The Semantic Derogation of Female

Major Research Paper

Master of Arts in Linguistics and Applied Linguistics

Andrew Ferley

York University

Supervisor: Ruth King

Table of Contents

1. Introduction

As Schulz (1975) observed decades ago, terms of reference can reflect a range of underlying ideological assumptions. One of her examples of this phenomenon is the lexical choice between *freedom fighter* and *terrorist* (p. 64), with the former reflective of positive appraisals and the latter a far more negative one. Schulz goes on to discuss the phenomenon she labels "the semantic derogation of women" whereby once neutral terms of reference undergo pejoration; part of the discussion compares terms like *lady* with its male counterpart *gentleman*, with the former undergoing pejoration in many contexts but not the latter. I first became aware of this phenomenon sometime during 2020, with regard to terms of reference for women when I observed people replying to online posts which described women as *females* with entirely textless responses consisting solely of pictures of an alien race from the television series Star Trek: The Next Generation and Star Trek: Deep Space Nine. These aliens are caricatures of capitalism and misogyny, and they themselves refer to women as females. I interpreted these interactions to mean those who posted these pictures of the fictitious aliens were signalling shared gender ideology with the Star Trek aliens.

However, "terminally online" behaviour like this often does not necessarily translate into real-world discursive patterns. I next asked many of the women in my life their reactions to hearing women referred to as female(s). While there was some variation in response, quite a few reported that this was a red flag for them, and men who did this, in their experience, were either dangerous or toxic. I followed up on these observations with an informal sampling of friends and coworkers, and the results were suggestive of *female* as a lexical variant which had undergone semantic derogation. Later in 2021, I conducted a qualitative analysis of the online communications of a misogynistic online Pickup Artist community which utilized this lexical variant frequently. However, while some findings were suggestive of such an analysis, they were ultimately inconclusive as within that insular community, one manner of referring to women seemed as hostile as any other.

The present study continues this examination of what I have come to refer to as the *conspicuous female*. Specifically, I characterise it as a lexical variant for women which, when it occurs outside of clinical contexts, seems to carry ideological baggage. To this end, I approach the question from a different direction than my earlier project, which was small in scale and purely qualitative in analysis. First, I re-administer an earlier survey on attitudes towards this lexical variant with a larger sample and a wider age range than in my previous research. Secondly, I mine Twitter using keyword searches for a reasonably large corpus of tweets containing the targeted variant, *female*, and a variant which is less negatively charged, *woman*. The choice of this second, more innocuous variant, was decided based on results from the attitudinal survey. The Twitter data are analyzed using the tools of critical discourse analysis (van Dijk 2005), for content which indexes prejudicial ideologies as well as quantitative variationist methodology (Bayley 2019).

2. Literature Review

2.1 The Derogation of Words for Women

That word choice reflects underlying ideological assumptions has been long documented in the literature on language and gender. As Eckert & McConnell-Ginet (1995) note, words such as *woman* draw upon "…reifications that emerge from and constitute conventional maps of social reality," (p.2). Schulz' (1975) study had confirmed that terms of reference for women in English almost invariably acquire sexual connotations and/or index misogyny (p. 70). Earlier still, Lakoff

(1973) likewise observed that while terms of reference for men may take on a variety of associated meanings, terms of reference for women inevitably are part of a far narrower field, and almost always take on a negative sexual connotation (p. 61). For example, Schulz cites *dog* as a clear example of this phenomenon: when applied to men, it is usually at worst playful chiding; however, when applied to women it implies “a fat, slovenly woman” or a sex worker (p. 70).

A useful theoretical tool for conceptualizing this gendered duality in language is that of McConnell-Ginet’s (2008) idea of conceptual baggage (hereafter CB). Under this framework, utterances or individual words can come to carry additional semantic content which can trigger inferences which speakers may not intend (McConnell-Ginet 2008: 513). It can include “...what traditional lexicographers and others have called connotations, but also encyclopedic knowledge, stereotypes or prototypes, and background assumptions, as well as knowledge about social practices in the course of which the word gets used” (McConnell-Ginet 2008: 512). However, while the information content within the conceptual baggage of an utterance may be extremely important interactionally, it can also be quite difficult to precisely pin down. In her explanation of CB, McConnell-Ginet references Kitzinger’s (2005) study of phone calls to a doctor’s office, in particular the conversational turn, “My husband has a terrible headache and is running a high fever.” From this utterance, the speaker’s interlocutor makes a number of assumptions that logically follow from that statement, including that the patient is an adult male and married to the caller. However, the use of the word *husband*, the interlocutor may also suggest both that this person cohabits with the patient and is aware of many details of the patient’s medical history (McConnell-Ginet 2008: 513), which do not immediately follow as readily from the relationship implied by *husband*. This is an incisive demonstration of how CB is activated, “…with no overt awareness on the part of either the speaker or hearer…" Indeed, it can unintentionally carry semantic content of which the speaker is unaware. This content can be anything up to and including racial labelling or other ideologically troubling inferences which the listener may unpack, and which would shock the speaker to have them pointed out. Indeed, CB does not necessarily operate under shared agreement between the speaker and listener, and the CB of a word or utterance will "…not only be what the interpreter takes a speaker to have meant by the word but the overall impact on subsequent developments, including inferences other interpreters might draw" (McConnell-Ginet 2008: 515).

Lexical items such as *girl*, *woman*, and *female* all appear to be subject to semantic derogation. While grating to many, in some communities of practice *gal*, derived from *girl*, may be ideologically neutral and taken to be the female equivalent of *guy*. More broadly, *girl* often indexes childishness or immaturity (McConnell-Ginet 2003:93) in a woman and is often used as a particularly insulting way to address a man. Likewise, *woman* can be similarly derogative depending on the speaker and context. As McConnell-Ginet (1989) observes, "A man who means to insult me by saying "you think like a woman" can succeed. He succeeds not because I share his belief that women's thinking is somehow inferior but that I understand that he is likely to have such a belief…" (p. 45). Thus, while for some speakers, English *woman* may indeed be an ideologically neutral noun, it is as prone to undergo derogation as any other semantically related word. Interestingly, while some acquaintances have told me that they consider *female* as the most neutral means of referring to women, couched in science and divorced from issues of gender and derogation, this has not been universally the case. Schulz (1975) reported that, at the time of writing, *female* had already replaced *woman* in common parlance because *woman* had come to be seen as associated with sexual impropriety and taboo. However, by 1975, *female*, too, was declining in use in public discourse, having come to be regarded as "degrading and indelicate" and disparaging to women. Schulz further noted that by 1975 the Oxford English Dictionary referred to it as a synonym for *woman* which was "avoided by writers" (p. 71). This process is in line with Lakoff's earlier (1973) comment that words which refer to women inevitably acquire sexual connotation. This early literature leads me to investigate if the recycling of what is considered an appropriate term of reference for women is still ongoing.

2.2 Later Research on Derogation and Slurs

Burnett's (2020) article, *A persona-based semantics for slurs*, is concerned with the difference between slurs and terms of reference, which are otherwise perceived to be more "neutral." The comparison she utilizes is *dyke*, a slur, and *lesbian*, which is considered a "neutral sister"; however, this distinction is broadly applicable to the derogation of *female* compared to other terms of reference for women. Burnett's analysis fundamentally rejects the premise that either member of the pair is more politically neutral than the other; rather, both are seen as semantic constructs associated with different personae (Agha, 2003, Eckert 2008) or social types (Burnett 2020: 33). Burnett proposed that personae can be modelled as locations within an ideological space, with *dyke* locating the referent as being outside the mainstream ideologically and, in most circles, with *lesbian* locating the referent as being inside the mainstream of that ideological space (p. 33).

Burnett notes that stereotypes and beliefs (which constitute personae) attached to *dyke* and *lesbian* are highly contextual and insular to the communities of practice in which either is used (Burnett 2020: 41). For speakers which Burnett describes as possessing a traditional or "bigot" (Burnett 2020: 48) ideology, both *dyke* and *lesbian* could actually occupy the same semantic space and thus be synonymous. This observation provides insight into why my earlier study which targeted the online discourse of the Pickup Artist community showed little variation (Ferley, 2021). Within that particular CoP, the personae ascribed to *female*, *woman*, and likely any term of reference for women were fundamentally identical and shared the same negative semantic space.

Communication under Burnett's framework is a sort of signalling game, whereby the speaker attempts to communicate the location of a target idea within their ideological space to the listener, and the listener then attempts to use this information to locate the target within their own ideological space. This process is generally accurate in instances where the ideological distance between speaker and listener is narrow; however, as ideological distance increases, so too, does miscommunication (p.50). This is also found for those speakers who use *female* as a lexical variant, either without forethought or as a deliberate attempt to seem inoffensive. For such speakers, they may be unaware of the ideological distance they have from those who have reported that hearing women referred to as *females* as making them uncomfortable. This has consequences as those who hear *female* in an interlocutor's speech as a cautionary sign are more likely to interpret its usage not only as possibly derogatory (Burnett 2020: 54) but potentially as indexing other ideologies they find distasteful or dangerous.

Another analysis of slurs is presented in Beaton & Washington's (2015) article *Slurs and the indexical field: The pejoration and reclaiming of favelado 'slum-dweller'*, which utilizes Eckert's (2008) conception of the indexical field to understand how slurs function and may be reclaimed. For example, *favelado* is a term which refers to someone who lives in a *favela*, a slum or shanty town found in large Brazilian cities, and exists alongside *morador de favela* which literally means "inhabitant of the slum". While the former is used as a slur and is pejorative, the latter is not (Beaton & Washington 2015: 12).

Eckert (2008:463), drawing upon the work of Silverstein (2003), introduced the indexical field to third wave sociolinguistics using the example of /t/ release in English. She illustrates how a listener can assign social types, stances, and permanent qualities to a speaker on the basis of a single phonetic articulation. Beaton & Washington (2015) argue that lexical items can invite similar indexical connections. As they argue, lexical items "…create a link between the referent and the sociocultural implications built upon the nth order of the word…" (p. 14). While the n-th referential meaning of *favelado* (see above) is merely someone who lives in a slum, the authors suggest that the existence of multiple options with the same referential meaning but more neutral connotations, in this case *morador de favela*, allows for the pejoration of *favelado* in the same way as do other slurs (p.16).

These authors construct an indexical field for *favelado* similar to the one Eckert (2008) did for /t/ release. Included are permanent qualities such as "illiterate," "dishonest," and "street-smart" and social types such as "slut," "redneck," and "warrior," as well as fans and players of a soccer team from Rio de Janeiro. As these examples should demonstrate, the qualities which *favelado* indexes are not purely negative (p.16).

Beaton & Washington utilize the concept of indexical field to illustrate how *favelado* has been ameliorated. They discuss how, while most second-order meanings of *favelado* are negatively valenced due to social marginalization, there are ameliorating indexical meanings associated with the aforementioned soccer team. In fact, those fans who self-identify as *favelados* are not simply identifying as someone hailing from a *favela*, but instead explicitly identifying themselves with the indexical meanings of *favelado* which make it a slur. The authors describe this practice as such: "…They respond by aligning themselves with the insult. By making the connection between themselves, the players, and the slur, they foster an in-group identity alongside slum-dwellers." (p.18) Under this interpretation, the fan is reclaiming the slur and "…recycling it with positive valence for solidarity in common with lexical items used against marginalized groups." (p. 18). This argumentation may hold explanative power for the observation made by Schulz (1975) that *woman* has previously possessed negative connotations involving sexual impropriety but had at some point been reclaimed into "acceptable" discourse. It also provides an explanation for the attitudes of the women who I have encountered for whom *female* not only exists in their speech but is also deemed ideologically neutral. This may have occurred either as a result of a process of amelioration similar to the one described above, or the variant may have always existed within a positive indexical valence for them.

2.3 Computer Mediated Communication

Androutsopoulos's (2006) chapter serves as a valuable introduction to research on computer mediated communication research (hereafter CMC) and addresses some of the differences between CMC and traditional modes of communication which can provide challenges in research. He illustrates how early research in CMC privileged medium-specific features like emoticons or a reliance on acronyms which led to preconceptions of internet communications as being "…distinct, homogeneous, and indecipherable to outsiders…" This focus has led to, at least at the time of the chapter's publication, a lack of attention to the socially situated discourses in which the aforementioned medium-specific features, and perhaps the perpetuation of language myths about CMC, not dissimilar to how gender stereotypes have sometimes been unintentionally replicated in language and gender research (Androutsopoulos: 2006: 420). Thus, he places significant value in then-ongoing research which sought to demythologize CMC and replace what were then typical searches for "…typical features of 'netspeak'…", such as the aforementioned smiley faces and externally inscrutable acronyms, with analyses that holistically examine the full contextual parameters which are evoked in various types of CMC as rich interactional genres.

Androutsopoulos' discussion of such analyses includes an examination of how certain spaces on the internet such as forum boards or chat servers may be viewed as communities. He notes that for many genres of internet communication, being considered a community is of marginal importance. Indeed, CMC often lacks the stable group membership, long-term commitment, and social accountability for any given internet space to qualify as a community sociologically (Androutsopoulos: 2006: 422). As a corollary of this, sociolinguists who wish to engage with CMC data must continually engage with the issues springing from a fundamental aspect of CMC, namely its anonymity and the ephemeral nature of individual internet authors participating in a given interaction for any length of time. Because of this, quantitative variationist sociolinguistics was, at least at the time of Androutsopoulos' (2006) publication, quite rare due to the unreliability of demographic speaker data privileged by sociolinguists (p. 424-425). Indeed, researchers even today must generally rely on self-description by their subjects. However, Androutsopoulos does provide a valuable insight in that in some instances, age and gender can be inferred directly from the source of the data. He suggests that those online communities which make explicit their social profile, such as teen chat rooms, or forums related to geographic area or ethnicity, provide self-identification as a group. Thus, in the absence of explicit demographic information tied to each individual speaker's token, such community self-identifications can provide a useful contrastive base for comparison (p.425).

Herring (2000), observes in her article, *Gender differences in CMC: Findings and implications*, that the generalization regarding anonymity in CMC is least somewhat exaggerated. She claims that achieving true anonymity is actually somewhat difficult since media platforms require names and email addresses attached to online identities. Further, she claims that bypassing these requirements requires specialized software and software skills that are outside the purview of most users (p.2). Admittedly, these specifics are largely out of date; however, the core of Herring's argument is still holds. While the intervening two decades since Herring's (2000) article have seen the ability to generate false credentials including disposable email addresses, become quite accessible, and there is allowance for anonymous participation on social media platforms, Herring's (2000) central argument that personal information can still be discerned remains valid. Indeed, even to her point, while accessing the internet through a VPN and utilizing social media accounts linked to temporary email addresses is fairly simple, there is, to my knowledge, no evidence that these practices are employed by any significant portion of internet users. Thus, Herring's (2000) claim remains useful; demographic information is accessible, even if it is not immediately apparent. However, while the ease of anonymity should encourage some skepticism on the part of a researcher, the additional pragmatic barrier that identity spoofing involves means that, on a large enough scale, a researcher can rely on self-reporting for variables such as gender and country of origin from their subjects.

Herring (2000) also provides some guidelines for at least provisionally assigning gender to Twitter speakers which may be otherwise obfuscated. She notes that stereotypical patterns of gendered speech differences are often found in CMC, including tendencies for men to assert opinions as fact, post longer messages, use profanity, or manifest an adversarial orientation towards their interlocutors. She argues that women, in contrast, apologise more frequently, hedge statements, and express support among others (p.3). Herring (2001) expands on these ideas in the chapter *Gender and power in online communication*, in which she claims that men and women display radically different styles in online postings, with men making use of sarcasm and insults (p.8), and women making greater use of appreciation, support and qualified assertion (p.9). While such stereotypes are not useful (or confirmed, see Cameron 2007) in language and gender research on a large scale, for individual interesting tokens within a corpus, such metrics are potentially useful for tentatively assigning gender when it is not readily apparent from a user's name or pronouns on their homepage.

In fact, for a linguist with a background in language and gender studies, Herring's generalizations seem dangerously close to reproducing reified gender differences with an academic veneer. She does, however, note that gender differences in CMC do disfavour women, and thus such patterns may simply be a reproduction of the "real world" gender status quo (p.19). Herring suggests that in mixed sex online discussions such as Twitter, women tend to post less when their messages receive little or negative reaction, whereas men are more likely to persist as well (Coates 2015). This discrepancy has the potential to cause the online communication contributions of men to be at least over-represented in any given corpus (Herring 2001:9). Herring notes that some evidence suggests that active moderation can correct for this variation in participation (p.10). Twitter possesses a fairly negative reputation on moderating against harassment and other kinds of abuse, and this leaves a very real concern that the online speech of male users may be overrepresented. Since the present study deals in part with Twitter discourse, this point must be kept in mind.

In her article *Computer-mediated discourse analysis: an approach to researching online behavior*, Herring (2004) provides valuable insight into a number of issues involved in designing research in CMC, of which the most relevant involve different techniques of data selection. She presents a number of sampling techniques which go beyond analysing whatever a researcher happens upon by chance. Herring suggests that temporal sampling provides the richest context for analysis, capturing "…coherent threads [and] thereby incorporating the advantages of thematic sampling as well" (p.11). I integrate this sampling technique into the present study with all tweets harvested between the dates of June 2nd, 2022 and June 4th 2022. Due to the nature of hashtags and the discourse utilizing femme-indexing language which tended to occur on those dates, this has the added benefit of integrating thematic sampling techniques into this study. Taken together, these two techniques provide a high level of coherence in the data gathered, though possibly with a remote risk of creating a sense of false positive correlations of red flag ideologies with either or both variants if the discourse sampled is unusually toxic at the moment in time for gender and sexuality discourses on Twitter.

D'Arcy & Young's (2012) book chapter *Ethics and social media: Implications for sociolinguistics in the networked public* raises important ethical considerations with reference to online research. While the authors' analysis is limited to Facebook and other social media platforms which follow similar network configurations (p.536), they propose a number of ethical CMC research quandaries which deserve consideration. Among these, the most relevant to this study are expectations of privacy with regard to nominally public-facing data. The authors discuss the conflation of a researcher's ability to access data online with that data being public facing. D'Arcy & Young reference a study performed at Harvard entitled *Tastes, ties, and time: A new social network dataset using facebook.com* (Lewis et al., 2008) which they go as far as to label "infamous". The researchers involved in *Tastes, ties, and time* incorrectly assumed that because they could access Facebook profiles and posts that such posts and the associated personal information were entirely public facing. However, as the researchers eventually discovered, their Facebook accounts shared network ties with the authors of some of the posts they used as data, and thus these data were in fact not public facing. This meant that the extraction and utilization of their social media data was done without the authors' consent and outside of their expectations of privacy (D'Arcy & Young, 2012: 536). It must be recognised that privacy settings are idiosyncratic to each online platform and are occasionally obfuscated by platform creators. In this regard, Facebook is an excellent example as its privacy settings are notoriously obtuse, hidden behind multiple menus depending on the computer platform being used to access Facebook. Thus, depending on the social network, a subject may have shared or posted information under the misapprehension that the activity was private when it in fact was not.

Finally, D'Arcy & Young (2012) invoke Bell's (1984) Audience Design theory in their conceptualization of the roles of participants in a Facebook exchange, with both specific addressees and what the authors describes as the "invisible audience". Within CMC research, the invisible audience includes friends of friends, networks and those able to read posts set to be read by everyone. To these participants, they apply Bell's roles of "…auditors, overhearers, and eavesdroppers," respectively. D'Arcy & Young advise the researcher to not undertake the role of an eavesdropper in research on Facebook. Due to the nature of Facebook's network of friends and private groups, there is an expectation of privacy, and a researcher acting as an overhearer is acting no differently than Lewis et al. (2008), doing the equivalent of "…surreptitiously recording conversations in non-virtual spaces" (D'Arcy & Young 2012: 537). Due to the degree of nuance involved in Facebook's privacy and sharing settings, the authors devote a great deal of the discussion to articulating guidelines for maintaining ethical integrity in CMC research on Facebook, including but not limited to the creations of Facebook groups which are explicitly tied to each specific instance of research (p. 538), appropriate methods involved in soliciting and recruiting participants into those groups (p.539-540), and even advice on severing all social media ties to former participants after a study's conclusion (p.541-542). Twitter, however, which is used in my own research, is fundamentally different in structure from Facebook. Where Facebook has labyrinthine privacy and sharing settings, on Twitter all posts are inherently publicly facing. Additionally, the only networking option is a non-obligatorily reciprocal "following" relationship. Thus, the researcher on Twitter should not be constrained from taking on the role of the overhearer as that is the standard audience role (Bell 1984) for any participant on Twitter. Additionally, while content on Facebook is at least nominally obligatorily linked to a person's "real world" name and identity, Twitter does not require participation under this identity. Tweets on Twitter can readily be published under user-selected pseudonyms and the public display of demographic or locational data to one's Twitter profile requires intentional effort. Indeed, we might well regard the study of Twitter data as being more akin to the rapid and anonymous survey methodology employed in Labov's (1966) landmark department store study than to Facebook data analysis.

3. Methodology

3.1 The Attitudinal Survey

Developed in 1932 by Rensis Likert, Likert scales are a psychometric response scale often used in questionnaires which aim to measure participants' preferences or degree of agreement with statements (Bertram 2007: 3) The most commonly encountered Likert Scale is a five-point agreement scale which ranges from (1) "Strongly Agree" to (5) "Strongly Disagree". It has been commonly used in both psycholinguistic and sociolinguistic research on language attitudes, beginning with pioneering research by Lambert et al. (1960). Likert Scales are also useful for measuring the frequency of a habit, the perceived importance of a topic, how similar items are, or how likely an occurrence of the studied variable is (Siegle 2015).

Likert Scales are useful in the context of this study as they are readily applicable to speakers' pragmatic knowledge as well as their affective dispositions (Nemoto & Beglar 2014: 2). However, as a tool for gathering data they do have some methodological problems that the researcher must take into consideration. First, odd-numbered Likert Scales that allow for neutral answers (e.g., 3 on a 5-point scale) risk participants' "fence sitting" in that respondents may shy away from taking a stance one way or another on an experimental stimulus (Brown 2000:28, Bertram 2007: 7). This fence sitting can be partially remedied through creating even-numbered Likert Scales which remove a neutral option, and thus forcing respondents to take a side. However, this can result in compromising the integrity of the data as respondents without a strong opinion on certain stimuli are forced to report an opinion they simply do not have (Brown 2000:28). Additionally, Likert Scales also face problems arising from a social desirability bias whereby respondents desire to be perceived more favourably or liked by the researcher (Bertram 2007: 7); however, this problem is somewhat mitigated by anonymizing questionnaires. There are also issues with stimulus construction, as improperly created stimulus may not measure what the researcher aims to measure. As an example, "double-barreled" questions which include conjunctions such as *and*, *or*, and *but* should be avoided as they are confusing. As Nemoto & Beglar (2014) note, some respondents, when given a statement like, "I can understand written and spoken academic texts", may incorrectly latch onto either "written" or "spoken" and answer one or the other, tainting research data. (p. 3) Nemoto & Beglar recommend that, if at all possible, constructions like these should be separated into two distinct stimuli.

With regard to the analysis of Likert Scale data, it is generally accepted that the median and mode responses are the most illuminating metrics to extract from coded data. However, there exists some controversy on whether the mean average of ordinal data has any relevance. While some suggest that the mean average of your data must never be considered (Allen & Seaman 2007, Bertram 2007) since, for example, the mean of “agree” and “strongly disagree” is fundamentally meaningless, others (e.g., Sauro 2016) argue that once these values have been coded numerically as ordinal data, insights can sometimes be obtained in this manner, depending on the research context. Statistical analysis may then be performed upon Likert Scale data which may include simple significance or multivariate regression analysis; however, this should be considered provisional at this time, as the ultimate method of analysis will depend on the nature of the data gathered.

3.2 The Twitter Data

While mining internet data can provide enormous corpora of data and thus be an invaluable resource for researchers, there are certain problems which must be kept in mind in the design and execution of any research which relies on such data. The following three studies work specifically with Twitter data and together shine a light on the major challenges this resource presents. Ilbury’s (2020) article *Sassy queens: stylistic orthographic variation in twitter and the enregisterment of AAVE* examines the usage of variant spellings commonly linked to AAVE to enact the persona of the *sassy* queen by queer Twitter users. The author argues that queer users rely on the conceptualization of Black women as “fierce” and employ variant orthographic spellings which enregister Black identity in the production of *sassy queen* personae (Agha 2003, Podesva 2007). Ilbury gathered data from a corpus of 15,804 tweets collected over a year from the timelines of ten gay British men. The author notes that duplication, retweets, and advertisements had to be culled to get down to this number. Additionally, while all subjects were prolific Twitter users, the author notes that individual contributions had to be capped for some users at 3000 tweets (p. 250). This corpus of data was then manually inspected for instances of orthographic variation, with non-standard spellings individually extracted and examined for the presence of AAVE features. From this corpus, the author was able to extract just 307 tokens, which were present in only 1.9% of the Tweets extracted (p.251). Despite this, the author contends that these misspellings represent an intentional effort to invoke the indexical meaning of Black women as “sassy” and “fierce” in the users practicing of their sassy queen personae (p. 260), an interpretation that quantitative researchers would probably not find compelling, given the relatively small amount of data that proved usable.

A second Twitter-based study is Memon et al.’s (2020) article *Characterizing sociolinguistic variation in the competing vaccination communities*, which compares pro-vaccination and anti-vaccination communities on Twitter. The authors utilized Twitter researcher privileges (p.2) to extract 588,110 tweets by 6262 Twitter users and performed both linguistic analysis and social network analysis to characterise each group of users. Linguistic analysis entailed examining tweets based upon three variables the authors suggest identify a user as either pro or anti-vaccination: usage of intensifiers, pronominals, and expressions of uncertainty. The authors claim that use of intensifiers correlates with pro-vaccination sentiments as such usage is associated with convincing others of the efficacy of vaccines. High levels of pronominal usage is associated with narratives, which may be produced due to the anecdotal nature of most anti-vaccination rhetoric. Finally, the authors suggest that use of words associated with uncertainty is negatively correlated with what they describe as echo-chamberness (p.5) and use this variable as a means of comparison between the two groups.

The authors found that their initial hypothesis that pro-vaccination users would utilize more intensifiers was entirely incorrect as self-identified anti-vaxxers actually used greater numbers of intensifiers; however, their hypothesis about pronominal usage held. Additionally, these authors found that, contrary to their expectations, Twitter users in anti-vaccination communities which were found to have greater degrees of echo-chamber-ness actually utilized more expressions of uncertainty (Memon et al., 2020: 8). The authors did notice that their usage of hashtags in data collection represented a limitation in their work as usage of hashtags was unequal between the two communities. The authors note that this likely affected the integrity of their data, but they do not state in what manner. (Memon et al., 2020: 10).

Along with the difficulties observed in Ilbury (2020) and Memon et al. (2020), namely the necessity to filter out irrelevant Twitter data which drastically lowered their amount of usable data as well as faulty pre-experimental hypotheses on the expected habits of Twitter users, Simaki et al's (2016) article *Age identification of twitter users: classification methods and sociolinguistic analysis* is also an important demonstration of one of the fundamental difficulties in working with Twitter data, namely extracting any demographic information that can be indexed with a particular Twitter token. As the title suggests, Simaki et al's article deals solely with attempting to discern the age of Twitter users. As Twitter users are not obligated to tweet under their real names, it naturally should follow that their ages are similarly obscured behind the anonymity of the internet. The authors attempt to get around this limitation through the application of machine learning algorithms (p.7). The algorithms applied to examine tweets measured a wide range of variables, including the number of characters per tweet, minimum word length, number of emoticons, number of sentences, and number of capitalised letters, as well as literally dozens of other metrics. With these variables in mind, the best set of algorithms, named Random Forest, managed to achieve an accuracy rate of 61% for independently verified ages for Twitter users (p.8). The authors appear enthusiastic about the efficacy of this program but the results underscore how unreliable this resource is for making demographic comparisons.

While my own research does not employ machine learning or other complex programming tools, these three papers illustrate many of the difficulties which must be taken into consideration when utilizing Twitter data mining in sociolinguistic research: specifically, large corpora of data have been shown to be difficult to analyze meaningfully, experimental bias can lead to constructing a corpus that doesn't show what you intended to study, and gathering even basic demographic information can be extremely challenging.

These studies do provide valuable insight on how to avoid common pitfalls in analyzing corpuses of online social media data. First, Ilbury (2020) cast too narrow a net in their selection of Twitter data since they limited their analysis to the Tweets of under a dozen users. While the corpus of Tweets they amassed was impressively large, the number of tokens which corresponded to the sociolinguistic practice they were studying, namely whether gay men employ language and spellings which index "fierce" personae, was frankly unimpressive. Given my personal knowledge of queer social media practices, I am inclined to believe that their hypothesis was not incorrect, but rather that their methodology was malformed. Casting a wider net of users and capping individual user token contributions even lower than the authors had done might have resulted in a better data set. Secondly, Simaki et al. illustrate the difficulty in ascertaining certain demographic information for online data. Fundamentally, there may simply be no feasible way of resolving this issue and researchers who deal with this sort of social media data may have to limit themselves to users who have previously volunteered information regarding geographic location or gender. Finally, Memon et al. (2020) did not encounter problems due to the nature of the online data itself. Rather, the difficulties they encountered stemmed fundamentally from problems with hypothesis formation during the planning phases of the study. Ultimately, researchers working in this still relatively new field of study need to be mindful that methodologies which were applicable to in person or traditional telecommunications may not be directly applicable to research on CMC. A successful CMC researcher will, first and foremost, need to be mindful of the evolving nature of best practices for research.

3.2.1 Exclusions

Twitter data extraction required a number of exclusions. First, as discussed above, retweets are pragmatically complicated and thus were excluded. A conservative estimate placed retweets at over eighty percent of all tweets gathered by the Fireant software program. Duplicated tweets, either as a result of posting errors or users copy/pasting other tweets were also excluded. Secondly, this study focuses on terms of reference used as nouns and thus instances in which they were used adjectivally were excluded. Thirdly, tweets which included non-English text were excluded as code-switching and the possibility of other-language influence were beyond the scope of this work. Finally, a significant number of harvested Tweets turned out to be advertisements which represented a fundamentally different kind of discourse and were thus not included in the analysis.

Initially, it had been planned to exclude tweets which also used the masculine-indexing terms in tandem with the targeted lexical items, specifically *men* and *women*, and *males* and *females*. While it was workable to exclude tweets which included *men*, the same was not the case for *male*. Twitter users who utilized *female* almost invariably also used *male*, possibly due to the semantic content of some tweets was Trans-exclusionary and constituted biologically essentialist discourse which was extremely common. Ultimately, I decided that these pairs would be allowed, as removing them would colour the data to show usage of *women* to be massively overly represented with red flag ideologies relative to *females*. Still, in a similar vein, the exceedingly rare tweets which used both *women* and *females* interchangeably were excluded.

3.3. The Present Study

This study draws upon data from two sources to investigate present-day derogation of *female* as a term of reference, both in terms of perception and actual usage. To this end, both an attitudinal questionnaire (see attached) and online data mining of Twitter discourse are employed. While prior informal conversations had suggested to me that there is an existing negative perception of *female/females* as a term of reference for women, this association remained to be verified for a wider range of present-day speakers through the usage of the attitudinal survey. Secondly, a total of 645 Twitter tokens of *female/females* are analyzed qualitatively for occurrences of misogyny, transphobia, or other potential "red flag" ideologies which previous respondents have informally indexed to hearing *female*. An example of a tweet containing what I describe as red flag ideology is, "Cause unlike these other females on here you have a since [sic] of humor Ik [I know] it made you laugh right?" (Twitter User @00unces). In this tweet, the author is expressing the misogynistic belief that women are not funny, which of course falls under this study's definition of red flag ideologies. Twitter mining on this scale demonstrated that only gender was a viable independent variable in the analysis of usage of these lexical variants. Tweets which featured easily accessible and reliable information on speaker ethnicity or country of origin are present in the corpus; however, they occur in such relatively small numbers that constructing a viable corpus solely with tokens which also feature these independent variables was beyond the time constraints for this study. Finally, 548 tokens for *women*, determined from the results of the attitudinal survey as the control variant have also been mined from Twitter. This second variant is the most perceptually neutral, based upon the results of the attitudinal survey; specifically, it was the variant for which the mode average response is "3", as well as which also had a median result closest to "3." The use of the two variants was compared for overall incidence of red flag ideologies to determine if *female* is markedly more negative than the control variant utilizing logistic regression analysis.

Drawing on my previous experience designing and administering sociolinguistic questionnaires, I created the *Language and Gender Survey* using Google Forms, which is a free online application and part of the larger *Google Workspace*. Google Forms allows a user to quickly create and disseminate surveys, as well as provide an easy means of exporting resulting data to Excel or other spreadsheet programs. Once in Excel, both Likert Scale attitudinal ratings and descriptive terms collected can be easily graphed and tabulated. *The Language and Gender Survey* studies perceptions of terms of reference in both subject and object positions by providing participants with two template sentences and replacing blanks with targeted stimuli. These template sentences are as follows, with the stimuli being inserted into the blank spaces,

1) “I recognize that ____.”

2) Those ____ live around here.”

The stimuli are, in order: *man, boy, male, woman, girl, female, lady*. Respondents were asked to assess each sentence provided on a five-point Likert Scale based on the reaction they would have to hearing each gendered noun in a non-clinical setting. Here, a 1 or a 2 indicates a negative reaction and a 4 or a 5 a positive one with 3 indicating a neutral or lack of reaction. Additionally, after each sentence participants were provided a text field in which to list any adjectives or qualities (‘terms’ is the prompt used to aid non-linguists) they associate with usage of each gendered noun. It is interesting to see how much commonality exists among respondents in adjective choice used to describe their impressions, and to see if there is suggestive patterning in the qualities and stances participants associate with any of the terms of reference. In addition, indexical fields similar to those created by Beaton & Washington (2015) are constructed.

Both template sentences were designed to as far as possible avoid implicit bias as well as avoiding ascribing qualities to either gender in the question composition. The first sentence template creates singular examples of gendered nouns, while the second creates plural. There was initially some concern that the plural may engender implicit negativity in some respondents through stereotyping where the singular template may not. However, that did not appear to be a factor, as the results discussed in Chapter 4.1 show. As I will demonstrate, respondents did not appear to respond to singular or plural stimuli differently. While additional stimuli would have been desirable, an important consideration in creating online surveys is length, and at fourteen stimulus questions, associated comment boxes, and five demographic questions (English native proficiency, country of birth, current age, generation, and gender) the survey is already at the upper limits for a general online audience. The omission of ethnicity is an unfortunate oversight which is discussed in Chapter 6.1. The initial target was 300 completed questionnaires. Based on previous experience, this is a realistic number to gather relying on known social network brokers and social media. The "refer a friend" method is employed whereby participants are asked to pass along the survey to family members and/or friends. Additionally, those with social media presences are encouraged to post the survey for their followers in order to cast the widest net possible. Social media websites such as Facebook.com and Reddit.com were also utilized as a means to spread the survey, utilizing such online communities as *r/samplesize*, or *Survey Exchange* and *r/takemysurvey* with the former being a Reddit community wherein a survey link is posted for those who enjoy taking surveys, and the latter two being communities of those seeking data will complete one another's surveys in a *quid pro quo* manner. Finally, as mentioned above, this survey also informs the choice of which variant serves as an experimental control against which *females* will be compared. From the possible selections of *women*, *girls*, and *ladies*, the variant with the median and mode Likert Scale responses closest to “3”, or neutral, was selected, and that variant was *women*.

The second half of this study draws on Twitter data to analyse usage of *female* and the expression of ideologies that it co occurs with. Twitter is a microblogging social media platform wherein users communicate through *tweets*, which at present is two hundred and eighty words maximum. Users can follow one another, which allows for the tweets of followed users to display in the follower’s feed on their Twitter homepage, though this follow relationship is not necessarily reciprocal (Gruzd et al. 2011). Finally, Twitter users can respond to other tweets or *retweet*, which essentially reposts and recreates the previous tweet while giving attribution to the original author. This can often signal agreement and, given the comparatively low effort it represents, a significant portion of mined twitter data. However, given that the act of retweeting is contextually varied (people can often retweet things they find repugnant to raise awareness), retweets are not studied in this work.

Tokens are selected through Fireant (Anthony & Hardaker 2022), software which, once attached to a Twitter account which possesses researcher privileges, allows a user to gather up to one hundred tweets at a time using searches for targeted words and phrases. This process can be repeated eighteen times in a single minute, and after an additional minute to refresh, this may be done repeatedly. The searches target tweets which contain either *females* or *women*. The target is 500 each for both sets.

Once the data are mined and organized into Excel spreadsheets, the tweets are qualitatively analysed for indications of ideologies which include those that index what have been described as red flag ideologies, which include but are not limited to: misogyny or transphobia as well as other negatively associated ideologies which may occur but are not currently predicted. Each tweet is assigned a binary variable indicating whether or not it contains red flag ideology. The number of red flag tags for both *female* and the control variant is then analysed statistically to determine if usage of *female* is significantly more correlated with these red flagged ideologies than the variant the questionnaire respondents indicated as more neutral.

4.0 Results

4.1 The Language Attitude Survey: Consultant Ratings

The Language and Gender Survey received 407 completed questionnaires, of which 325 involved native speakers of English. Sample questions are provided in Section 3.3 and the entire survey document is found in the appendix. Respondents were predominantly Millennials and members of Gen Z, with older generations making up just over 10% of the total. While this overrepresentation of younger speakers limits the generalization of its findings, it is likely still reasonable for comparison with the Twitter data. It is additionally not surprising, given that the survey was predominantly disseminated on various social media. A majority of the Baby Boomer responses required the survey to be printed out and returned via hard copy, so few Baby Boomers had the ability or opportunity to circulate the survey within their peer groups. Women are slightly overrepresented, with men making up just under 40% of the corpus. Finally, non-binary and trans individuals appear to be overrepresented, comprising 11.8% of the corpus.

Likert-sclae responses overwhelmingly confirmed the initial hypothesis that usage of *female* and *females* was negatively perceived among most populations, with 63.1% of all respondents providing a “Very Negative” (“1”) response to the stimulus “I recognise those females." Adding those who responded with a "2" brings the negative response rate to 83.7%. Conversely, only 15 participants responded favourably to this stimulus. Results for the singular variant stimulus, "That female lives around here," is virtually identical.

As mentioned in the Chapter 2, one of the two primary goals of *The Language and Gender Survey* was to determine which variant was perceived as the most neutral, and thus the most suitable to serve as a point of comparison with *female* or *females* in the Twitter data analysis. This turned out to be *woman* and *women* with the singular and plural forms receiving neutral ratings at 53.8% and 54.1% respectively. Therefore the plural forms *females* and *women* were selected for comparison within the Twitter data analysis. The analysis of Twitter data was limited to the plural forms as the inclusion of singular forms introduced additional contextual complications outside the scope of this study.

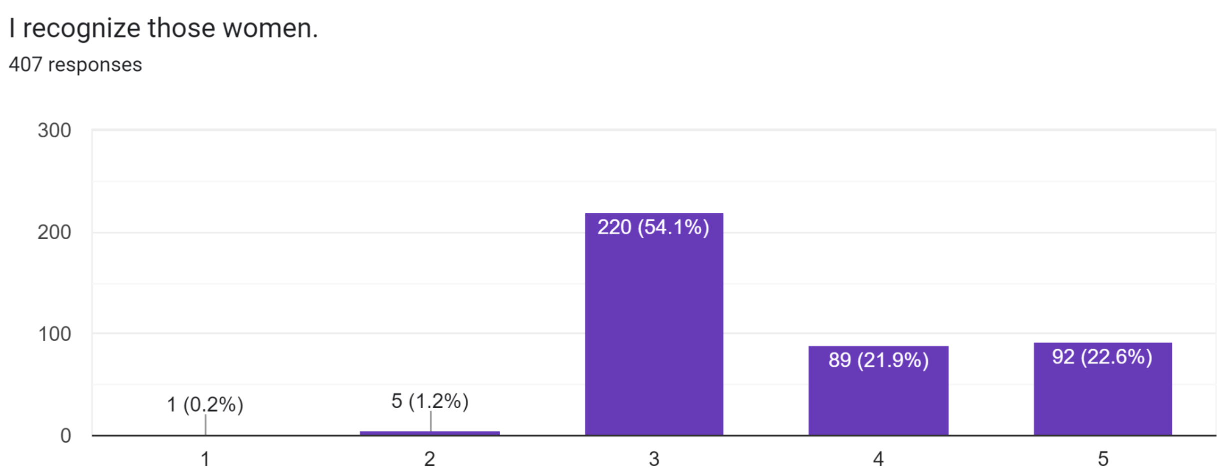

Figure 1: Likert scale results for "I recognize those women."

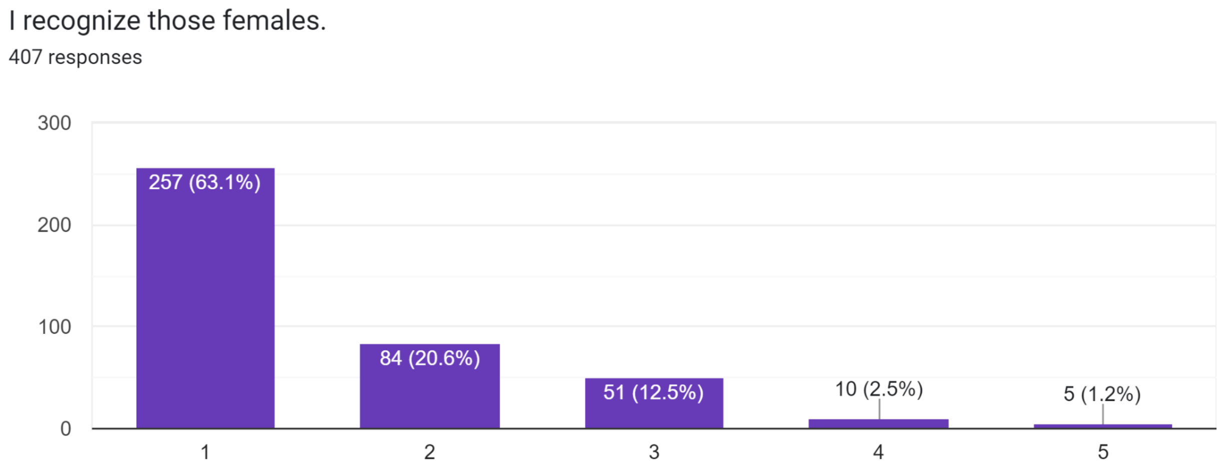

Figure 2: Likert scale results for "I recognize those females."

These results show neutral attitudes to *women* and highly negative attitudes towards usage of *females*.

4.2 The Language Attitude Survey: Descriptors

While the Likert-scale questions on *The Language and Gender Survey* required participants to provide responses to every scale, the corresponding fields which prompted participants to provide up to three descriptors were intentionally left optional (see Appendix). This inclusion of descriptors varied considerably by term of reference. For the 325 native-English speaking participants, 181 (56%) provided at least one descriptive term for either *female* or *females*. In comparison, that number drops to 125 (39%) for those who provided any descriptor for *woman* or *women*. Additionally, among those who provided descriptors,the *female* variants received an average of 3.4 descriptors per respondent out of the possible 6, while *women* variants received 2.9. This is in line with the scalar data, that the *women* variants are perceived neutrally and thus appear less likely to garner specific reference to identities or qualities, whereas the *female* variants evoke stronger responses among respondents which are more likely to be tied to particular behaviours or persons. The individual descriptors and their rates of occurence further illustrate this point. Consider Tables 1 and 2 below, which display the most commonly repeated descriptors provided by respondents, namely those descriptors which occurred 5 or more times. These tables pool the responses to both the singular and plural stimuli. Note that fewer than 10 of 325 respondents varied in the positivity or negativity of their descriptors between the singular and plural forms.

Word	**Number of Occurrences**
Normal	58
Neutral	44
Respectful	23
Polite	13
Average	11
Casual	9
Formal	9
Mature	8
Older	8
Adult	7
Nice	5
Straightforward	5
Young	5

Table 1: Descriptors provided for *women* and *woman*.

Word	**Number of Occurrences**
Sexist	45
Weird	42
Incel	40
Misogynist	38
Creepy	27
Awkward	22
Odd	19
Man	18
Clinical	15
Dangerous	15
Rude	14
Dehumanizing	13

Disrespectful	13
Gross	13
Ignorant	13
Strange	13
Police	9
Scientist	9
Non-Native Speaker	8
Alien	7
Ferengi	6
Robotic	6
Insulting	5
Older	5
Predatory	5
Unnatural	5

Table 2: Descriptors provided for *females* and *female*.

Perhaps unsurprisingly, a strong plurality of the descriptive items provided for *woman* and *women* consist of "normal", "neutral", and "average". There are also a fair number of stances which do not generally index specific gender identities such as "respectful", "polite", and "straightforward". Additionally, there are conflicting descriptions surrounding age and style such as "older" and "young" as well as "casual and "formal". This in inaddition to the approximately 24 descriptors which have only a single or small number of occurrences which involve an extremely wide range of indexical values. Overally, these results serve to illustrate that *woman* and *women* are considered neutral by most respondents..

In contrast, *female* and *females* are strongly associated with negative stances and characteristics, as well as very specific identities which are also often negatively perceived. With 45 and 38 occurrences respectively, "sexist" and "misogynist" represent a substantial proportion of the provided descriptors and clearly illustrate that usage of this lexical variant is, in the minds of a large number of respondents, associated with misogyny. Similarly, other negative descriptors include "odd", "awkward", "creepy", "dehumanizing", and, rather starkly, "dangerous". In addition, unlike the case of *woman* and *women*, usage of *female* and *females* is associated with very specific identities. These range from neutral to highly negative. For this variant, the descriptor with the third highest rate of incidence was “incel” (or “involuntary celebate”). It should be noted here that the primary method of data collection for this study was through online dissemination of the survey, and thus the awareness of this online identity and its association with usage of *female* is probably heightened. Law enforcement and science-based identities such as “police”, “cop”, “scientist”, and “biologist” also occur, as shown in Table 2. The word clouds presented in Figures 3 and 4 group both of these pairs into *police* and *scientist* respectively. *Police* co-occurred with negative Likert ratings while *scientist* was more neutrally perceived, though not universally so.

In addition, the clouds presented in Figures 3 and 4 demonstrate the indexical values of *women* and *females* are different. “Normal” and “neutral” dominate the cloud for *woman* and *women* as Table 1 also showed. However, what is visible in Figure 3 below which was omitted in Table 1, are all remaining descriptors. Each of these often occur with only a single token in the corpus, tend to vary considerably, and are often contradictory. It may be that age or power has some indexical relevance to *women* relative to *females*, with the relational descriptions “older” and “younger” as well as “mature” and “formal” present. Additionally, “middle-aged”, with four tokens, or “Boomer” with two, also seem to pattern alongside these. Such responses still represent a pittance of the data compared to the incidence rate of “normal” and “neutral”, which comprise a plurality of the indexical space for *women*, with the remaining descriptors for *women* functioning as something of an indexical scattershot. This may be expected given the overwhelming neutrality for this variant held by the majority of respondents. For a picture of this indexical scattershot, consider the word cloud depicted in Figure 3 below, which contain "blonde", "authoritarian", "estranged", and "child", which do not point to a consistent conception of an identity or behaviour and seem to be a case of particular respondents grasping at straws to provide any response at all.

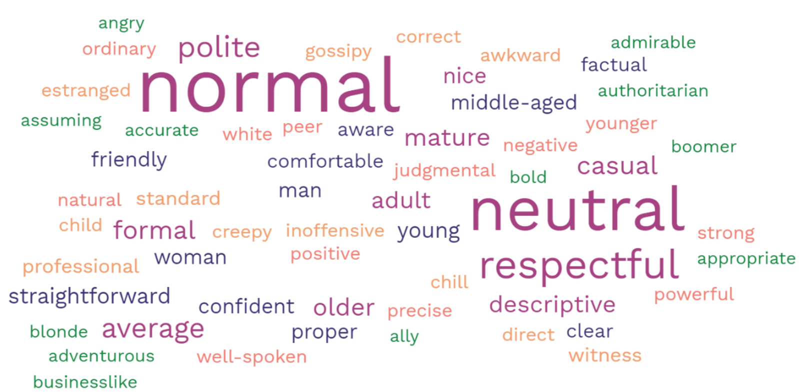

Figure 3: A word cloud displaying all descriptors provided by participants who are both native English speaking and originating in North America. These speakers were isolated from the rest of the corpus to attempt to control for conflicting variables. Note that the colours within the word cloud are automatically generated by the software which created it to aid readability, and do not correspond to any method of categorization.

Conversely, Figure 4 illustrates the strong and commonly repeated indexical connections respondents provided for the *female* variant. Given the much larger number of responses relative to what was depicted in Figure 3, Figure 4 only shows descriptors with at least two tokens within the corpus. Note that even with this low bar of only two instances, neither "normal" nor "neutral" manages to make it onto this word cloud. Additionally, where some of the associations for *women* and *woman* were so varied that generalization became difficult, Figure 4 shows that there is a shared conception of the sort of person who utilizes this variant. This person is, at the most charitable interpretation, a scientist, often a biologist. Next, and somewhat less charitably, the speaker may have been imagined to be in law enforcement, or simply a non-native speaker of English lacking pragmatic awareness. Regardless, these charitable interpretations still carry an association with being clinical despite the survey instructions specifying that the speech stimulus occurred in a non-clinical setting, suggesting at minimum a baseline of inappropriateness. Beyond these outliers, the respondents clearly communicate that this variant is associated with “misogynists”, “chauvinists” and “incels”. Its usage is “dehumanizing”, “rude”, “insulting”, and “judgemental”. Those who say it are seen to be inherently “dangerous”, “odd”, and “robotic”. Further, by the usage of this variant the speaker distances themselves and becomes like an “alien”, with some respondents even specifying to “ferengi”, the enormously eared, orange and ultra capitalist misogynists of Star Trek: The Next Generation and Deep Space Nine mentioned in Chapter 1.

Figure 4: A word cloud constructed from the responses of native speakers of English who originate from North America. Note that given the larger pool of provided descriptors this cloud, unlike figure 3, this cloud depicts descriptors with at least two tokens. As with the previous word cloud, text colours are for readability, rather than any method of categorization.

4.3 The Twitter Data

Utilizing Fireant's keyword search function, 5000 tweets containing either *females* or *women* were harvested. After the exclusions outlined in 3.2.1 were eliminated, the corpus comprised 548 tweets containing *women* and 645 tweets containing *females*, totaling a corpus of 1193 tweets. For these data, there were 634 (53%) tweets written by authors deemed likely to be men, and 559 (47%) deemed likely to be women. Gender would ultimately be the only demographic variable that could be ascertained with any degree of certainty, precluding the possibility of a robust multivariate analysis. Age was, unsurprisingly, given the difficulties discussed in previous literature (Simaki et al., 2016), essentially impossible to discern with any degree of accuracy. Similarly, while a small percentage of users within the corpus listed a country of origin, or had geolocation active while tweeting, this did not represent a sufficient percentage of the corpus to warrant analysis, even ignoring the problematic nature of grouping geolocation data and stated country of origin together as a single variable. Ethnicity was functionally impossible to discern for the vast majority of users. While a very small number of users appeared to self-identify their ethnicity in their Twitter bio, this did not represent anything beyond a negligible portion of tweets harvested by Fireant.

Author gender was determined to a fair level of confidence through author self-identification, or to a lesser degree through the examination of their screen name or display picture, tweets were assigned a binary probable author gender. Tweets where the author's gender was impossible to determine, or were posted under the auspices of an organization were not considered. This unfortunately also precluded tweets written by a very small number of self identified trans and non-binary authors, and admittedly, may have lead to the unintentional misgendering and mistaken inclusion of some tweets.

Tweets were first evaluated in a binary fashion as to whether they contained red flag ideologies. Within these data, there are 468 (%) red flagged tweets. This was further broken down with a system of tags which could be applied to a tweet. Broadly, the discourses observed within these data could be broken down into five such tags; "Misogyny" and "Transphobia" which, unsurprisingly all also received a "red flag" designation" and "Scientific/Technical" which also uniformly did not. "AAVE" tagged tweets could variably contain red flags or not, as well as "General" tweets. As an example, a tweet which was tagged as "AAVE" might be communicating misogyny, transphobia, or in some cases overt conspiracy theories and would be labelled a "AAVE" and "red flag". Similarly, while "General" tagged tweets which usually lacked any sort of ideological content and were often merely conversational, there were still a very small number (fewer than 10) of tweets which utilized *women* while expressing vaccine conspiracies, thus these tweets were labelled "General" and "red flag", which the majority of "General" tweets were "No red flag". "General" tweets comprise a plurality of these data (45%). "Science" tagged tweets make up the smallest portion of the corpus (4%). This tag was applied to tweets which discussed scientific, often regarding natural biology and was largely omitted from most analyses. Tweets which received the "Misogyny" (26%) and "Transphobia" (14%) label communicated ideas or beliefs consistent with their eponymous labels and comprised the majority of all "Red Flag" identified tweets.

Most complicated of the tags, and separate from the aforementioned issue of identifying author ethnicity, is the probable identification of AAVE usage in tweets within these data through the identification of features such as the absence of copula *be*, utilizing *ain't* in negation, multiple negation or vernacular lexical features such as *finna* (Helgotsson 2021). However, the probable identification of AAVE does not automatically reveal the ethnicity of a tweet's author. Tweets found to be utilizing two or more of these metrics were tagged with the label of "AAVE" for the purposes of analysis.

Mixed-effects regression analysis using Rbrul was used to find the best model for the data. There did not appear to be a statistically significant correlation between the usage of either lexical variant and the general incidence of red flag ideologies (see Table 7 below), nor does it appear as though speaker gender conditions the incidence of red flag ideologies to any significant degree.

			Rbrul Results with Women as the Application Value		
Input Probability					.459
Deviance					1619.722
Total n					1193
		Factor Weight	Log-odds	%	n
Presence of Red Flags	p=2.71e-05				
No		0.565	.26	60.1	723
Yes		0.435	-.26	39.9	468
Gender	p=.18				
Women		0.52	.0811	46.7	558
Men		0.48	-.0811	53.3	533
Range			.52		
r^2			0.024		

Table 3: Rbrul results with *women* as the application value.

Females, however, appears to be strongly correlated with speakers communicating on Twitter in AAVE, though it must be noted that AAVE-tagged tokens represent a comparatively small number of tokens within this corpus, and the identification of AAVE speakers was difficult to discern purely from a user's Twitter behaviour. Regardless, this appears to be a highly suggestive correlation. Comparatively, *females* occurs less frequently than *women* in casual and conversational discourse styles in standard English which comprise the majority of the *general* tag. Both of these correlations appear to be statistically significant.

			Rbrul results with Females as the application value		
Input Probability					0.524
Deviance					1381.212
Total n					1143
		Factor Weight	Log-odds	%	n
Tag	p=2.08e-42				
AAVE		0.234	1.186	15.0	171
Misogyny		0.467	0.134	23.5	269
Transphobia		0.508	-0.0306	14.4	165
General		0.784	-1.2894	47.1	538
Range			2.4754		
r^2			0.204		

Table 4: The incidence of f*emales* compared to *women* by likely author gender and tag.

Finally, when the red flag identified tweets are broken down and instances of misogyny are separated from transphobia there does appear to be a statistically significant correlation between men expressing misogynistic sentiments and utilizing the f*emales* variant. It should be stated that this is a fine grained analysis and it requires the exclusion of a full third of the corpus, specifically tweets which include transphobia, in order for this correlation to become apparent. Thus, while it *is* of interest that men do appear to use the *females* variant more often to express misogyny, this analysis requires both a considerable amount of data to be excluded. Additionally, it is perhaps less than revolutionary to say that “men tend to express misogyny more than women,” rendering the utility of this finding questionable.

5.0 Discussion

During the design of this MRP, *The Language and Gender Survey* was initially an afterthought, a way to find the control variant to pit against *females*. In many ways, its limitations, such as the exclusion of an obvious question about participant ethnicity, or the lack of a corresponding pair of stimuli related to *gentleman* to pair with those stimuli related to *lady*, reflect this. *The Language and Gender Survey* stemmed from the desire to do due diligence by experimentally confirming the informal discussion in my social circles that speakers utilizing the *female* variant were off-putting. From there, fundamental experimental praxis required this one question to be at least nominally disguised, which led to the inclusion of other variables. It then became apparent that these other variables could be used to select an experimental control for the Twitter data analysis. Finally, I included the questions which called for participants to optionally provide descriptors to provide more than simply numerical results.

I provide all this to background how the descriptors themselves which were provided by my participants very quickly eclipsed the rest of the study in terms of the amount of insight provided. Consider Figures 5 and 6 below, which present indexical fields (Eckert, 2008) constructed for *women* and *females* respectively. This level of detail would have been impossible had the *Language and Gender Survey* merely been a series of Likert-scale questions. Instead, we are able to glean that, for the overwhelming majority of respondents consisting mostly of Millennial and Gen-Z participants, *females* is not merely more negatively perceived than *women*, but also carries radically different indexical values.

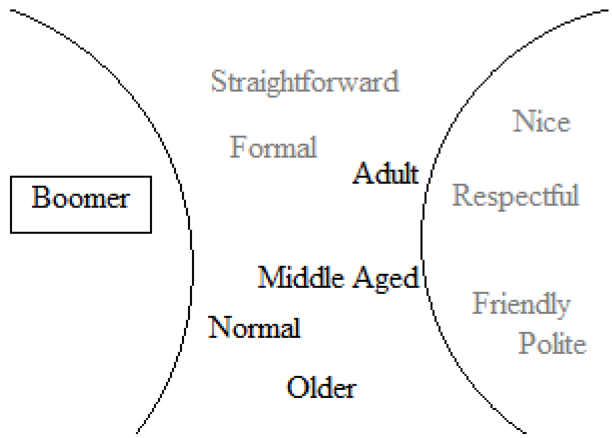

Figure 5: An indexical field for the lexical variant *women*. Stances are in grey, permanent qualities are in black and boxed items represent social types. The left semi-circle represents items associated with a negative polarity, and the right circle represents those with a positive polarity. Observe that only "Boomer" possesses negative polarity as it, unlike the neutral "Middle-Aged", co-occurred with more negative Likert ratings.

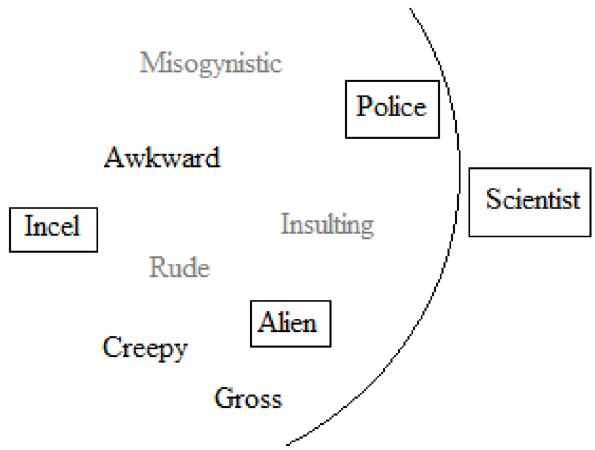

Figure 6: An indexical field for the *females* variant. As above, stances are in grey, permanent qualities are in black and boxed items are social types. Unlike in Figure 5, the vast majority of items possess negative polarity, with only "scientist" being neutral. *"Police"* was the next least negative of the remaining social type, however, it still co-occurred with mostly negative Likert ratings.

That said, this study asked its two fundamental questions: "Is usage of *females* more negatively perceived relative to other variants for *women* ?" and "Does usage of the *females* variant co-occur with red flag ideologies more often than a control variant?" These are, in turn, answered with a "yes", as seen in Figures 1 and 2, as well as Figures 3 and 4; and a "sometimes", as seen in Tables 3 and 5 respectively. This discrepancy begs a third further question: namely, why the disconnect?

While the methodological issues inherent in this instantiation of the *Language and Gender Survey* make drawing firm conclusions difficult, I surmise that the key to understanding this mismatch is in the frequency at which "normal" and "neutral" were provided as descriptors for *women* and *woman*. Consider for a moment that my previous research on the usage of these variants in "Pickup Artist" communications found that, at least within the community I was studying, neither *females*, nor *women* were associated with expressions of misogyny (Ferley, 2021). To paraphrase Deborah Cameron (1985), any word when in the mouth of a sexist may be sexist. So too, it seems possible for any word uttered by a transphobe to become transphobic. Add Schulz' (1975) observations on the consistent derogation of all femme-indexing language, and the solution becomes more likely to be pragmatic than lexical in nature.

Thus, I propose that, in the face of Twitter data showing that both variants are more or less equally likely to be associated with red flag ideologies, that listeners are reacting more to the oddness of hearing or reading *female* used in non-clinical settings. This pragmatic strangeness, I suspect, leads to a greater salience of anything else that is "off" in the speech of the person utilizing *females* as a lexical variant. Recall that "weird" and "odd" had 42 and 19 repetitions within the corpus respectively and 8 respondents imagined the speaker to be a "non-native speaker" of English. Further, at least one respondent explicitly found the usage of the *female* variant to be ungrammatical. Thus, in the face of both variants equally co-occurring with red flags in the Twitter data, it seems unlikely that listeners are attending to differences in semantic content between the variants. Rather, this salience causes listeners to create very specific and consistent personas for those who use *females* based upon the red flags that they are now actively searching for in the speech of those people.

There is one other finding that these data within this study suggests. As mentioned previously, it would have been enormously helpful to have more demographic information for each respondent of the *Language and Gender Survey.* My knowledge of the identities of those with whom the survey was initially administered, and of their social circles, suggests that at minimum a considerable portion of the respondent pool were affluent, possessing post-secondary education, and caucasian. This becomes problematic when the Twitter data suggests that usage of *females* as a lexical variant is strongly correlated with AAVE, and survey respondents described hearing usage of that variant as “weird”, “odd” and most importantly, “dangerous”.

AAVE speakers within the Twitter corpus used *females* as a neutral polarity lexical item to express all manner of sentiments. The overwhelmingly negative reaction the survey respondents had to usage of this variant suggests that, at least in some forums and spaces wherein AAVE and mainstream English speakers interact, usage of this variant could be the source of, at minimum, misunderstanding or inappropriately assessing a speaker as *dangerous* or *weird*.

6.0 Next Steps and Conclusions

6.1 Next Steps

With regard to *The Language and Gender Survey*, as discussed previously, the omission of a question related to participant ethnicity was at minimum, an unfortunate oversight. While it should again be stated that the survey was in itself conceived as a vehicle to isolate a control variable for the Twitter data analysis, it quickly grew into the portion of this study that was the most interesting and exciting. Thus, a reduction in the utility of this corpus of information due to the omission of ethnicity as an independent variable for analysis is, in retrospect, a disappointing oversight.

In future research, while I have not run statistical analysis on whether question order affected participant results both in Likert ratings and descriptive items provided, it certainly appears that it may well have. Specifically, participants more or less “figured out” what task was being asked of them as they went along. As discussed above, the polarity of descriptors provided remained the same between the plural and singular stimuli; however, it appears as though participants “fence sat” somewhat less in the later questions with the singular stimuli for the “female” variant garnering fewer 5’s than its plural counterpart. To this end, randomizing the question order would be advisable.

In another iteration, it would behoove the researcher to be much more explicit in the survey instructions. While many participants, especially in an unpaid online survey, often do not read instructions, or fail to read closely enough to understand the task required of them, there were many participants who followed up with me after they took the survey to say that they didn’t really understand what was being asked of them. One participant appears to have unintentionally exemplified this confusion when they wrote, “I love women " in the descriptor field provided for that variable and wrote “I’m scared of men” for the stimulus for *men*. While it is safe to assume junk responses will always occur in surveys, this still suggests that greater care in survey construction could be attended to in further research.

Finally, in regards to the survey, a criticism I received from some of those who took the survey, in particular, from some who possessed a relevant academic background, was that they wanted the stimuli to specifically state the gender and racial identity of the speaker the question asked them to imagine. I would caution a future researcher attempting to replicate this work against incorporating this suggestion, especially if they are relying on similar free methods of data collection. Providing speaker identity will require asking duplicated questions for multiple gender and ethnic combinations and will dramatically and negatively affect the completion rate of participants. If a future researcher is able to compensate participants, these inclusions could be useful. However, I do wonder if the extra labour will reveal anything significantly different than this work did given the general absence of racial cognisance in respondent associations between *females* and AAVE.

Moving now to the second half of this study, the Twitter data analysis also presents a number of confounding issues. These exist both in experimental design as well as those of the discursive environment in which the tweets were collected. First, and perhaps the most regrettable, was that I alone was the one deciding whether an individual tweet included red flags. While I made an attempt to remain objective in these assessments, I was and am still prejudiced against a speaker utilizing *females* instead of *women* when I am agnostic of their identity. With this knowledge, it is entirely possible I was uncharitable towards some, or perhaps even many Twitter users, in my assignment of red flags. A simple method of correcting for this would be to locate a person more neutrally disposed towards usage of the "females" variant and have them independently analyze the tweets in parallel with me. Indeed, such independent rating methodologies are common in impressionist phonetic analysis and could well be employed in research such as I have done.

It also cannot be ignored that the tweets this study utilized were gathered during a particularly tumultuous period in public discourse. The demonization of transgender and queer individuals as "pedophile groomers" by significant portions of even the mainstream of the North American right wing may have been especially problematic as transphobia was one red flag that was specifically looked for. Large numbers of tweets referred to a transgender woman who had just won a swimming competition. More still mentioned a cis-woman who was angry about losing to a (different) transgender woman at a skateboarding competition. Finally, many more tweets were attempting to entertain serious discussion around a transphobic documentary called "What is a Woman?" by the theocratic fascist (as self-described in his Twitter bio) contributor to The Daily Wire, Matt Walsh. These events represent a confluence of topics contributing to transphobic discourse and may have caused this study to present a "false positive" in regards to the levels of transphobia attached to both variables studied in the Twitter data analysis. It remains to be seen, however, whether this level of transphobic expression in the public discourse represents a momentary inflammation and future analysis will show different results, or if this level of transphobia is a new normal in which case future study may confirm these results.

That said, this research has shown the utility of studying lexical items in the construction of indexical fields for which phonological variables figure predominantly. In variationist research, the study of the lexicon has been overshadowed by research in phonology and morphosyntax. This research is in part to call for more research along the lines I have presented here, which complements research on other levels of language.

6.2 Conclusions

Schulz (1975) observed that terms of reference for women had almost invariably become imbued with negative connotations. *Women* had become negatively associated with sex work, so the zeitgeist adopted *female* as the preferred variant. However, that too, in time, became perceived as impolite, and once again, negatively associated with sex work. This led to the widespread adoption of *ladies* as the preferred variant. If, as you read *ladies* in this context, you made the connection with the euphemism “lady of the night”, you may, depending on your age, feel some suspicion that *ladies*, had, in turn too, become associated euphemistically with sex work. The results of *The Language and Gender* survey suggest that this progress is actually cyclic and that in the intervening years between Schulz’ chapter and the present day, it has continued. Respondents overwhelmingly communicated that *women* was, as of the time of this writing, perceived most neutrally among all variants.

The word clouds in chapter 4.2 and the indexical fields constructed in chapter 5 do in fact demonstrate that there is variation in the perceptions of consultants for this study. Moreover, the results also demonstrate that, at least among the sampled population, there is an incredible gulf in lexical preference, and a fair amount of variation involving lexical choice. The usage of one variant could imply specific power dynamics in a conversation, or the relative ages of its interlocutors. The usage of another variant might imply extreme politeness or dismissive condescension. The usage of *females* as a variant clearly implies that for many participants in the study the “speaker” is someone to be extremely wary of.

Ultimately however, the variant being employed by a given speaker or author is very open to resignification and they may all be used to express red flag ideologies. Transphobes and misogynists whose Twitter data was analyzed appeared more likely to use *females* when engaging in biological essentialism, or *women* while broadly fumbling toward some reductive definition of womanhood. To close, I will again paraphrase Deborah Cameron: Any word can become the vehicle for bigotry when it comes from the mouth of a bigot.

Appendix

Language and Gender Survey

Section 1- Online Consent Information

Researcher

Andrew Ferley

aferley@gmail.com

Purpose of research: This study, which represents as a portion of a study to be completed in accordance with fulfilling a Master of Arts in Linguistics, studies the ways in which gendered language affects speaker perception.

What you will be asked to do in the research: First, you will be asked to answer five questions about yourself, including your age, generational affiliation, gender, ethnicity, and whether or not you are a native speaker of English. Afterwards, will be provided with a series of sentences which vary by only a single word, specifically a gender-related noun. You will be asked to rate how the sentence would make you feel about a person if the sentence was said to you in a non-medical setting. These ratings are on a five point scale, where a (1) is very negative, and a (5) is very positive. Additionally, each sentence will have an associated comment box and you are encouraged to provide up to three descriptive words which you feel could describe the person saying the sentence.

Risks and discomforts: No risks or discomforts are anticipated during your participation in this study.

Benefits of the research and benefits to you: By participating in this research, you are aiding in the study of gender-related discourse and may find that you have interesting reflections on the language you use in your day-to-day life.

Voluntary participation and withdrawal: Your participation in this study is entirely voluntary. In addition, you may withdraw your consent and participation at any time during the survey by simply closing the window. Be advised, however, that by clicking the final submit button, you are giving your consent that your provided data will be used within the study.

Confidentiality: All data gathered through this study will be held in the strictest confidentiality. Further, at no point will you be providing any personal identifying information such as your name, date of birth or address, All the data that you provide will be cited only with your provided demographic information. It will be stored securely and shared only in summary form.

Further questions about the research? If you have further questions about the research or your role in it, you may contact Andrew Ferley at the email provided at the beginning of this document.

Section 2- A little about you!

In this section you are asked to provide information on whether you are a native speaker of English, what generation you belong to, the year of your birth, your gender identity, and where you were born. This information will not be shared with any personally identifying information but will still be kept in the strictest confidence

1. Are you a native speaker of English?
   Yes
   No
2. Which of the following generations do you belong to?
   Gen Z (Born 1997-2012)
   Millennial (Born 1981-1996)
   Gen X (Born 1965-1980)
   Baby Boomer (Born 1946-1964)
   Pre-War (Born Pre-1945)
3. What year were you born in?
   (Open answer field)
4. Which of the following genders do you identify as?
   Man
   Woman
   Non-Binary
   Other (Open answer field)
5. What country were you born in?
   (Open answer field)

Section 3- Sentence Judgements

In this section, you will read seven nearly identical sentences- which vary only in the final word. Imagine yourself hearing these sentences in casual, non-clinical conversation. Imagine that the speaker is referring to an adult.

For each sentence, please consider the following:

1. How would hearing this sentence make you feel about the speaker? Please rate the speaker on a scale between 1 (Very Negative) and 5 (Very Positive).

2. If you have an impression of the speaker, please provide up to three terms you would use to describe a person saying that sentence.

a) I recognise that man.
b) Please provide up to three descriptive words for the person who said "man."

a) I recognize that boy.
b) Please provide up to three descriptive words for the person who said "boy."

a) I recognize that male.
b) Please provide up to three descriptive words for the person who said "male.”

a) I recognize that woman.
b) Please provide up to three descriptive words for the person who said "woman.”

a) I recognize that girl.
b) Please provide up to three descriptive words for the person who said "girl.”

a) I recognize that female.
b) Please provide up to three descriptive words for the person who said "female.”

a) I recognize that lady.
b) Please provide up to three descriptive words for the person who said "lady.”

Section 4- Sentence Judgements 2

In this section, you will read another seven nearly identical sentences. This time they occur at the beginning of the sentence. Imagine yourself hearing these sentences in casual, non-clinical conversation. Imagine that the speaker is referring to adults.

For each sentence, please consider the following:

1. How would hearing this sentence make you feel about the speaker? Please rate the speaker on a scale between 1 (Very Negative) and 5 (Very Positive).

2. If you have an impression of the speaker, please provide up to three terms you would use to describe a person saying that sentence.

a) Those men live around here.
b) Please provide up to three descriptive words for the person who said "men."

a) Those boys live around here.
b) Please provide up to three descriptive words for the person who said "boys."

a) Those males live around here
b) Please provide up to three descriptive words for the person who said "males."
a) Those women live around here.

b) Please provide up to three descriptive words for the person who said "women."

a) Those girls live around here.
b) Please provide up to three descriptive words for the person who said "girls."

a) Those females live around here.
b) Please provide up to three descriptive words for the person who said "females."

a) Those ladies live around here.
b) Please provide up to three descriptive words for the person who said "ladies."

Section 5- Thank You For Your Participation!

Your responses will be extremely helpful. If you have a minute, it would be really great if you could pass this survey along using this link (https://forms.gle/HwnquDCTPweAxLUC6) to anyone you think might be interested. Posting the link on social media is great, too! I will be accepting responses until at least (INSERT DATE HERE).

Thank you again!

Works Cited

Software

Anthony, L. and Hardaker, C. (2021). *FireAnt (Version 2.0.5)* [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

Sankoff, David, Sali A. Tagliamonte, and Eric Smith (2005). *Goldvarb X: A variable rule application for Macintosh and Windows*. Department of Linguistics, University of Toronto.

Scrivner, O., & Díaz-Campos, M. (2016). Language Variation Suite: A theoretical and methodological contribution for linguistic data analysis. *Proceedings of the Linguistic Society of America*, *1*, 29-1.

Literature

Androutsopoulos, J. (2006). Introduction: Sociolinguistics and computer‐mediated communication. *Journal of sociolinguistics*, *10*(4), 419-438.

Allen, I. E., & Seaman, C. A. (2007). Likert scales and data analyses. *Quality progress*, *40*(7), 64-65.

Banet-Weiser, S., & Miltner, K. M. (2016). # MasculinitySoFragile: Culture, structure, and networked misogyny. *Feminist media studies*, *16*(1), 171-174.

Barua, A. (2013). Methods for decision-making in survey questionnaires based on Likert scale. *Journal of Asian scientific research*, *3*(1), 35-38.

Bartlett, J., Norrie, R., Patel, S., Rumpel, R., & Wibberley, S. (2014). Misogyny on twitter. Retrieved from https://www.demos.co.uk/files/MISOGYNY_ON_TWITTER.pdf

Bayley, R. (2019). Variationist sociolinguistics. *The Oxford handbook of sociolinguistics*.

Beaton, M. E., & Washington, H. B. (2015). Slurs and the indexical field: The pejoration and reclaiming of favelado 'slum-dweller'. *Language sciences*, *52*, 12-21.

Bertram, D. (2007). Likert scales. *Retrieved November*, *2*(10).

Bernhard, E. M. (2020). Because Internet: Understanding the New Rules of Language, Gretchen Mcculloch (2019). *Explorations in media ecology*, *19*(2), 225-228.

Boyle, K., & Rathnayake, C. (2020). # HimToo and the networking of misogyny in the age of# MeToo. *Feminist media studies*, *20*(8), 1259-1277.

Brock, A. (2018). Critical technocultural discourse analysis. *New media & society*, *20*(3), 1012-1030.

Brown, J. D. (2000). What issues affect Likert-scale questionnaire formats. *Shiken: JALT testing & evaluation SIG newsletter*, *4*(1).

Burnett, H. (2020). A persona-based semantics for slurs. *Grazer philosophische studien*, *97*(1), 31-62.

Cameron, D. (2007). *The Myth of Mars and Venus*. Oxford University Press, USA.

Cameron, D. (1985) Feminism and Linguistic Theory. London. Routledge. Revised second edition, 1992.

Coates, J. (2015). *Women, men and language: A sociolinguistic account of gender differences in language*. Routledge.

D'Arcy, A., & Young, T. M. (2012). Ethics and social media: Implications for sociolinguistics in the networked public 1. *Journal of sociolinguistics*, *16*(4), 532-546.

Eckert, P. (2008). Variation and the indexical field 1. *Journal of sociolinguistics*, *12*(4), 453-476.

Eckert, P., & McConnell-Ginet, S. (1995). Constructing meaning, constructing selves. *Gender articulated: language and the socially constructed self*, 469-507.

Ferley, A. (2021). *Misogyny and conspicuous use of female(s)*. Unpublished Manuscript.

Gries, S. T. (2021). Statistics for Linguistics with R. In *Statistics for linguistics with R*. Mouton De Gruyten.

Gruzd, A., Wellman, B., & Takhteyev, Y. (2011). Imagining Twitter as an imagined community. American Behavioral Scientist, 55(10), 1294-1318.

Helgotsson, M. (2021). Grammatical features of African American Vernacular English in the movie Sextuplets: A sociolinguistics study of the speech of the two African American characters Alan and Dawn.

Herring, S. C. (2004). Computer-mediated discourse analysis. *Designing for virtual communities in the service of learning*, 338-376.

Herring, S. C. (2003). Gender and power in on-line communication. *The handbook of language and gender*, 202-228.

Herring, S. C. (2000). Gender differences in CMC: Findings and implications. *Computer professionals for social responsibility journal*, *18*(1),

Hoff, M. (2020). Cerca mío/a or cerca de mí? A variationist analysis of Spanish locative+ possessive on Twitter. *Studies in Hispanic and Lusophone linguistics*, *13*(1), 51-78.

Hoffman, M. F., & Walker, J. A. (2010). Ethnolects and the city: Ethnic orientation and linguistic variation in Toronto English. *Language variation and change*, *22*(1), 37-67.

Ilbury, C. (2020). "Sassy Queens": Stylistic orthographic variation in Twitter and the enregisterment of AAVE. *Journal of sociolinguistics*, *24*(2), 245-264.

Jones, T. (2015). Toward a description of African American vernacular english dialect regions using “Black Twitter”. *American speech*, *90*(4), 403-440.

Kitzinger, C. (2005). Heteronormativity in action: Reproducing the heterosexual nuclear family in after-hours medical calls. *Social problems*, *52*(4), 477-498.

Labov, W. (1986). The social stratification of (r) in New York City department stores. In *Dialect and language variation* (pp. 304-329). Academic Press.

Lakoff, R. (1973). Language and woman's place. *Language in society*, *2*(1), 45-79.

Lambert, W. E., Hodgson, R. C., Gardner, R. C., & Fillenbaum, S. (1960). Evaluational reactions to spoken languages. *The Journal of abnormal and social psychology*, *60*(1), 44.

Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A., & Christakis, N. (2008). Tastes, ties, and time: A new social network dataset using Facebook. com. *Social networks*, *30*(4), 330-342.

McConnell-Ginet, S. (2015). Ehrlich, S., Meyerhoff, M., & Holmes, J. (Eds.). (2014). The handbook of language, gender, and sexuality : Handbook of language, gender and sexuality. John Wiley & Sons, Incorporated.

McConnell-Ginet, S. (2008). Words in the world: How and why meanings can matter. *Language*, 497-527.

McConnell-Ginet, S. (2003). What’s in a name? Social labeling and gender practices. *The handbook of language and gender*, 69-97.

McConnell-Ginet, S. (1989). The sexual (re) production of meaning: A discourse-based theory. Language, gender, and professional writing: Theoretical approaches and guidelines for nonsexist usage, 35.

Memon, S. A., Tyagi, A., Mortensen, D. R., & Carley, K. M. (2020, October). Characterizing sociolinguistic variation in the competing vaccination communities. In *International conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior representation in modeling and simulation* (pp. 118-129). Springer, Cham.

Milroy, L. (2000). Social network analysis and language change: Introduction. *European journal of English studies*, *4*(3), 217-223.

Milroy, J., & Milroy, L. (1985). Linguistic change, social network and speaker innovation1. *Journal of linguistics*, *21*(2), 339-384.

Misiejuk, K., Scianna, J., Kaliisa, R., Vachuska, K., & Shaffer, D. W. (2021, February). Incorporating sentiment analysis with epistemic network analysis to enhance discourse analysis of Twitter data. In *International conference on quantitative ethnography* (pp. 375-389). Springer, Cham.

Mukherjee, S., & Bhattacharyya, P. (2012, December). Sentiment analysis in twitter with lightweight discourse analysis. In *Proceedings of COLING 2012* (pp. 1847-1864).

Nemoto, T., & Beglar, D. (2014). Likert-scale questionnaires. In *JALT 2013 conference proceedings* (pp. 1-8). Tokyo: Jalt.

Paolillo, J. C. (2001). Language variation on Internet Relay Chat: A social network approach. *Journal of sociolinguistics*, *5*(2), 180-213.

Jeff Sauro, P. D. (2016, May 24). *Can you take the mean of ordinal data?* MeasuringU. Retrieved May 10, 2022, from https://measuringu.com/mean-ordinal/

Schulz, M. R. (1975). The semantic derogation of woman. Language and sex: Difference and dominance. Ur. B. Thorne, N. Henley.

Simaki, V., Mporas, I., & Megalooikonomou, V. (2016, April). Age identification of twitter users: Classification methods and sociolinguistic analysis. In *International conference on intelligent text processing and computational linguistics* (pp. 385-395). Springer, Cham.

Siegle, D. (2015, June 30). *Likert scales*. Educational Research Basics by Del Siegle. Retrieved May 10, 2022, from https://researchbasics.education.uconn.edu/likert_scales/#

Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language & communication*, *23*(3-4), 193-229.

Simaki, V., Mporas, I., & Megalooikonomou, V. (2016, April). Age identification of twitter users: Classification methods and sociolinguistic analysis. In *International conference on intelligent text processing and computational linguistics* (pp. 385-395). Springer, Cham.

Tatman, R. (2015). # go awn: Sociophonetic variation in variant spellings on Twitter. *Working papers of the linguistics circle*, *25*(2), 97-108.

van Dijk, T. A. (2005). Critical discourse analysis. *The handbook of discourse analysis*, 349-371.