



## University of Pennsylvania Working Papers in Linguistics

---

Volume 25  
Issue 2 *Selected Papers from NWA47*

Article 17

---

1-15-2020

### Lexical Change as Sociopolitical Change in Trans and Cis Identity Labels: New Methods for the Corpus Analysis of Internet Data

Lal Zimman  
*University of California, Santa Barbara*

Will Hayworth  
*Google*

Follow this and additional works at: <https://repository.upenn.edu/pwpl>

---

#### Recommended Citation

Zimman, Lal and Hayworth, Will (2020) "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*: Vol. 25 : Iss. 2 , Article 17.

Available at: <https://repository.upenn.edu/pwpl/vol25/iss2/17>

This paper is posted at ScholarlyCommons. <https://repository.upenn.edu/pwpl/vol25/iss2/17>  
For more information, please contact [repository@pobox.upenn.edu](mailto:repository@pobox.upenn.edu).

---

## Lexical Change as Sociopolitical Change in Trans and Cis Identity Labels: New Methods for the Corpus Analysis of Internet Data

### Abstract

This paper uses corpus linguistic methods and general purpose computing tools to explore short-scale lexical change in the identity terminology used in an online community for transgender men and other transmasculine people. It focuses on the rapidly changing landscape of labels for trans people, cis people, and non-binary people in a trans community on LiveJournal.com, which was a popular social media venue among trans people in the 2000s. We consider a number of questions about lexical change, including when currently popular forms (e.g. cisgender, non-binary, transmasculine, etc.) were introduced; the decline of labels that have been problematized (e.g. transgendered, transsexual); and the sociocultural discourses that contextualize and account for these changes.

We also describe novel methods for social media data collection, which rely on simple custom software, which we call livecorpus. Livecorpus was built for use with widely-available cloud computing tools, meaning that it is serverless (i.e. does not require the provisioning of the analyst's own servers) and offers flexible configuration that can be modified as data collection progresses. These methods can be applied to other social media sources that are not pre-formatted in ways that facilitate automated analysis, which in practice means we can reach further back into the history of social media-based language use.

While scholars of language variation and change have tended to focus on phonological and morphosyntactic variables in unselfconscious vernacular speech rather than the lexicon, we argue that speakers' awareness of – and metalinguistic discourses about – lexical choices makes this level of language an ideal site for considering linguistic manifestations of sociopolitical change. Far from an unfortunate exception to the normal, non-conscious process of structural linguistic transformation, these types of intentional interventions into lexical usage must be recognized as a critical component of language change.

# Lexical Change as Sociopolitical Change in Trans and Cis Identity Labels: New Methods for the Corpus Analysis of Internet Data

Lal Zimman (he/him) and Will Hayworth (they/them)\*

## 1 Introduction

This paper uses corpus linguistic methods and general purpose computing tools to explore short-scale lexical change in identity terminology for transgender, cisgender, and non-binary individuals.<sup>1</sup> We approach this question using a social media corpus of approximately 17 million words that consists of posts and comments in a community for transgender men and others on the transmasculine identity spectrum on LiveJournal.com that was active in the 2000s.

While studies of lexical variation and change are less common than those on phonological and morphosyntactic variables, we embrace the lexicon as a site of politically motivated language change and agentive linguistic resistance. It is therefore useful to consider how shifts in identity terminology emerge alongside counter-hegemonic political frameworks. We also introduce innovative uses of popular computing tools for creating and querying corpora of data not formatted for automated analysis that can be implemented in studies of other semi-structured textual data.

## 2 The Place of the Lexicon

Despite early attention from dialectologists, modern sociolinguistics has often shied away from lexical analysis in favor of phonological and morphosyntactic variables that demonstrate the patterned regularity of language change in unselfconscious vernacular speech. Because the lexicon is “above the level of awareness” (Silverstein 1981), speakers are often highly conscious of their word choices. The lexicon can also be hard to fit into a variationist framework in which variables have clearly defined variants that can be quantified in terms of occurrence versus non-occurrence.

By contrast, the lexicon has always been central in the study of language, gender, and sexuality, and its status as above the level of awareness is an asset rather than a limitation. The lexicon encodes misogynistic, homophobic, and transphobic meanings referentially, as well as indexically, but it also allows for the intervention of those who wish to challenge such ideologies. This analysis supplements qualitative queer linguistics research on the use and re-signification of identity terminology (e.g. Chen, 1998, Wong 2005) by quantitatively tracking changes in their frequency of use.

Corpus sociolinguistics (Baker 2010) provides powerful tools for quantitative analysis of the lexicon, but existing corpora are limited in the varieties and communities represented and information provided about the context of the data. There is a high cost associated with the creation of corpora, particularly when it is comprised of data structured in ways that are not ideal for automatic parsing. We therefore share our methods for creating and using *livecorpus*, a crawling pipeline we used to create the TransLiveCorpus from contributions to a trans community on LiveJournal.com.

## 3 The Study, Community, and Platform

This study is focused on lexical variation in a corpus of public posts and comments from a trans community on LiveJournal. This community was selected because of the first author’s prior digital ethnographic fieldwork on metalinguistic commentary and talk about the body in the same group (Zimman 2014). We take a community of practice-based approach (Eckert & McConnell-Ginet 1992), in which the data source is defined not by speakers’ demographic categories but by their joint activity: in this case, talking about the experiences of transmasculine people. We make no attempt to ascertain the gender identity of users, and it is clear that membership in this group includes partners, family members, friends, and allies. Importantly, all who participate are held to the same

---

\* Thanks to Google for providing trial credits for this analysis.

<sup>1</sup> While transgender people have gender identities different from the genders assigned to them at birth, cisgender people are those who identify with their assigned gender. Non-binary people are those who do not identify as exclusively female or male. Many, but not all, non-binary / genderqueer people also identify as trans.

standard when it comes to community-sanctioned terminology. Whether the offending party is trans or cis, using language judged to be problematic often leads to direct confrontation and, occasionally, spectacular conflict. Community members tend to share a common orientation to the importance of language in fighting transphobia, making it an ideal site for the investigation of lexical change as an element of sociopolitical change.

The community of study is a group for “FTMs,” or female-to-male individuals,<sup>2</sup> on the blogging site LiveJournal.com, which also hosts communities in which members post and receive comments from others. Though no longer in widespread use, LiveJournal is a useful source of online interactions among members of communities of practice prior to the rise of current social media giants. LiveJournal was particularly popular among trans people in the 2000s, and many of the discourses and terminological norms that have risen to some prominence – e.g., the preference for *transgender* over *transgendered* – can be seen in early form here. There was a major drop in traffic on LiveJournal starting in around 2007. Figure 1 depicts the number of posts and comments during the lifetime of the transmasculine community under study, with each dot representing one month.

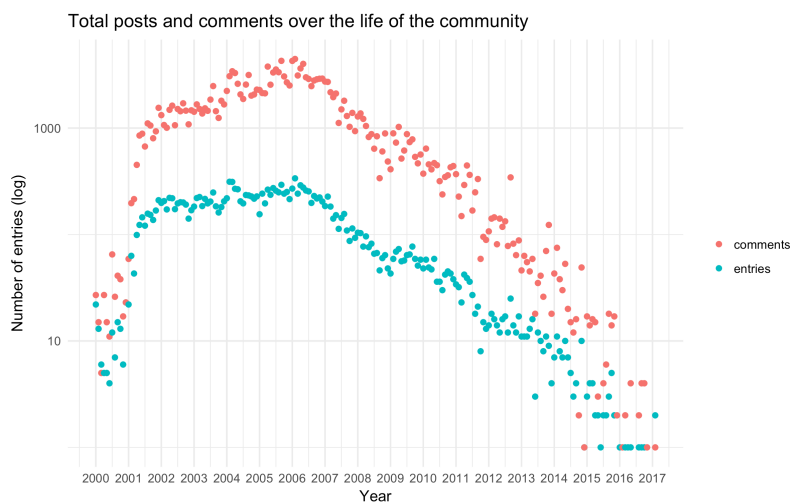


Figure 1: Number of posts and comments to the LiveJournal community over time.

Activity in the community was at its peak between 2002 and 2006. The sharp decline that can be observed from 2007 onward may be associated with the emergence of other social network platforms such as Twitter in 2006, Tumblr in 2007, and Facebook, which opened to non-college students in 2007. The TransLiveCorpus analyzed here contains approximately 20,000 posts, each with anywhere from 0 to 290 comments (11,382 in all), for a total of approximately 17 million words.

## 4 Methods

### 4.1 The Corpus

To collate our data, we used Google’s Cloud Platform (GCP) to build *livecorpus*, a crawling pipeline for LiveJournal (see Figure 2). The crawler is written in Python and runs on App Engine, which requires minimal setup and supplies computing capacity as needed. When a user sends a request to *livecorpus* to scrape a particular community, it fetches HTML from LiveJournal’s servers and parses it, extracting both content and metadata. The crawler reads pages listing entries and enqueues the individual entries’ links in Cloud Tasks, which dispatches them to the crawler to be processed. Using task queues allows us to ensure that entries are processed successfully and limits the rate of our crawl to comply with LiveJournal’s bot policy, which permits no more than 5 connections per second.

<sup>2</sup> Despite being for “FTMs”, in practice the group included many members who would today likely identify as non-binary. At the time, *FTM* sometimes functioned as an umbrella category like *transmasculine*.

After the entries and their comments have been successfully scraped, they are stored in a NoSQL document database called Cloud Firestore.<sup>3</sup> Document databases do not require a schema to be defined before usage, making it easy to get started and progressively iterate (i.e., to scrape and store more fields as you go). Cloud Firestore is also hierarchical, which means collections have documents which have sub-collections of documents, matching LiveJournal’s data structure (community → entries → comments).

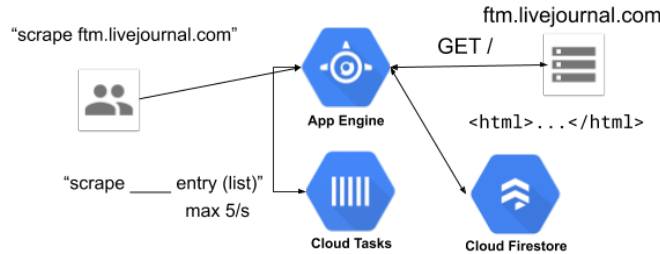


Figure 2: The *livecorpus* crawler.

Although Cloud Firestore provides a convenient repository for storing our parsed results, it does not support the types of queries we needed to run for our analyses. For this, we used GCP’s analytical data warehouse, BigQuery. Figure 3 illustrates this process. Cloud Firestore’s managed export feature writes files to Cloud Storage which BigQuery can import. BigQuery executes our SQL code and returns results that can be converted to CSVs or Google Sheets. To look for particular terms, we used regular expressions in combination with string functions like REGEXP\_CONTAINS, which matches those expressions to fields in the database.

There are several things that make the *livecorpus* method distinctive. First, *livecorpus* is *serverless*: none of the resources our crawler uses – App Engine for code execution, Cloud Tasks for work distribution, Cloud Firestore for storage, BigQuery for analysis – requires provisioning or configuring servers. We can optionally modify parameters to trade off throughput for cost once the application is running. This high-level approach means that the Google shares resources on our behalf (multitenancy); we pay only for the computing and storage we actually use. Crawling the transmasculine community and performing data analysis cost about \$20 over a couple of months.

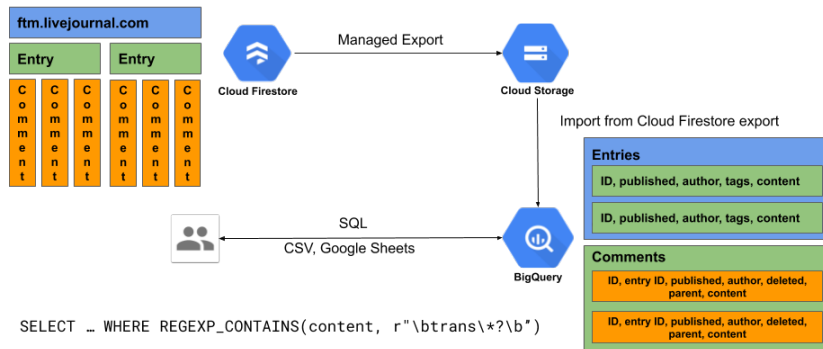


Figure 3: The querying process.

One downside of our implementation is that it is not trivial to run on another cloud; it is optimized for Google’s infrastructure, not platform independence. Fortunately, the underlying architecture is portable and extensible. *Livecorpus* is also built entirely on “general purpose” programming languages (Python, SQL) and cloud-computing services, which are used in many disciplines for diverse applications. Support and functionality extension via libraries is easier for widely used tools than for more specialized and discipline-specific ones. Additionally, these components scale automatically, so the corpus can grow from megabytes to terabytes with minimal intervention.

<sup>3</sup> The second author is part of the Cloud Firestore team at Google.

## 4.2 Querying and Coding TransLiveCorpus

The lexical items in our queries of the TransLiveCorpus for the analysis below fall into three categories of widely used identity labels. The first, terms for trans people, include *transgender*, *transgendered*, *trans*, *trans* + [group], *transgenderist*, *transsexual*, *transfeminine*, and *transmasculine*. The *trans* + [group] category consisted of the word *trans* followed by one of the following words for a gender category or person in general: *man*, *woman*, *male*, *female*, *boy*, *boi*, *girl*, *guy*, *chick*, *dude*, *fella*, *folk*, *people/person*, and *individual*. In the second group, words for cis people, parallel terms were analyzed with *cis* in place of *trans*, except where no parallel option exists: *cisgender*, *cisgendered*, *cis*, *cis* + [group] (with the same groups), and *cissexual*. The low frequency of cis- words allowed us to search for all words beginning with *cis*, which resulted in a residual “other” category consisting of words like *cissexed* and *cissy* (a tongue-in-cheek insult for a cis person). The non-binary terms investigated were *agender*, *bigender*, *genderfuck*, *genderfluid*, *genderqueer*, *non-binary*, and *polygender* (see section 5.3 for definitions). A regular expression was used to identify the desired forms, their plural and possessive derivations, common alternate spellings, and variants in the *trans/cis* + [group] categories in which morphemes are separated by spaces (e.g. *trans man*), nothing (*transman*), hyphens (*trans-man*) or stars (*trans\*man*). Only forms that referred to people were included; forms referring to abstract concepts and other non-human referents (e.g., *transgenderism* and *cissexism*) were excluded.

Each token was coded for several factors, many derived from crawled data, including whether it appeared in an entry or a comment, the ID of the relevant entry or comment, the username of the author, the date of use (binned by month); the referential category (i.e., trans, cis, or non-binary people), the lexical category (e.g., *trans* + [group]), and the form used (e.g., *transmen*).

## 5 Analysis

### 5.1 Terms for Trans People

We begin with an analysis of identity labels for trans people, with a focus on three questions:

- (1) What are the most common terms to refer to trans people, and how did this change over time?
- (2) What is the distribution of the terms *transgender* and *transgendered*?
- (3) What is the distribution of *transgender* or *transgendered* versus the short form *trans*?

Table 1 contains the frequency of the labels for trans people we queried from the TransLiveCorpus. The most popular way of referring to trans people over the life of this community is the shortened form *trans*, which was used primarily as an adjectival predicate and accounts for 46.2% of all tokens. The second most popular lexical category is *trans* + [group], which accounts for 32.5% of tokens. Together, these realizations of the shortened form, *trans*, constitute the great majority (78.7%) of the tokens. The full form, *transgender*, is less common (8.6%), and the full form with the -ed ending, *transgendered*, is even rarer (3%). *Transsexual* is about as frequent as *transgender* (9%). *Transfeminine*, *transmasculine*, and *transgenderist* each constitute less than 1% of the tokens. Figure 4 introduces a time dimension by plotting the frequency of lexical categories according to their date. Time is represented on the x-axis, while the number of tokens is represented on the y-axis, with different colors indicating different lexical categories. Each bar represents six months. Figure 4 shows a drop in frequency for all terms starting in 2007, when overall LiveJournal traffic began to decline. This decline makes it difficult to see changes in relative frequency over time, so Figure 5 presents the same data as a percentage of the total number of tokens. In Figure 5, several patterns are evident.<sup>4</sup>

Category	Number of tokens	Percentage of total
<i>trans</i>	23,393	46.2%
<i>trans</i> + [group]	16,434	32.5%
<i>transfeminine</i>	22	<1%

<sup>4</sup> Note that the final two years represented in the four rightmost histogram bars had very few posts, which explains the seemingly extreme changes in the distribution of forms across the rightmost bars of Figure 6.

<i>transgender</i>	4,365	8.6%
<i>transgendered</i>	1,532	3%
<i>transgenderist</i>	20	<1%
<i>transmasculine</i>	273	<1%
<i>transsexual</i>	4,547	9%
Total	50,586	

Table 1: Frequency of terms for trans people.

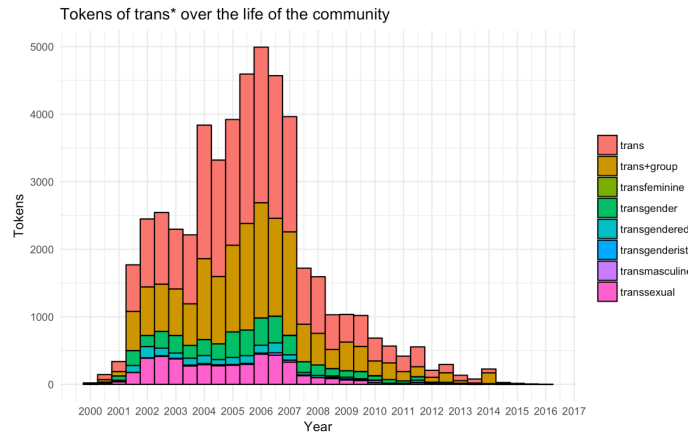
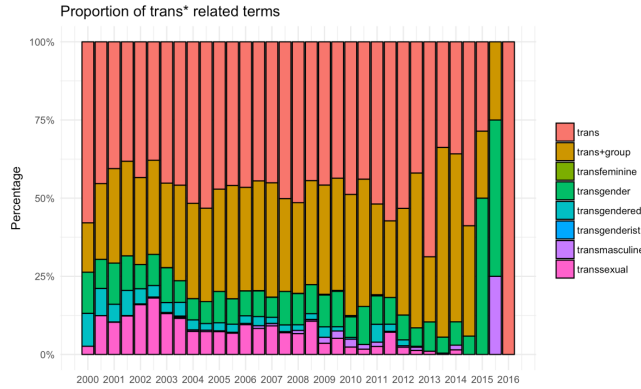


Figure 4: Terms for trans people over time.

First, it is clear that the short form *trans* (in red/the topmost chunk of the histogram bars) is the most popular of the terms we queried, especially when combined with the *trans* + [group] category (in mustard/the second topmost chunks). Second, there is a downward trend in the frequency of *transsexual* (in pink/the bottommost chunks), which reflects the way many trans people have coalesced around the problematization of this word on the basis of its association with pathologized models of trans identity and the associated demands for gender normativity (Stryker 2008). Though *transgendered* (in blue/the second bottommost visible chunks at the left end of the histogram) was never all that popular in this community, Figure 6 shows a similar decrease in frequency for the word starting in 2003 and again in around 2010. *Transgendered* has also been rejected by many trans people, though for more nebulous reasons than *transsexual*. *Transgendered* was widely used in trans communities throughout the 1990s and into the 2000s, as these data confirm, but today a common language ideology in trans communities is that the -ed ending makes it sound like being trans is something that happens to the person, as if one “gets transgendered” by an external force that renders one “a” transgender, rather than a characteristic. As *transsexual* and *transgendered* decreased in frequency, the short forms *trans* and *trans* + [group] increased in use. In fact, one potential benefit of the short form is that it obviates the need to choose between *transgender*, *transgendered*, and *transsexual*. These data allow us to identify key time periods for changes in group-internal norms for identity labels. Finally, in purple we can see the emergence of *transmasculine* as an umbrella label for anyone who has assigned female at birth but who identifies as male or masculine. The term first appeared in 2002 but didn’t become popular enough to be visible on the plot until around 2008. *Transfeminine* occurs in the data as well, but with fewer tokens, which makes sense given the focus of the community. *Transmasculine* has since replaced older umbrella terms like *FTM*. Also included in this plot is *transgenderist* (not visible in figure), which was popular in the

1990s but only occurred 20 times in the corpus, mostly before 2006. While *transsexual* has continued to be a common term of identification, *transgenderist* has not.

Figure 5: Terms for trans people over time as percentage.



On the basis of these observations, statistical analysis was performed to test the hypothesis that forms using the -ed ending (*transgendered*, n = 1,532 [26%]) decreased in usage over the lifespan of the community relative to the plain form (*transgender*, n = 4,365 [74%]). Figure 6 visualizes the relative frequency of the two terms over time, with the -ed ending in red and plain form in blue. During the first few years of the community’s existence, *transgendered* was just as common as *transgender*, but by late 2013 the form with the -ed suffix had disappeared from this group. Table 2 contains the results of a linear mixed-effects regression used to model the relationship between time (binned by month) and the presence of the -ed ending. Random effects included the author and the number of posts made that month, the latter to account for changes in overall community traffic. The regression shows a main effect of date for both forms. The fact that both variants decreased in frequency seems to relate to the increasing popularity of the short forms, *trans* and *trans+[group]*. However, the effect size for the decrease in the -ed form is almost twice that of the plain form, indicating a more dramatic decrease in usage of *transgendered* than *transgender*.

	Estimate	Standard error	df	t value	P(> t )
(Intercept)	90.0467	3.8543	173	23.363	<2e-16 ***
edending-ed	-0.4165	0.1199	50410	-3.474	0.000513 ***
edending-plain	-0.2247	0.0733	50410	-3.006	0.002173 **

Table 2: Results of mixed-effects linear regression of *transgendered* vs. *transgender*.

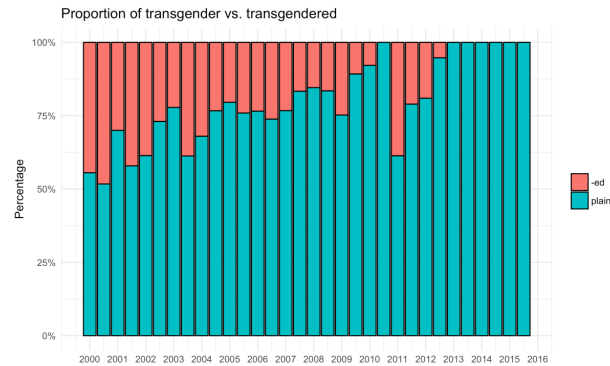


Figure 6: Comparison of -ed ending (*transgendered*) vs. plain form (*transgender*).

5.2 Terms for Cis People



The next set of terms we consider are those for cis people. *Cisgender* is a word for non-trans individuals derived from the Latin *cis-* ('on the same side'). Here too, we consider three questions:

- (1) When did the word *cis(gender)* emerge?
- (2) Does *cisgender* vs. *cisgendered* follow the same pattern as *transgender* vs. *transgendered*?
- (3) Does *cissexual* follow the same pattern as *transsexual*?

*Cisgender* was coined to mark the unmarked category, creating a label for a group whose identities had been so naturalized that no term was seen as needed. Without this term, cis people are often referred to as *biologically (fe)male*, which naturalizes cis identities and treats sex as fixed at birth, or simply called *(wo)men* while trans people are *trans (wo)men*, implying that trans people do not really count as members of their self-identified genders. Increasingly widespread familiarity with *cis* is a success story for trans linguistic activism, but its history is not well-documented.

Table 3 provides the frequency of labels in the TransLiveCorpus that implement the *cis-* affix. One similarity between these data and the words for trans people discussed in section 5.1 is that *cis* and *cis* + [group] together comprise a plurality of the tokens, though stand-alone *cis* (12.2%) is far less frequent than stand-alone *trans* (46.2% of trans tokens). The shortened form of *trans* may be preferable in part because of its salience and frequency of use, as well as its relative age compared to *cis*. One difference from the trans data is that the -ed form, *cisgendered*, is more common (26.4%) than *transgendered* (3%). Like *transgendered*, however, *cisgendered* decreased in frequency of use.

Category	Number of tokens	Percentage of total
<i>cis</i>	258	12.2%
<i>cis</i> + [group]	762	36.1%
<i>cisgender</i>	448	21.3%
<i>cisgendered</i>	557	26.4%
<i>cissexual</i>	74	3.5%
Other	9	<1%
Total	2,108	

Table 3: Frequency of terms for non-trans people using the *cis-* affix.

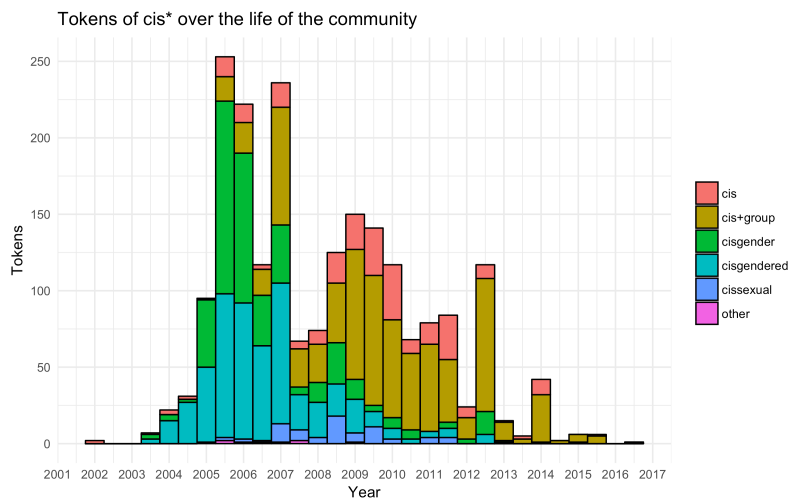


Figure 7: Terms for cis people over time.

Figure 7 shows that *cis* did not begin to appear in this community with much frequency until 2004, with only two tokens of *cis* (from the same speaker) prior to 2003. *Cis* and its derivations became much more frequent in late 2005 and 2006. There is also a less extreme drop-off after 2007 than in the terms for trans people, suggesting that *cis* was increasing in use around this time.

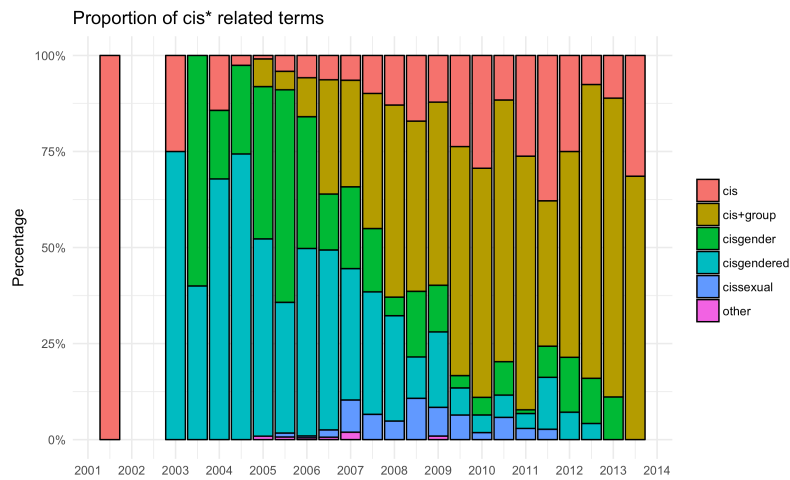


Figure 8: Terms for cis people over time as percentage.

Figure 8 offers a few additional insights. Even as *transsexual* decreased in use over time, *cissexual* first appeared in 2007 and declined after only a few years. This may be attributable to an influential book (Serano 2007) that advocated for the use of *cissexual* to refer to people who do not change the sexual characteristics of their bodies, regardless of how they identify (meaning some transgender people are cissexual, others transsexual). While popular for a time, Serano's distinction did not maintain. As with terms for trans people, the rise in the short forms, *cis* and *cis* + [group], obscures distinctions between different types of long forms (*cisgender*, *cisgendered*, *cissexual*).

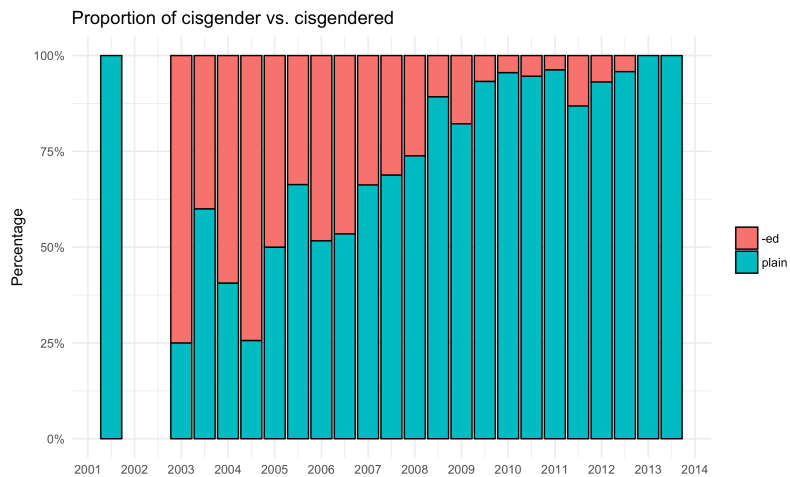


Figure 9: Comparison of -ed ending (*cisgendered*) vs. plain form (*cisgender*).

A plot of just the distribution of *cisgender* and *cisgendered*, which appears in Figure 9 shows a similar, yet slightly different, pattern than *transgender(ed)*. *Cisgender* was still popular as *transgendered* began to decrease in frequency: from 2003-2005, *cisgender* consisted of more than half of the total number of the *cisgender/cisgendered* tokens, while during the same period *transgendered* at times consisted of fewer than 25% of the *transgender/transgendered* tokens. Overall, though, both words indicate a disfavoring of the -ed ending.

### 5.3 Terms for Non-binary People

Our final set of questions concerned terms for non-binary people during a time with far less non-binary visibility, even within trans communities. The questions of focus in this section are:

- (1) When did the term *non-binary* emerge?
- (2) What is the distribution of *non-binary* versus other words for identities outside the binary?

These questions were informed by the observation that *non-binary* is a relatively new umbrella label, replacing *genderqueer* as the dominant term for anyone outside the gender binary. *Genderqueer* is still sometimes used in this way, but also as a term for those who emphasize the distinctly queer nature of their genders. Two other long-standing terms were also included for analysis: *genderfuck* (used in reference to confrontational gender non-normativity) and *genderfluid* (moving between genders across contexts or over time). Several newer terms were also included: *agender* (not identifying with any gender category), *bigender* (identifying with both binary genders), and *polygender* (identifying with multiple genders, but not necessarily the binary ones).

Plotting the non-binary data using the same schema as in section 5.1 and 5.2 presents a challenge because of the popularity of *genderqueer*, which at 91.6% of the 3,288 tokens of words for non-binary people, dominates plots renders other words invisible. Tokens of *genderqueer* have been removed from Figure 10, which shows change over time for all other terms for non-binary people.

Category	Number of tokens	Percentage of total
<i>agender</i>	19	<1%
<i>bigender</i>	14	<1%
<i>genderfluid</i>	63	1.9%
<i>genderfuck</i>	158	4.8%
<i>genderqueer</i>	3,014	91.6%
<i>non-binary</i>	16	<1%
<i>polygender</i>	4	<1%
Total	3,288	

Table 4: Frequency of terms for non-binary people.

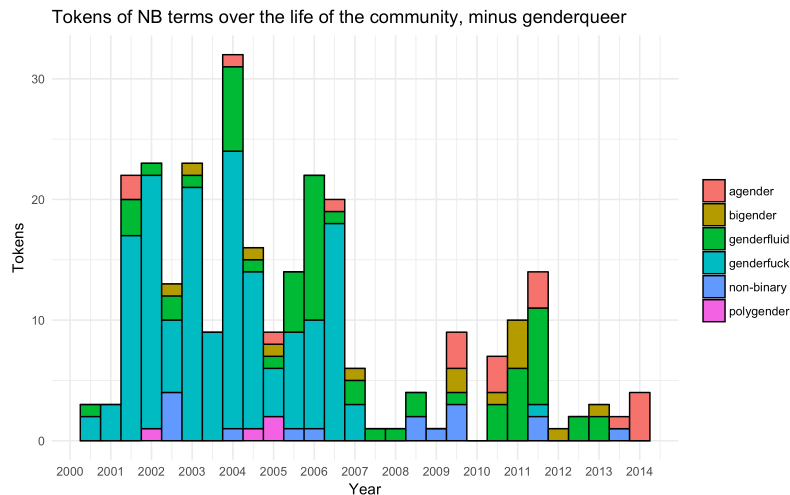


Figure 10: All terms for non-binary people (other than genderqueer).

Several observations can be made about the terms that are visible in Figure 10. First, *genderfuck* (in light blue) disappeared after 2007 despite being the most common item in the set before that year. More specific labels, including *agender* (in red), *bigender* (in mustard), *genderfluid* (in green), and *polygender* (in pink), appear as early as 2001-2002, though *agender* and *bigender* both have stronger representation in later years. *Polygender* only appears a few times in the corpus.

The word *non-binary* first occurs in late 2002, but was not common at any point in this corpus.

We had expected *non-binary* to be increasing in use during this time, but there are a few potential explanations for this unexpected result. The first, and simplest, is that *non-binary* was not yet being used. This resonates with the first author's experience in the field; as recently as 2012, research participants who now identify as non-binary were describing themselves as genderqueer. Another possibility is that *non-binary* was being used, but not in this male-dominated community: perhaps non-binary people were forming independent groups on LiveJournal at this time. A final explanation is that non-binary people were leaving LiveJournal in favor of other social media platforms, such as Tumblr (launched in Feb. 2007), which quickly became associated with trans people and politics.

## 6 Conclusions and Future Directions

As technologically-mediated language becomes a more significant target of sociolinguistic inquiry, computational and corpus methods are essential. Social media data provides a rich source of data for short-scale linguistic change, but many online data sources require significant transformation prior to analysis. This paper illustrated the use of livecorpus, software that uses general purpose computing tools for the creation and analysis of custom corpora from semi-structured textual data. This approach allowed us to analyze older social media data than might otherwise be possible and hence cover more of the rapidly changing landscape of gender-related identity labels, which is a central part of trans people's linguistic activism. Despite the challenges involved, the lexicon offers a rich domain for investigating language change, the complexities of which are rendered more workable with computational tools. Rather than being a shortcoming of lexical analysis, the fact that the lexicon is open to agentive intervention makes it a critical ground on which sociopolitical change takes place. The politically charged nature of language demands that we consider the role of agency, and not just linguistic structure, in the process of language change.

## References

- Baker, Paul. 2010. *Corpus Sociolinguistics*. Edinburgh: Edinburgh University Press.
- Eckert, Penelope & Sally McConnell-Ginet. 1992. Think practically and look locally: Language and gender as community-based practice. *Annual Review of Anthropology* 21:461–490.
- Serano, Julia. 2007. *Whipping Girl: A Transsexual Woman on Sexism and the Scapegoating of Femininity*. Emeryville, CA: Seal Press.
- Silverstein, Michael. 1981. The limits of awareness. *Working Papers in Sociolinguistics* 84:1–30.
- Stryker, Susan. 2008. *Transgender History*. Berkeley, CA: Seal Press.
- Wong, Andrew D. 2005. The reappropriation of tongzhi. *Language in Society* 34:763–793.
- Zimman, Lal. 2014. The discursive construction of sex: Remaking and reclaiming the gendered body in talk about genitals among trans men. In *Queer Excursions: Retheorizing Binaries in Language, Gender, and Sexuality*, ed. L. Zimman, J. L. Davis & J. Raclaw, 13–34. New York: Oxford University Press.

Lal Zimman  
 Department of Linguistics  
 University of California, Santa Barbara  
 Santa Barbara, CA 80306-3100  
 zimman@ucsb.edu

Will Hayworth  
 Google  
 345 Spear Street  
 San Francisco, CA 94103  
 wsh@google.com