexual Woman on Sexism and the Scapegoating of Femininity*. Seal Press.

Sankoff, G. & Blondeau, H. (2007). Language change across the lifespan:/r/in Montreal French. *Language* 83(3), 560–88.

Stolcke, A. (2002). SRILM—an extensible language modeling toolkit. *PROCEEDINGS OF THE 7TH INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING 2*, 901– 904.

Tao, K. (under review). *Claiming Transgender Identity: Contextualising the Linguistic Tension among trans women in Hong Kong* [Manuscript submitted for publication]. School of English, The University of Hong Kong.

Tudisco, J. (2022, April 7-9) “*But... If You Define your Gender through Binary Sexed Characteristics... How Can you Call yourself not Binary?*”: *Linguistic Self-Determination, Gatekeeping and Trans Identities in the French-Speaking Context*. Society for Linguistic Anthropology Conference 2022, Boulder, CO, United States.

Valentine, David (2007). *Imagining transgender: An ethnography of a category*. Duke University Press.

Zimman, L. (2012). *Voices in Transition: Testosterone, Transmasculinity, and the Gendered Voice among Female-to-Male Transgender People* [Doctoral dissertation, University of Colorado at Boulder]. ProQuest LLC.

Zimman, L. (2014). The Discursive Construction of Sex: Remaking and Reclaiming the Gendered Body in Talk about Genitals among Trans Men. In L. Zimman, J. Davis & J. Raclaw (Eds.), *Queer Excursions*. Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780199937295.003.0002>

Zimman, L. (2017). Gender as stylistic bricolage: Transmasculine voices and the relationship between fundamental frequency and /s/. *Language in Society*, 46(3), 339-370.
doi:10.1017/S0047404517000070

Zimman, L. & Hayworth, W. (2020a). Lexical change as sociopolitical change in trans and cis identity labels: New methods for the corpus analysis of internet data. *University of Pennsylvania Working Papers in Linguistics* 25(2), 143-152.

Zimman, L. & Hayworth, W. (2020b). How we got here: Short-scale change in identity labels for trans, cis, and non-binary people in the 2000s. *Proceedings of the Linguistic Society of America* 5(1), 499–513.

Appendix 1

Diagram showing how the first six posts from one time period would be divided up differently into training (blue/left arrow) and test (red/right arrow) sets for each of 1000 LMs. Each of the posts in the test set will get a cross-entropy value, which will be averaged for each LM. Then, the cross-entropies for each LM will be averaged, to get an overall cross-entropy value for the period.

