2025. Proc Ling Soc Amer 10(1). 5907. https://doi.org/10.3765/plsa.v10i1.5907.

When "anonymous" is not enough: methodological issues and the safety of human subjects in social media research

Allison Casar

Abstract. Drawing on data from a previous study on slur reclamation practices on Twitter/X, as well as scholarly discussion of context collapse and digital research ethics, I discuss the need to (re)evaluate how scholars engage with, publish, and present searchable language data online.

Even when a subject's social media persona is not linked –by name, location, and other identifying information –to their offline self, the distinction between the two is increasingly thin. Harm done to a person online –through harassment, dogpiling, suicide-baiting, other emotional abuse, and doxxing, among other tactics –is also harm done to their offline self. This is an especially salient risk for social media users from vulnerable or marginalized communities. I argue that stricter methodological and ethical standards should be established for research on social media language data, and present strategies myself and others have used (in various combinations) to tackle this issue: quotation with informed consent; discourse tallying; data aggregation; and focus on (in)famous public figures and organizations. I discuss the drawbacks and advantages of these methods, supplying examples from my work on metalinguistic attitudes towards slur reclamation on Twitter/X.

Keywords. research ethics; social media research; digital humanities; internet linguistics

1. Introduction. As linguists increasingly turn to social media for language data –and as social media landscapes continue to shift and change –conversations about research ethics must occupy a more central and dynamic place in the field. Halford (2018) points out that social media research presents a series of pressing disruptions to the norms of privacy and informed consent established by ethics committees and review boards. I argue that methodological creativity, greater safeguards to research subject privacy –especially in cases of searchable text data –and flexible, reflective research ethics are key to establishing a more just research paradigm in the era of big data.

In this paper, I discuss ethical issues inherent to the ways in which we as linguists engage with social media data, then offer sample approaches drawn from my own research on metalinguistic stances towards slur reclamation on Twitter/X[1]. My aim is not to critique individual scholars or studies, the majority of whom are working in compliance with institutional review boards or their local equivalents. Rather, I encourage discussion on best practices moving forward. To that end, I first review literature on social media research ethics, highlighting issues which are particularly salient to language research, and then offer my own experiences navigating privacy concerns while analyzing a corpus of tweets containing potentially sensitive content.

2. Literature review As Williams et al. (2018) note, ethical concerns in social media research fall into three broad and non-exclusive categories: privacy, confidentiality, and concerns about

[1] This research began before Twitter was renamed X, and for the sake of simplicity I continue to refer to "Twitter," "tweets," etc throughout.

Author: Allison Casar, University of Illinois Urbana-Champaign, IL (casar2@illinois.edu)

consent to data use and study participation. In this section, I give a brief survey of existing literature on ethics in online language research across a variety of disciplines. It's important to note at the top that many of these papers are, in some way or another "out of date" –which is to say that, while their general ethical principles and insights into user and researcher behavior remain extremely valuable, the exponential speed of cultural and technological change in the last decade means that specifics about communication norms, platform terms of service, platform popularity, and even human-technology interaction norms are constantly shifting. For a brief but insightful overview of the history and importance of research ethics in the age of digital communication and big data, see Samuel & Buchanon's 2020 editorial introducing a special issue on social media research ethics in the *Journal of Empirical Research on Human Research Ethics.*

2.1. PRIVACY, CONFIDENTIALITY, AND CONSENT The question of whether social media should count as a "public" or "private" space for the purpose of research ethics is as old as social media, and has no simple answer. Ahmed et al, writing for the edited volume *The Ethics of Online Research* (Halford 2018), note that there are differences in levels of privacy expected even between platforms. On Twitter, for example, shared content is considered public by default. Unless a user chooses to make their content private, it is publicly accessible via Twitter's API and search function. In contrast, "the majority of Facebook is considered private" on both a data-sharing and user experience level (Ahmed et al 2018: 86).

However, even a nominally public post may not be intended for all audiences, or (especially) for a research audience. Marwick & boyd (2011) note that "social media collapse diverse social contexts into one," a phenomenon aptly known as "context collapse." This makes it harder for social actors online to "engage in the complex negotiations needed to vary identity presentation, manage impressions, and save face." It flattens complex attitudes and richly varied social identities into a 280-ish character snippet that persists, beyond the moment of its creation, until deleted. It means that speech meant for a specific context and imagined audience – something as broad as "my followers" or are narrow as "other LGBTQ+ people currently watching season 4 of my favorite show" –might be received in unintended ways by audiences outside of the intended social or chronological context. I have previously argued (Casar 2024) that this can be a major cause of conflict online; it is also something that researchers absolutely *must* take into account if they want to treat social media speech as "public" data. Most of us can imagine something that we have posted online that we never intended to see carefully analyzed in an academic journal or on the big screen at a linguistics –or anthropology, or sociology, or epidemiology, or psychology, etc –conference.

Users' expectations of social media privacy also differ cross-culturally, between age groups, between gender and sexual orientations, and over time. Williams et al (2018; same volume) conducted two exploratory studies on social media users' understanding of and attitudes towards the use of their posts and content in academic research. The first study was conducted in 2013 with the National Centre for Social Research in the UK (see also Beninger et al 2014), and included a total of 34 participants –low, medium, and high social media users –in focus groups and paired interviews. The second study, conducted in 2015 with Cardiff University's Social Data Science Lab, comprised an online survey of 564 Twitter users. While attitudes have likely changed since these studies, they still offer a reminder that the attitudes about "public" and "private" information and the understanding of academic research with social media data varies significantly between researchers and social media users, and between different users.

Williams et al. found that, across both studies, participants showed "showed a general lack of concern from social media users over their information being used for research purposes (with university research attracting least concern)" but stress that "the majority of respondents stated that they would want to be asked for consent and to remain anonymous in publications reporting social media research." About 80% of respondents in the 2015 study, for example, expressed at least some desire to be "contacted for informed consent" before academic publications of their posts. About 90% expressed a desire for anonymity, an attitude which was strongest among female social media users, social media users of color[2], and users who posted photographs. Lesbian, gay, and bisexual respondents were more likely that heterosexual respondents to have concerns about their posts being used for government or commercial interests, but not for academic interests (Williams et al. 2018: 34). These findings support an understanding of privacy and control of personal data as particularly important to members of marginalized or historically-surveilled populations and communities.

Attitudes towards consent were also related to participants knowledge of the Twitter terms of service (with more knowledgeable users being less likely to expect informed consent) and towards participants attitudes towards "self-regulation" of online speech (with more "self-regulating" participants having less expectation of informed consent). Platform also played a role:

> *"websites with a 'social' purpose were thought to contain more 'personal' content, whereas content posted to 'professional' sites was less so. In light of this, participants thought that it would be acceptable for researchers to access the latter without gaining consent because the risks associated with being identifiable through personal information are lower." (Williams et al: 30)*

Methodological approaches to consent in social media research vary. In the case of studies that draw on a large quantity of posts –such as Ahmed et al.'s (2019) work on public, online information sharing during the 2009 Swine Flu outbreak –it can be impossible to gain consent from every individual represented in a 200,000+ item dataset. Smaller studies that deal with hostile participant populations –such as the "pick up artists" discussed in Rüdiger and Dayer 2017 –can also pose significant problems for the consent process (see further discussion of this in Section 3.3, below). However, some researchers and many social media users argue that social media data should never be used without some form of user permission, and informed consent remains at the core of review board requirements. There is an overall "lack of consensus in the published literature…regarding whether or not consent is required for [social media] research" (Samuel and Buchanon 2020: 6).

Researchers who seek informed consent for social media research may also find the consent process different from an in-person study. Salmons (2018) discusses issues such as the comprehensibility of participant information documents –especially when the researcher is not in the room to explain them –and the negotiation of what types of data –messages, posts, images, etc. –a participant will allow researchers to use. She also notes that, for participants accustomed to quickly scrolling and clicking through online terms and conditions documents, "informed" consent may be more rote than informed. She offers several case studies of consent processes in online research (Salmon 2018).

2.2: CONTEXT AND REFLECTIVE RESEARCH PRACTICES Cognizance of online research venue and consideration for user expectations of privacy –as shaped by social, cultural, legal, and platform

[2] Williams et al use the designation "BME" or "Black and Minority Ethnic" in their report

norms –are key components of the approach to ethical research advocated in the third revision of the Internet Research Ethics (IRE 3.0) guidelines published by the Association of Internet Researchers (AoIR). Rather than setting out top-down rules for ethical internet research, the IRE advocates a "process approach [that] is first of all *reflective* and *dialogical*[3] as it begins with reflection on own research practices and associated risks and is continuously discussed against the accumulated experience and ethical reflections of researchers in the field and existing studies carried out" (franzke et al. 2020).

The AoIR guidelines exemplify a "situational" approach to ethics, one which requires attention to the unique context and needs of each research project. This approach is *reflective* and *dialogical* in that it requires researchers to constantly consider and reconsider the ethics of a study. It requires close attention to the research site –i.e.: Twitter or Facebook –and the context in which language data are produced, circulated, gathered, de-identified (if applicable) and published. Many professional organizations and learned societies[4] have produced guidelines for internet research that take a broadly situational approach, though this is not always the case for Institutional Review Boards (IRBs)/Research Ethics Committees (RECs).

Samuel and Buchanon (2020) discuss the findings of two studies (Hokke et al 2020 and Sellers et al 2020) of Australian and UK review boards that there is an overall lack of "personal and professional experience of [social media]" and a lack of knowledge of expert guidelines among REC members. This is perhaps unsurprising, as the speed of technological advances and the constant advent of new tools with which to access social media data (and new ways to use data to glean insights into human subjects) can make it difficult for review boards and researchers –especially those whose primary fields are not computational –to stay on top of them. For further discussion of the "ethical disruptions" that social media data present for traditional IRB/REC review, see Halford (2018).

While institutional review remains a backbone for research ethics, I argue that linguists ought also to turn a reflective ethical eye to their research processes. Rather than treating ethical review as "a "one-off" tick-box exercise that is primarily an obstacle to research" (franzke et al. 2020: 4) we will need to reflect on ethical best practices consistently and iteratively throughout the entire research practice, potentially adjusting and reevaluating as we go.

3. Sample approach: *queer* on Twitter. In this section, I discuss my approach to participant privacy when gathering and analyzing a corpus of tweets. I reflect on the strengths and drawbacks of my approach, not as a template for future studies, but as an example of the self-reflective work that I believe is necessary for linguists to continue to work ethically with social media data.

The data discussed this this section come from a small corpus of posts scraped from the Twitter API over a one-month period between mid-July and mid-August 2023. These posts were analyzed as part of my previous research on the word *queer*, its historical trajectory as a reclaimed slur, and the discursive and pragmatic function that its use –or avoidance –serves in contemporary LGBTQ+ discourse. This research included both in-person interviews with self-identified LGBTQ+ people and mixed-methods analysis of social media data, but only the social media component will be discussed here.

---

[3] Emphasis in the original.

[4] Among them the AoIR, the British Psychological Society (BPS), the British Society of Criminology (BSC), the British Educational Research Association (BERA), the European Society for Opinion and Market Research (ESOMAR) (Williams et al 2018).

The Twitter analysis focuses on the word *queer* and the avoidance term *q-slur.* The goal of this analysis was to compare how each term was used on Twitter, and in conjunction with what stances towards LGBTQ+ identity and community.

3.1 STUDY METHODOLOGY Tweets were pulled from Twitter's API using keyword searches for *queer* and *q-slur*. *Q-slur* was chosen as a search term because it is an easily-searched and—in my personal experience with LGBTQ+ social media communities—relatively common avoidance term for queer. Other common avoidance terms and textual censorship strategies use asterisks in place of specific letters in the word, as in *q***r* or *qu**r*. These terms are extremely difficult to search using Twitter's API, because the asterisk is interpreted in the search as part of a regular expression.

Tweets using these keywords were pulled from the Twitter Application Processing Interface (API) using the tweepy package for Python (Roesslein 2023) and imported into R for data cleaning using the tidyverse package for R (Wickham et al. 2019). Tweets were organized into tidy dataframes (tibbles). All tweets in languages other than English, as well as several tweets whose text had been cut off or corrupted somewhere in the collection process and could not be manually reconstructed by searching their tweet ID, were removed from the corpus at this stage. I then manually removed all duplicates of tweets, including those that had been queried multiple times from the API and retweets that appeared multiple times in the corpus. The final corpus contains 479 unique tweets: 450 with the keyword queer (List 1) and 29 with the keyword q-slur (List 2). Simple sentiment analyses—using a modified version of Hu & Liu's opinion lexicon (2004) – and frequency analyses were carried out on each list independently. Common English stop words, such as *I, me, the, and* were removed for frequency analysis alone.

Finally, I used a random number generator to select 20 tweets from each list on which to perform an iterative discourse analysis. These tweets were imported as two separate datasets into Taguette, an open-source tool for tagging and analyzing qualitative data (Rampin et al., (2021). For each dataset I went through four rounds of analysis, tagging discourse topics as follows:

- Round 1: Tag creation. Tags were emergent from the data, rather than pre-constructed.
- Round 2: Tag consolidation. On the second round of analysis, tags with the same or similar themes (ie: "celebrity" and "popular media") were consolidated.
- Rounds 3 and 4: Reevaluation. Tweets were re-read to make sure that the tags created in rounds 1-2 had been applied wherever they were applicable.

I then analyzed the tags for the core discourses that emerge from and connect them, and the ways that individual Twitter users invoke these discourses. I focused on slurs used or discussed, discourses about age, time, and homophobia; the tweet author's overall stance towards *queer*; and usage of *queer* or other identity terms (*LGBT, homosexual, gay*, etc) in discursive context. I also looked at whether tweets contained a direct address to another Twitter user (using the @ sign.)

3.2 REPORTING DATA, MAINTAINING PRIVACY This study was designated "not human subjects research" by the Office for the Protection of Research Subjects at the University of Illinois. However, following reflection on the data collected and discussion with my dissertation director and committee[5], I had significant concerns about the privacy of tweet authors in my corpus. Of particular concern were the following factors:

[5] With thanks to Professors Krystal A. Smalls, Rakesh M. Bhatt, Michele Koven, and Anna Mendoza at the University of Illinois Urbana-Champaign.

1. the persistence and searchability of tweet text, even when the author's username and profile image are omitted.
2. the heightened privacy concerns of LGBTQ+ people.
3. the potentially controversial nature of tweets containing (reclaimed) slurs and metalinguistic critique

I settled on the following approach to reporting data, excerpted from my thesis:

> "Throughout this paper, I will discuss social media posts and broader social media discourses without reproducing the full text of any one post. The only exception is for truly public figures and public-facing organizations, such as the President of the United States, Elon Musk, J.K. Rowling, or the LGB Alliance. I have made this choice because even "anonymized" posts are easily searchable when the full text is available. Many researchers and institutional review boards take the stance that because these posts are made publicly, on public-facing accounts, they are not subject to consent practices or general human subjects data standards. However, I take the view that most social media users posting—i.e. speaking—online do not do so under the assumption that their speech will one day be gathered, decontextualized, analyzed, and disseminated through scholarly research, much in the same way that someone speaking to a friend in a busy town square would not make that assumption. What's more, social media speech is persistent and searchable in ways that fleeting public conversations of the past were not…As none of the people whose social media posts I have scraped and analyzed have gone through a traditional informed consent process, I have chosen to be as careful as possible to safeguard even their pseudonymous online identities. In this paper I…heavily redact individual posts, and discuss them in aggregate when dealing with wide-spread discourses. Any quotes from individual tweets will be subject to the "google" test, i.e. can searching the text on Google lead me back to the original tweet? Only text that is sufficiently shortened or redacted as not to be searchable will be included…When referring to the contents of specific tweets I use an identification number that specifies the list the tweet appears in, and then its individual designation within that list (i.e. 1_22 or 2_13)"

\

What this looks like, in practice, is a series of tables and charts aggregating word frequencies, sentiment scores, and discourse themes and frequencies across both keyword lists. For example, Figures 1 and 2 show the sentiment scores and discourse themes for List 1 (keyword: *queer*). Discourses emergent from the close analysis of tweets are also discussed in aggregate, with tweet designations in which those discourses are central cited alongside a paraphrase or description of the tweet.

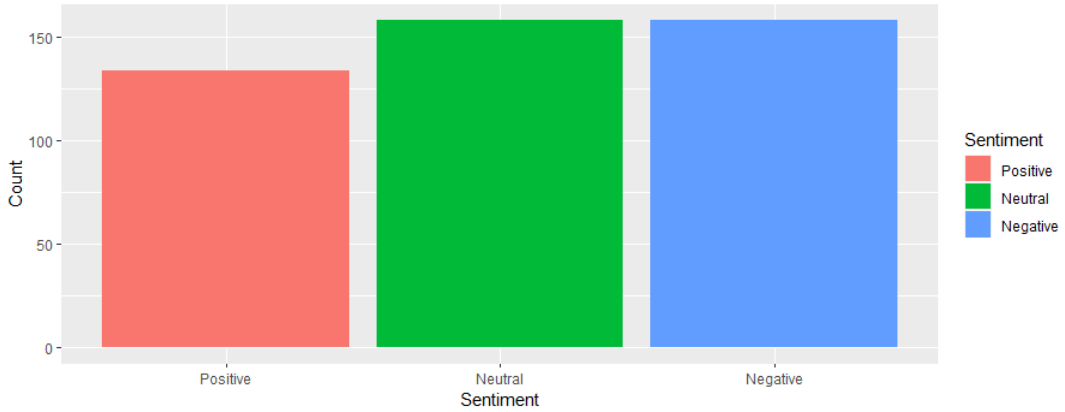

Figure 1. Sentiment scores in List 1

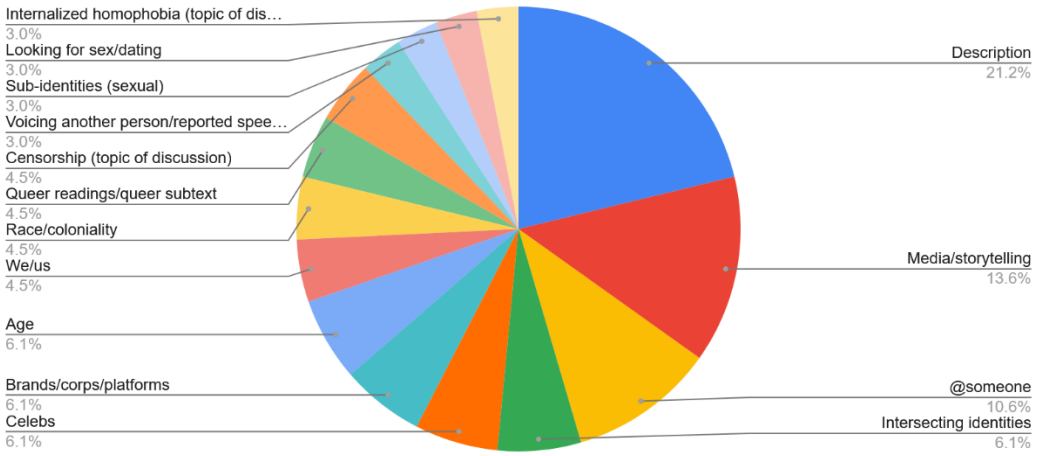

Figure 2. Discourse themes in List 1

3.3 REFLECTION This approach has several key strengths, as well as some methodological (and potentially ethical) weaknesses. In terms of strengths, this approach significantly decreases the searchability of tweets. All data is completely deidentified and the short snippets of tweets presented are not searchable on Google or the Twitter/X mobile application.

The lack of a standard consent process or any direct researcher involvement with subjects also allows for a much broader variety of viewpoints on the central topic –the reclamation of *queer* and its use as a community or individual identity label –than other approaches. Rüdiger and Dayter (2017) reflect extensively on their research on a publically accessible forum whose participants –pickup artists, or PUAs –is historically hostile to feminist or sociological critique. The researchers understood that forum members might refuse study participation or reject the lens through which their language was interpreted. Rather than a standard consent process, they chose to observe the forum without any kind of intervention, and "in lieu of fully informed consent, the researchers assigned pseudonyms over PUA forum members' chosen nicknames (resulting in double anonymization for users who chose an anonymous nickname in the first place), besides the usual anonymization measures such as removing references to names and geographical location."

While I do not agree with all of Rüdiger and Dayter's methodological choices –they, for example, publish very extensive and potentially searchable quotes from the forum –I do agree that a lack of direct researcher intervention or informed consent process makes it possible to study speech and subjects that might refuse participation based on the identity or positionality – woman, social scientist, queer-identified, queer theorist, etc –of the researcher. I also agree with them that, when writing up sociocultural research of any kind, transparency and reflexivity about a researcher's position and views are extremely important. As they note, "images of the investigator as a seeker of absolute truth, and of the subject as a reliable stimulus-response machine, sprout the myth of the unbiased researcher. It is, however, just that – a myth."

It is better, therefore, to be upfront about potential bias in both data collection and data presentation. I, for example, strongly disagree with both the anti-*queer* and homophobic viewpoints in my corpus, both from people who use *queer* as a homophobic insult, and from those who disparage "*the q-slur"* as a valid identity term. In framing the latter within discourses of homonormativity, transphobia, and LGB separatism –and by making my disagreement with all three ideologies clear –I am providing a necessarily biased analysis, informed by my own academic training within queer sociolinguistics and by my life experiences. I do not consider this a weakness in my analysis, but an inevitable result of doing social research as a social being.

That is not to say that this approach is infallible. Some pervasive issues remain. The most obvious is methodological; it is fundamentally difficult to present close discourse or text analysis without publishing the texts being analyzed. In weighing the importance of increased subject privacy against greater analytical transparency in the discourse analysis stage, I chose the former. This was made easier by the inclusion of supplementary aggregate[6] analysis, such as sentiment analysis, which I felt did not present as much of a privacy concern[7]. I was also able to compare the Twitter analysis to a previous study I had done on *queer*, which featured interviews with fully consented participants. Taken together, these different approaches to research reinforce each other, filling in some of the gaps left by the omission of full tweets.

Another approach is demonstrated by Ahmed et al (2018; see also Ahmed et al 2019). Their study looks at information sharing on Twitter during the 2009 and 2014 Swine Flu outbreaks, rather than controversial speech. While they did not obtain informed consent to collect their initial dataset of tweets –which is noted to contain approximately 214,784 items in the 2019 study –they chose to reach out via Twitter to obtain permission from users whose posts they wanted to quote directly. In order to make this process easier, they "argued that it may be necessary to accept a Tweet or saying "Yes" as evidence of consent, rather than a traditional informed consent process with "participant information sheets and consent forms." This approach may be easier when dealing with less sensitive or controversial topics.

Another issue I faced was the deletion of tweets that were already in my corpus. The tweets in my corpus were pulled from the Twitter API in summer 2023, at a time in which Twitter's new owner, Elon Musk, was instituting major changes to research, moderation, and hate speech policies on the platform (Barnes 2023; CCDH 2024; Edison Research 2024). With a

---

[6] Note that aggregation does not remove researcher bias. For a simple example: I chose *q-slur* as the keyword to compare to *queer* because it was the most common, searchable avoidance term *in my personal experience as a social media user.* The entire corpus was built using these two keywords. Another term might have yielded very different results, both in discourse analysis of individual tweets and in frequency and sentiment analysis of the entire corpus.

[7] Though there is further discussion to be had on whether even an aggregate approach is appropriate when subjects do not intend their social media data to be used in an academic study, or do not fully understand that it *can* be (see, for example, Williams et al 2018; Samuel & Buchanon 2020; Walsh 2023.)

measurable increase in hate speech and other controversies plaguing Twitter, many people – including or especially many LGBTQ+ people –left the site. Over the five months it took to analyze data and write up results, many of the tweets in my corpus were deleted from Twitter. In this case, Williams et al (2018), recommends that a researcher ought to "consider [deleted tweets] content removed from the public domain, and do not publish (unless they are from an organizational or public figure account)." Because of the level of anonymity in my study and the fact that my small corpus has not been and will not be made public (as well as the pressure to complete a dissertation) I chose to keep those tweets in the corpus and include them in my analysis.

4. Conclusion I reiterate here that this account is not meant as a template for further research. Rather, it is an example of the self-reflective approach to ethics that I believe will be crucial to navigating the shaky, every-changing grounds of social media as researchers. I offer it as a way of opening dialogue about how we approach research with language data online. As interactions with technology continue to be important (and ever-changing) parts of human social and linguistic life, and as online language data –individual language examples and "big data" alike – become easier for researchers to collect, analyze, and publish, we may need to shift from a one-size-fits-all approach to ethical research design and towards one that maintains a central commitment to human privacy, dignity, and safety above and beyond the safeguards mandated by review boards and platform terms of service.

To me that requires a "*reflective* and *dialogical*" approach (franzke et al. 2020), one in which the uniqueness of each research context and the specifics of platform and subject privacy needs are kept in mind through every step of study design. In this case, that means understanding that even anonymized tweets are easily searchable, and that such searchability, combined with the controversial nature of research on sexuality and slurs and the increased privacy needs of many LGBTQ+ people, means that I should not publish or present whole tweets in my research. It also means reflecting openly on how I came to this decision, and on the strengths and weaknesses of the resultant research.

I hope that this reflection, and the resources presented in this paper and below, will open a door for further conversation on our role as linguists in our increasingly datafied world. As Halford (2018) and Samuel and Buchanon (2020) point out, we are working in a space increasingly dominated by the computer sciences, and by interests –government, marketing, corporate –that are not always trained in or subjected to the same human subjects research ethics as linguists. It is paramount, then, that we openly and wholeheartedly step up to the task of ensuring the highest possible ethical standards in our work.

References

Ahmed, Wasim, Peter A. Bath, Laura Sbaffi & Gianluca Demartini. 2019. Novel insights into views towards H1N1 during the 2009 Pandemic: a thematic analysis of Twitter data. *Health Information & Libraries Journal* 36(1). 60–72. https://doi.org/10.1111/hir.12247.

Ahmed, Wasim, Peter A. Bath & Gianluca Demartini. 2017. Using Twitter as a Data Source: An Overview of Ethical, Legal, and Methodological Challenges. In *The Ethics of Online Research* (Arei), vol. 2, 79–107. Emerald Publishing Limited. https://doi.org/10.1108/S2398-601820180000002004.

Barnes, Jenae. Twitter Ends Its Free API: Here's Who Will Be Affected. *Forbes*. https://www.forbes.com/sites/jenaebarnes/2023/02/03/twitter-ends-its-free-api-heres-who-will-be-affected/. (14 November, 2023).

Center for Countering Digital Hate. New report: X Content Moderation Failure. *Center for Countering Digital Hate | CCDH*. https://counterhate.com/research/twitter-x-continues-to-host-posts-reported-for-extreme-hate-speech/. (25 April, 2024).

Davidson, Brittany I., Darja Wischerath, Daniel Racek, Douglas A. Parry, Emily Godwin, Joanne Hinds, Dirk van der Linden, Jonathan F. Roscoe, Laura Ayravainen & Alicia G. Cork. 2023. Platform-controlled social media APIs threaten open science. *Nature Human Behaviour*. Nature Publishing Group 7(12). 2054–2057. https://doi.org/10.1038/s41562-023-01750-2.

Edison Research. 2024. Twitter/X Usage Sees Sharp Decline. *Edison Research*. https://www.edisonresearch.com/twitter-x-usage-sees-sharp-decline/. (17 April, 2024).

Freelon, Deen. 2018. Computational Research in the Post-API Age. *Political Communication*. Routledge 35(4). 665–668. https://doi.org/10.1080/10584609.2018.1477506.

Halford, Susan. 2018. The Ethical Disruptions of Social Media Data: Tales from the Field. In Kandy Woodfield (ed.), *The Ethics of Online Research* (Advances in Research Ethics and Integrity), vol. First edition. Bingley, UK: Emerald Publishing Limited. https://proxy2.library.illinois.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1577428&site=ehost-live&scope=site.

Kelly-Holmes, Helen. 2024. Artificial intelligence and the future of our sociolinguistic work. *Journal of Sociolinguistics* 28(5). 3–10. https://doi.org/10.1111/josl.12678.

Marwick, Alice E. & danah boyd. 2011. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*. SAGE Publications 13(1). 114–133. https://doi.org/10.1177/1461444810365313.

Rampin, Rémi & Vicky Rampin. 2021. Taguette: open-source qualitative data analysis. *Journal of Open Source Software* 6(68). 3522. https://doi.org/10.21105/joss.03522.

Rüdiger, Sofia & Daria Dayter. 2017. The ethics of researching unlikeable subjects: Language in an online community. *Applied Linguistics Review*. De Gruyter Mouton 8(2–3). 251–269. https://doi.org/10.1515/applirev-2016-1038.

Samuel, Gabrielle & Elizabeth Buchanon. 2020. Guest Editorial: Ethical Issues in Social Media Research. https://doi.org/10.1177/1556264619901215.

Sellers, Chelsea, Gabrielle Samuel & Gemma Derrick. 2020. Reasoning "Uncharted Territory" : Notions of Expertise Within Ethics Review Panels Assessing Research Use of Social Media. *Journal of Empirical Research on Human Research Ethics: An International Journal* 15(1/2). 28–39.

Walsh, Melanie. "Chapter 18" in "Debates in the Digital Humanities 2023" | Debates in the Digital Humanities. https://dhdebates.gc.cuny.edu/read/debates-in-the-digital-humanities-2023/section/a57b98ab-0f10-45d0-b205-3e563aab7ea8. (3 March, 2025).

Williams, Matthew L., Pete Burnap, Luke Sloan, Curtis Jessop & Hayley Lepps. 2017. Users' Views of Ethics in Social Media Research: Informed Consent, Anonymity, and Harm. In *The Ethics of Online Research* (Arei), vol. 2, 27–52. Emerald Publishing Limited. https://doi.org/10.1108/S2398-601820180000002002.

.

Ethical Guidance Needed for Social Media Data in Research. *Medical Ethics Advisor*. Morrisville, North Carolina: Relias LLC 39(12). 1–16.