Title
The theoretical and methodological implications of bisexuality in language and sexuality studies

Author
Willis, Chloe Marie

Publication Date
2023

University of California
Santa Barbara

The theoretical and methodological implications of bisexuality in language and sexuality studies

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Linguistics

by

Chloe Marie Willis

Committee in charge:

Professor Lal Zimman, Chair
Professor Tania Israel
Professor Argyro Katsika
Professor Simon Todd

September 2023

The Dissertation of Chloe Marie Willis is approved.

Professor Tania Israel

Professor Argyro Katsika

Professor Simon Todd

Professor Lal Zimman, Committee Chair

August 2023

The theoretical and methodological implications of bisexuality in language and sexuality studies

by

Chloe Marie Willis

For my fellow bisexuals.

Acknowledgements

I have tremendous gratitude for everyone who made this dissertation possible.

I would especially like to thank the members of my committee: Tania Israel, Argyro Katsika, Simon Todd and Lal Zimman. Thank you for knowledge, guidance, time and support. To Tania Israel, thank you for your curiosity towards linguistics and for the haikus. Your perspective as a non-linguist was invaluable for shaping this dissertation. To Argyro Katsika, thank you for your discerning eye for detail and for sharing your well of phonetic and experimental knowledge. You were always a safe person with whom I could polish rough ideas into something beautiful.

To Simon Todd, thank you for your collaboration and technical mentorship. The skills you taught me defined my career path beyond academia. I deeply appreciate your patience and thoroughness when answering my many questions.

To Lal, who has been with me through every step of my graduate career. Thank you for sharing your knowledge, for being my soundboard, for supporting my shower-thought-turned-research ideas. I feel incredibly lucky to have had an advisor who was always in my corner and saw the best in me.

This dissertation would not have been possible without all of my participants and the users in the Twitterverse. Without your efforts, insights, and musings, this dissertation would have never come to pass. Thank you, and please keep being you.

To my fellow Linguistics graduate students, who shaped my graduate experience for the better. Thank you all for the laughs and camaraderie. To Jamaal Muwwakkil, for the thought-provoking conversations and shenanigans. To Karen Tsai, for your mentorship and kindness. To Chris Brendel, for your technical and emotional support. To Julia Fine, for the insightful (and fun) research we built together. To Giorgia Troiani and Jazmine Exford, it was an honor to grow together as scholars through our peer review sessions. To Jordan Douglas-Tavani, for being a great neighbor and friend.

To the friends I made along the way, thank you for being in community with me. I would like to extend a special thanks to Eugene Riordan, Keene Morrow, Carl Gabrielson, Karen Rice, Megan Koth, and Mary Michael. Playing Dungeons & Dragons with you was a precious and hilarious escape. To Jordan Tudisco, for enduring the trials of leadership with me. To Naomi Joseph, for finding my signature look. To Justin Kelz and Erin Horowitz, for making me and Adam feel at home. You all made Santa Barbara a better place to live.

To my friends since always, thank you for grounding me through this experience. To Sam Dennis, Troy Edwards, Jessica Edwards, Claire Jadulang, David Sefo, Diana Lustig, Lani MacLean, Halan Guedi, Sami Muscio, and Kerri Blake-Cavanaugh, thank you for reminding me that I am more than my degree.

To my family, thank you for weathering this storm with me. To my husband Adam, thank you for sharing my triumphs and my failures. Your intentionally incorrect use of linguistic terminology never failed to put a smile on my face. To my parents, thank you for fostering a love for learning and a desire to achieve from the very beginning.

Last but not least, I would like to thank my corgi, Tank. Your sweet smile kept me going when I felt like giving up. You gave me love and purpose. Thank you, I love you.

Curriculum Vitæ
Chloe Marie Willis

Education

2023	Ph.D. in Linguistics (Expected), University of California, Santa Barbara.
2020	M.A. in Linguistics, University of California, Santa Barbara.
2014	B.A. in Language Studies (Japanese), University of California, Santa Cruz

Publications

Willis, Chloe. Forthcoming. Sôshokukei kara asuparabêkon made (‘From herbivores to bacon-wrapped asparagus’): Binary gender taxonomies and neoliberal self-making in modern Japan. Gender and Language 17(3).

Willis, Chloe. Forthcoming. Bisexuality in Experimental Sociophonetics: Ideologies and Implications. Journal of Language and Sexuality 14(1).

Willis, Chloe & Chadi Ben Youssef. 2023. Random bisexual forests: Intersections between gender, sexuality, and race in /s/ production. Proceedings of the Linguistic Society of America 8(1): 5504. `https://doi.org/10.3765/plsa.v8i1.5504`

Willis, Chloe. 2021. Bisexuality and /s/ production. Proceedings of the Linguistic Society of America 6(1): 69-81. `https://doi.org/10.3765/plsa.v6i1.4942`

Babel, Molly, Grant McGuire & **Chloe Willis**. 2021. The Role of Voice Evaluation in Voice Recall. In: Weiss B., Trouvain J., Barkat-Defradas M., Ohala J.J. (eds) Voice Attractiveness: 101-124. Springer, Singapore. `https://doi.org/10.1007/978-981-15-6627-1_6`

Professional appointments

2022 - present	Consultant, Textio, Inc.
2020, 2022	Teaching Associate, Department of Linguistics, University of California, Santa Barbara
2018 - 2023	Teaching Assistant, Department of Linguistics, University of California, Santa Barbara
2018, 2020	Teaching Assistant, Department of East Asian Languages and Cultural Studies, University of California, Santa Barbara

Abstract

The theoretical and methodological implications of bisexuality in language and sexuality studies

by

Chloe Marie Willis

Language communicates not just what we say, but also who we are. Research demonstrates that speakers use linguistic variation to construct sexual identities and communities. Driven by the question of whether stereotypes about lesbian and gay speech have "real" bases in speech production or perception, the literature in this area is historically preoccupied with differences between lesbian, gay, and straight people. White and cisgender positionalities are overrepresented in this work. As a result, the understanding of sexuality that emerges in this area is white, monosexual, and often mapped onto gender expression. Bisexuality and its intersections with other aspects of identity are not widely considered, leaving the theorization of sexuality incomplete. This dissertation extends the theorization of sexuality by positioning bisexual practices and subjectivities as central to understanding sexuality. The first and second content chapters present quantitative analyses of acoustic data. Both chapters explore how bisexual people produce /s/–the sound associated with the "gay lisp" stereotype–relative to lesbian, gay, and straight people. The first content chapter focuses on the theorization of the gender-sexuality interface, whereas the second content chapter focuses on the intersection between sexuality, gender, and race. The third content chapter examines how people talk about bisexuality online using a statistical approach called keyness analysis. The chapter provides insights into the concepts and ideologies that characterize discourses of bisexuality online, and identifies numerous avenues for further research on bisexual identity construction. All three content chapters illustrate the value of centering peripheral identities in linguistic research. This dissertation represents the largest body of research on bisexuality within linguistics to date.

Contents

Vita vii

Abstract viii

1 Introduction 1

1.1 Permissions and Attributions . . . 9

2 Bisexuality in experimental sociophonetics: Ideologies and implications 10

2.1 Introduction . . . 10
2.2 Sexuality and the voice . . . 12
2.3 Methodology . . . 17
2.4 Statistical results . . . 21
2.5 Qualitative analysis of post-test surveys . . . 29
2.6 Discussion . . . 31
2.7 Conclusion . . . 34

3 Modeling the social dimensions of /s/ production: (Bi)sexuality, race, and gender in random forests 36

3.1 Introduction . . . 36
3.2 /s/ . . . 42
3.3 Data . . . 47
3.4 Results . . . 51
3.5 Discussion . . . 64
3.6 Conclusion . . . 68

4 *Happy*, *#horny*, *valid*: A keyness analysis of bisexual discourses on Twitter 70

4.1 Introduction . . . 70
4.2 Data . . . 72
4.3 Analysis . . . 79
4.4 Results . . . 82
4.5 Discussion . . . 98

4.6 Next steps: Dispersion and alternative measures 101
4.7 Conclusion 105

5 Conclusion **106**

Chapter 1

Introduction

A central goal within sociolinguistics and linguistic anthropology is to investigate how speakers construct, perform, understand, and negotiate identity through language. People position themselves as particular kinds of speakers—through their words, their grammar, and their voices—every time they talk. Listeners, in turn, use linguistic variation to interpret and assign identity categories. For example, there is a widespread belief that English-speaking gay men "talk like women" (e.g., speak with higher pitch, pronounce /s/ in a way that sounds "lispy") and that lesbian women "talk like men" (e.g., speak with lower pitch, use monotone intonation). Linguists, anthropologists, and cognitive scientists have invested considerable effort in attempting to prove or disprove whether these stereotypes have "real" perceptual or productive bases (Campbell-Kibler 2007, 2011; Gaudio 1994; Levon 2006; Moonwoman-Baird 1997; Munson et al. 2006b, 2006a; Pierrehumbert et al. 2004; Smyth et al. 2003; Zimman 2013). As a result, the interdisciplinary field of language and sexuality studies is historically characterized by a preoccupation with how gay men (and sometimes lesbian women) differ from straight people. What these studies fail to recognize is that this specific understanding of sexuality—one that is binary, white, and mapped onto gender expression—comes from a cultural, rather than theoretical, framework of sexual identity. This dissertation expands the theorization of sexuality beyond this specific cultural understanding through an investigation of bisexuality and its intersections with other axes of identity.

First, it is useful to define bisexuality. As a self-identified bisexual person, my working definition is that bisexuality is being attracted to two or more genders. This definition is common, though not uncontested, in the in-person and online queer and trans communities in which I participate. Looking outside the communities with which I engage, the crowdsourced online dictionary *Urban Dictionary*, which often provides a (playful) perspective on public views, includes the following definitions (*Urban Dictionary* 2021):

1. The gift of nature to love both genders indiscriminately.

2. Attraction to two or more genders. This can sometimes be confused with pansexuality. Pansexuality is attraction to all genders or attraction regardless of gender; bisexuality means the person is attracted to at least two, but not all genders. A bisexual person does not have to be attracted to males and females specifically; for example, one may be attracted to females and nonbinary people.

3. Being attracted to all women and 4 very specific types of men only.

The first definition situates bisexuality relative to a binary-based understanding of gender with the use of “both genders” to describe the potential objects of desire. In contrast, the second definition describes bisexuality as “attraction to two or more genders” and further specifies that “a bisexual person does not have to be attracted to males and females.” This definition also establishes a distinction between bisexuality and *pansexuality*, another plurisexual identity term which is described as “attraction to all genders or attraction regardless of gender.”[1] The third definition provides a humorously narrow definition that suggests bisexual people may not be attracted to different genders equally.

[1] *Plurisexual* refers to people who are, or have the potential to be, attracted to multiple genders.

These *Urban Dictionary* definitions exhibit similar patterns to those reported in the literature on bisexuality. Galupo et al. (2017) report a mixed-methods analysis of how people under the "bisexual umbrella" (i.e. people who identify as bisexual, pansexual, or queer) define their sexualities. They find that bisexual participants describing their sexualities in the following ways (among others) (Galupo et al. 2017):

4. I find myself romantically and physically attracted to both men and women, albeit not equally, I would say that my attraction level is about 65% toward men and 35% toward women.

5. I am romantically and sexually attracted to all genders and orientations except for cis straight men.

6. I'm attracted to my gender and other genders. I'm not attracted to all genders, therefore not identifying as pansexual.

7. I fall in love with the heart not the anatomy. I'm an equal opportunity employer.

These definitions, too, indicate variable understandings of gender (e.g. "both men and women" versus "all genders") and ways of experiencing attraction ("not equally" versus "equal opportunity employer"). All in all, these examples illustrate that there is not a clear, singular definition of bisexuality. It can refer to a range of attraction styles, desires, and understandings of gender and the gender-sexuality interface. For the purposes of this dissertation, however, I am not interested in defining what is or is not bisexual (versus pansexual or any other plurisexual identity term). Rather, I am interested in how bisexuality in all its forms informs how people experience and navigate the social world. I understand a bisexual person to be anyone who identifies with the label "bisexual". Being a linguist, I situate my exploration of this question within what people say and how they say it. Before I present the chapters exploring this question, however, I address the state of bisexuality in language and sexuality studies in order to establish what it contributes to the literature.

A critical examination of methodologies used in perception studies reveals theoretical shortcomings in language and sexuality research. Listeners are typically prompted to evaluate decontextualized voices in order to identify what makes someone sound lesbian/gay. In the forced-choice paradigm (e.g. Smyth et al. 2003), listeners are typically limited to a binary decision between lesbian/gay or straight. This paradigm precludes the possibility of a voice sounding anything other than lesbian, gay, or straight. In a Likert scale paradigm (e.g. Munson et al. 2006a), listeners are generally asked to evaluate voices on an odd-point scale in which one end represents lesbian/gay-sounding and the other straight-sounding. This paradigm too renders unintelligible voices that are not clearly lesbian/gay- or straight-sounding, except perhaps in the undefined middle of the scale. It is not clear what, or rather who, the middle of the scale represents, nor is there any empirical reason to assume all participants interpret the scales in the same way. The uncritical way these methods are applied highlights the extent to which research in this area is couched in binary, monosexist ideologies of sexuality. *Monosexism* refers to the ideology or assumption that people are, or should be, attracted to only one gender (Eisner 2016). Bisexual erasure is institutional and systematic, even in sexuality research (Alarie & Gaudet 2013; Israel 2018; Wilkinson 2019).

Language and sexuality research is dominated by whiteness on multiple fronts. Given that one of the primary motivations driving this research area is related to the stereotype that gay men “talk like women”, the variables examined in these studies are generally associated with speech styles linked to “women’s language” (c.f. Lakoff 1973) which historically focuses on white speakers (Bucholtz & Hall 1995). In turn, the linguistic features examined in language and sexuality studies are frequently linked to gendered whiteness, as opposed to other racialized speech varieties such as African American Language. Additionally, quantitative language and sexuality studies rarely report the ethnic and racial identities of their participants. As a result, ethnoracial identity is not included as a factor in studies using statistical modeling, precluding quantitative results demonstrating a relationship between ethnicity/race and sexuality. The exclusion of queer people of color indexes an understanding of sexuality that is not intersectional, indicating the ductilack of theorization of the interface between sexuality and race in this area of research.

The intersection between sexuality and gender is also undertheorized. The interplay between the two is often left implicit in language and sexuality studies, rendering the theorization of their intersection incomplete (Zimman 2013). Lesbian and gay speech styles are typically interpreted as adherence to or deviation from the straight norm. Some studies specifically investigate whether lesbian and gay speech styles are “whole-scale adoptions of sex-opposite speech patterns” (Munson et al. 2006a: 234), effectively conflating sexuality with gender expression. As a result, lesbian and gay participants are positioned as gender non-normative in such studies and heterosexuals as gender normative. Bisexual participants too are positioned in this way, as we are typically grouped with lesbian and gay speakers a priori. However, this positioning does not reflect how bisexuality is negotiated in everyday life. Bisexual people are often misunderstood as straight when they express different-gender attraction. Those same individuals may be perceived as lesbian/gay when they express same-gender attraction. Of course, there are bisexual people who are perceived as non-normative based on their own gender presentation or other indexes of sexuality. These heuristic situations illustrate that the fluid, co-constructed, and context-dependent nature of sexuality is particularly salient for bisexual people. The way bisexuality is understood in daily life is not adequately captured in language and sexuality studies, such that more theoretical attention to the gender-sexuality interface is necessary.

What is needed to address these theoretical and methodological issues is a theory of (sexual) identity that is disentangled from ideology. My dissertation answers this call, extending the theorization of sexuality beyond monosexist, white ideologies. Drawing from trans linguistics

(Zimman 2020), my theoretical framework takes as central identities that are normally situated as peripheral in language and sexuality research, i.e., bisexual, trans, and non-white positionalities. Additionally, the contextual, co-constructed, and embodied nature of identity (Bucholtz & Hall 2004, 2005, 2016) is critical to my theorization of sexual identity. This kind of investigation requires a return to the critical origins of intersectionality theory, which stems from Black feminist epistemologies (Crenshaw 1989). In doing so, I better align how sexuality is operationalized and understood in experimental research in this area with broader theories of identity and power originating in interdisciplinary social theory.

Not only is the importance of this work a theoretically and methodologically necessary endeavor, it is also a matter of the ethics of representation and inclusion. My dissertation generates the largest body of linguistic work on bisexuality to date (though see Murphy 1997; Thorne 2013; Turai 2021; Wilkinson 2019), directly addressing the lack of bisexual representation in the field. It is also constructed to address applied concerns about bisexual well being. Bisexual people face disproportionately high rates of poverty, employment discrimination, intimate partner violence, and poor physical and mental health compared to lesbian, gay, and straight people, depending on the measure (Israel 2018; Project et al. 2014). Researchers in clinical psychology and advocates argue that the development of cultural competency and visibility are crucial for addressing these disparities (e.g. Israel 2018). An essential component of building cultural competency on part of researchers, healthcare professionals, and the general public is developing critical awareness of bisexuality. This dissertation develops critical awareness, and by extension cultural competency, by centering the experiences of bisexual people as essential for understanding sexuality (Zimman 2020). Inspired by my experiences as a bisexual person and the needs of my communities, my dissertation develops the language and analytic tools needed to name and dismantle the problematic stereotypes, institutions, and structures that marginalize and erase us.

This dissertation contains three content chapters in an article-style format. The first content chapter presents a statistical analysis of acoustic data that compares how cisgender bisexual women and men produce /s/–the sound associated with the "gay lisp"–relative to lesbian, gay, and straight cis people. Results suggest that there are indeed significant differences in how bisexual people say /s/, though the complexities of the results resist simple conclusions about what a bisexual voice sounds like. The chapter focuses on the lack of theorization of sexuality (and its intersection with gender) by attending to the assumptions of some commonly used methods. It also introduces the focus of the next chapter, namely how bisexuality intersects with other aspects of identity in speech production.

The second content chapter presents an analysis of the same acoustic data that uses machine learning, specifically random forests. My collaborator Chadi Ben Youssef and I add a number of factors into the /s/ production model, which in the previous chapter contained only gender and sexuality. The study finds that race and ethnicity play a particularly important role in predicting /s/ production among English speakers. Ultimately, the chapter focuses on two recommendations: first, a re-examination of quantitative research practice in light of intersectionality theory and second, the addition of random forests to the toolkit of sociolinguists (Tagliamonte & Baayen 2012).

The first and second content chapters focus on the relationship between bisexuality and language at the phonetic level and /s/ in particular. Both chapters are more concerned with a methodological and theoretical critique of the field, and recommendations to address those critiques, than they are in the idea of a bisexual-sounding voice. However, both studies find significant differences between how bisexual and non-bisexual people produce /s/, suggesting that a bisexual-sounding voice is potentially emerging.

While the first two content chapters focus on the lowest level of linguistic structure (phonetics), the final content chapter shifts its focus to the highest level of language structure: discourse. What it means to sound bisexual is yet unclear, but there are a number of prevalent stereotypes and ideologies related to bisexuality in other ways. A common misconception is that bisexual people are "confused" about their sexuality and will eventually identify as either lesbian/gay or straight. This stereotype is clearly rooted in monosexism. Consider the following Tumblr post:

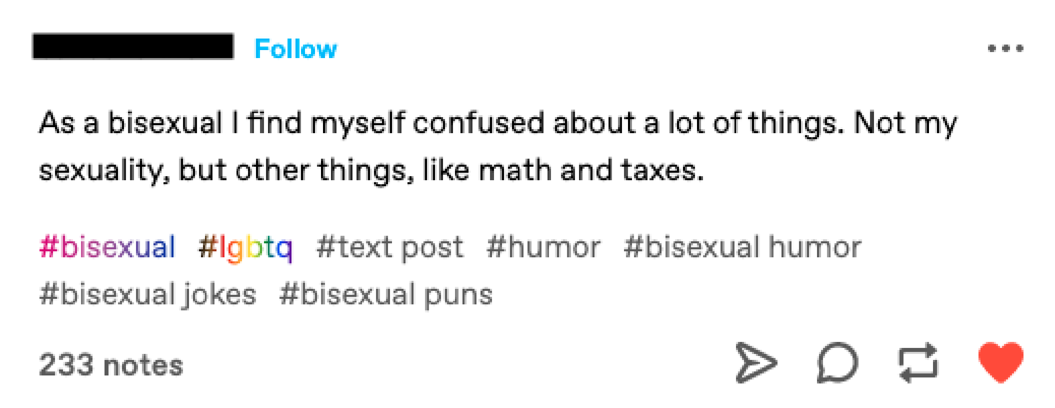

Figure 1.1: Tumblr post featuring a joke about bisexual people being confused

This poster uses humor as a discursive tactic to denaturalize monosexist ideologies of sexuality. By refocusing the target of bisexual confusion to math and taxes–both of which are widely complained about as being difficult to understand–the author frames bisexuality as easy to understand. In other words, being attracted to more than one gender comes naturally to this poster, but other realities of adult life, like math and taxes, do not.

As bisexual visibility in public life increases, and as the number of Americans who identify as bisexual increases (Jones 2022), discourses about bisexuality, too, are increasing. When people talk about bisexuality, what exactly are they saying? The third content chapter of this dissertation, a collaborative effort with Simon Todd, addresses this question through a statistical analysis of bisexual discourses on the social media platform Twitter. The results uncover some of the key concepts and ideologies important to discussions of bisexuality online. They also suggest that bisexuality itself is constructed and understood in multiple ways, one of which may present challenges for the ways sexuality is theorized in linguistics and related disciplines.

Finally, the dissertation concludes with a brief afterward. This section summarizes the key findings of each content chapter, and then discusses what each chapter brings to bear on the others. It also presents directions for expansion and future research, including brief previews of follow-up studies born out of this work.

1.1 Permissions and Attributions

1. The content of Chapter 2 is set to appear in a forthcoming issue of the *Journal of Language and Sexuality* (tentatively issue 13.1 in 2024). The content is reprinted here with permission from John Benjamins Publishing Company, Amsterdam/Philadephia.

2. The content of Chapter 3 is the result of a collaboration with Chadi Ben Youssef. Portions of chapter 3 have been published (under a CC BY 4.0 license) as an article in the *Proceedings of the Linguistic Society of America* (Willis & Ben Youssef 2023).

3. The content of Chapter 4 is the result of a collaboration with Simon Todd. A preliminary version of the paper was presented at the 61st Annual Meeting of the Association for Computational Linguistics (ACL) in Toronto, Canada.

Chapter 2

Bisexuality in experimental sociophonetics: Ideologies and implications

2.1 Introduction

One of the most vigorously pursued questions in sociophonetics concerns the idea that sexuality is indexed through the voice. This particular research focus attends to the widespread idea that there are systematic differences in speech production as a function of sexual orientation and that these differences are detectable by listeners. Empirical evidence has emerged to support such claims; studies of speech production report distinctive features of lesbian voices (e.g. Munson et al. 2006a; Pierrehumbert et al. 2004; Van Borsel et al. 2013) and gay voices (e.g. Crist 1997; Linville 1998; Pierrehumbert et al. 2004; Zimman 2013). From a perception standpoint, research suggests that listeners are able to judge speakers' sexual orientation at above-chance levels based on speech samples alone (e.g. Gaudio 1994; Linville 1998; Munson & Babel 2007). This research program has established a canonical literature on sexuality and the voice that is largely based on a dualistic contrast between lesbian/gay and straight speakers.

Despite calls to action as early as Gaudio (1994), bisexual speakers are effectively erased within sociophonetic studies of sexuality and the voice. The paradigms used to elicit sexuality judgments are heuristic of how this erasure happens. Listeners are typically prompted to evaluate a decontextualized voice as lesbian/gay- or straight-sounding in order to identify the acoustic correlates of sexuality. In the forced-choice paradigm (e.g. Smyth et al. 2003), listeners are limited to a binary decision between lesbian/gay or straight. This paradigm precludes the possibility of a voice sounding anything other than lesbian, gay, or straight and reinforces a binary understanding of sexuality.[1] In a Likert scale paradigm (e.g. Munson et al. 2006a), listeners are asked to evaluate voices on an odd-point scale in which one end represents lesbian/gay-sounding and the other straight-sounding. This paradigm too renders unintelligible voices that are not clearly lesbian/gay- or straight-sounding. It is not clear what, or rather who, the middle of the scale represents, nor is the implicit assumption that all participants interpret the scales in the same way supported with empirical justification. Additionally, this paradigm reinforces an understanding of bisexuality that positions it as an amalgam of lesbian/gayness and straightness. The uncritical way these methods are applied highlights the extent to which research in this area is couched in binary, monosexist ideologies of sexuality.[2]

How participant labels are assigned is also instructive. The conflation of bisexuality with lesbian and gay identities is widespread in sociophonetic studies of sexuality and the voice. Bisexual people are often uncritically grouped with lesbian and gay speakers in the few studies that do include us (e.g. Munson & Babel 2007; Munson et al. 2006b, 2006a).[3] The assumption that bisexual participants pattern with lesbian and gay speakers is based on intuitive understandings of "LGB". If you're not straight, you're everybody else. Alternatively, Smyth et al. (2003: 334) report that one of their groups of listeners "explicitly identified as gay males" whereas "the

[1]A binary approach may be appropriate depending on the research question. However, the choice to employ such a design should be measured and intentional.

[2]*Monosexism* refers to the ideology or assumption that people are, or should be, attracted to only one gender (Eisner 2016).

[3]However, see Barron-Lutzross (2015) who treats bisexual women as a separate category and Pierrehumbert et al. (2004) who conflate bisexual and lesbian women based on initial data exploration.

remainder formed a mixed group, by which we mean that we did not ask about their sexual orientation and we presume that most identified as heterosexual". If you're not gay, you're straight. Bisexual erasure is institutional and systemic, even in a research area that specifically studies sexuality.

My intention here is not to undermine the value of previous work, which has provided valuable insight into monosexual identities.[4] Rather, my goal is to highlight the extent to which binary, monosexist ideologies of sexuality permeate the theoretical underpinnings of many sociophonetic studies of sexuality and the voice. Until a more diverse range of sexual identities are included in the field, researchers cannot conclude that sexuality has acoustic correlates that go beyond cultural ideologies of lesbian/gay versus straight difference.

To that end, this chapter centers bisexual people in the analysis with the goal of complicating intuitive understandings of "LGB". I present an acoustic analysis of how bisexual English speakers produce the voiceless alveolar fricative /s/ relative to lesbian, gay, and straight speakers to test whether an a priori LGB grouping is empirically justified. Findings indicate that bisexual women and men do not consistently pattern with lesbian, gay, or straight speakers, or even with each other. These results problematize monosexist, cultural understandings of sexuality that often inform experimental research practice. Ultimately, I argue that bisexuality highlights the need for sociophonetic studies of sexuality and the voice to engage more deeply with the theorization of sexuality and its intersection with gender normativity.

2.2 Sexuality and the voice

A variety of potential acoustic correlates of sexuality have been examined in production and perception studies of English-speaking voices. The most thoroughly analyzed feature is pitch

[4]*Monosexual* refers to people who experience romantic, sexual, or affectional desire for only one gender (*LGBTQIA+ GLOSSARY* 2021).

or F0 (e.g. Gaudio 1994; Levon 2011; Smyth et al. 2003; Van Borsel et al. 2013). The results of these studies are inconsistent and occasionally contradictory. Gaudio (1994) reports that pitch range and variability do not reliably predict whether a man sounds gay. Likewise, Linville (1998) finds that differences in mean F0 are not reliably correlated with sounding gay, nor do gay men produce significantly different mean F0 than straight men. Contrary to the stereotype that gay men speak with higher pitch, Zimman (2013) reports that cisgender gay men speak with significantly lower mean F0 relative to straight cisgender men and that transgender men of varying sexual orientations fall in between these two groups. As for lesbian women, Waksler (2001) reports no significant differences in pitch or pitch variability between lesbian and straight women, whereas Van Borsel et al. (2013) find that lesbian women produce significantly lower mean F0 and fewer pitch fluctuations compared to straight women.

The acoustic properties of English vowels have also been frequent objects of interest in sociophonetic studies of sexuality (e.g. Jacobs et al. 2000; Linville 1998; Munson et al. 2006a; Pierrehumbert et al. 2004; Rendall et al. 2008; Zimman 2010, 2013). A perception study by Munson (2007) reports a positive correlation between relatively high-frequency F1 values and sounding gay. However, studies by Linville (1998), Jacobs et al. (2000), and Zimman (2010, 2013) do not corroborate these findings. As for production studies, Pierrehumbert et al. (2004) report significant differences in vowel quality between lesbian/bisexual and straight women for the back vowels /ɑ/ (F1 and F2) and /u/ (F1 only).[5] They also find that the gay and straight men in their study differ in terms of F1 and F2 for the /i/, /ɑ/, and /æ/ vowels.

Other than pitch and vowels, /s/ is perhaps the most thoroughly studied acoustic correlate of sexuality in English. Many sociophonetic studies of sexuality and /s/ focus on center of gravity, the first spectral moment.[6] Also known as the centroid, center of gravity (COG) is the

[5]Pierrehumbert et al. (2004) conflate bisexual and lesbian women based on exploratory analyses.

[6]Spectra of fricative noise can be analyzed in terms of four spectral moments: center of gravity, standard deviation, skew, and kurtosis. The standard deviation (also known as spread) refers to the variance and range of energy in the spectrum. Kurtosis refers to the peakedness of the spectrum (Forrest et al. 1988; Jongman et al. 2000; Thomas 2017). To my knowledge, correlations between sexuality and spread or kurtosis have not been found in any

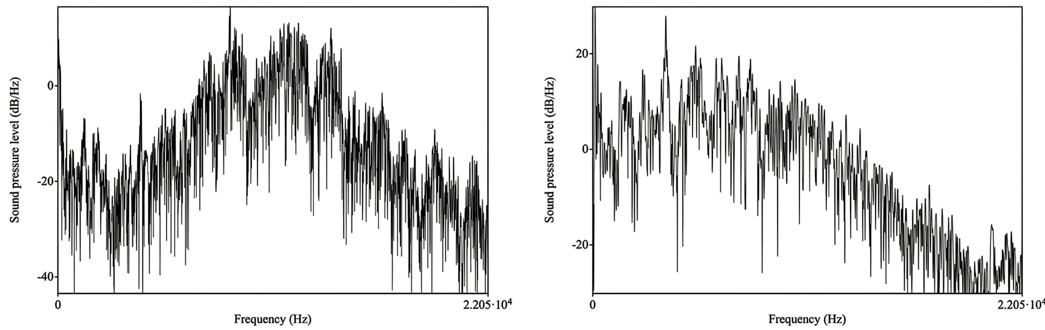

Figure 2.1: A high COG /s/ produced by a bisexual woman (left) and a low COG /s/ produced by a bisexual man (right)

weighted mean of frequencies in the spectrum. This acoustic measure is correlated with the frontedness of /s/ in articulatory terms, such that higher center of gravity correlates with fronted /s/ (Figure 2.1)(Shadle 1990, 1991). Comparing across genders, studies indicate that women produce a more fronted /s/ than men (e.g. Fuchs & Toda 2010; Hazenberg 2015; Munson et al. 2006a). English-speaking adult women's center of gravity typically ranges from 6,400 to 8,500 Hz while men's range from 4,000 to 7,000 Hz (Zimman 2017). Production studies suggest that gay men produce fronter /s/ tokens than straight men (e.g. Linville 1998; Munson et al. 2006a; Podesva & Hofwegen 2014). Likewise, perception studies of men's speech suggest that fronter /s/ tokens are more gay-sounding (e.g. Campbell-Kibler 2011; Linville 1998; Munson 2007; Rogers & Smyth 2003; Zimman 2013).[7] Studies of women's speech have been less conclusive. Barron-Lutzross (2015) reports no correlation between /s/ production and sexual orientation in women. In contrast, Hazenberg (2015) finds that straight women produce fronter /s/ tokens than lesbian women. Similarly, Munson et al. (2006a) report that straight women produce /s/ with a higher center of gravity than lesbian/bisexual women.[8] Overall, previous research finds a consistent correlation between /s/-fronting, as measured by center of gravity, and sexual

previous work on speech production or perception and as such these measures are not discussed at length here.

[7] Some studies use the term (average) peak frequency rather than center of gravity (e.g. Linville 1998; Rogers & Smyth 2003).

[8] Munson et al. (2006a) do not provide theoretical or empirical justification for conflating bisexual and lesbian women in their study.

orientation in men, though less so in women.

Another measure of /s/ considered in previous studies of English is skew, the third spectral moment. Skew or skewness is a measure of spectral tilt, or how fast energy falls off as frequency increases (Thomas 2017). Positive skew indicates prominence in lower frequency ranges, whereas negative skew indicates prominence in higher frequency ranges (Figure 2.2). A skew of zero indicates a symmetrical distribution around the mean (Jongman et al. 2000). Munson et al. (2006a) report that English-speaking women produce /s/ with a more negative skew than men. Likewise, the authors find that gay men produce /s/ with a more negative skew than straight men, though lesbian/bisexual and straight women do not differ in this respect. They also report that men who produced /s/ with a more negative skew were more likely to be rated as gay-sounding.

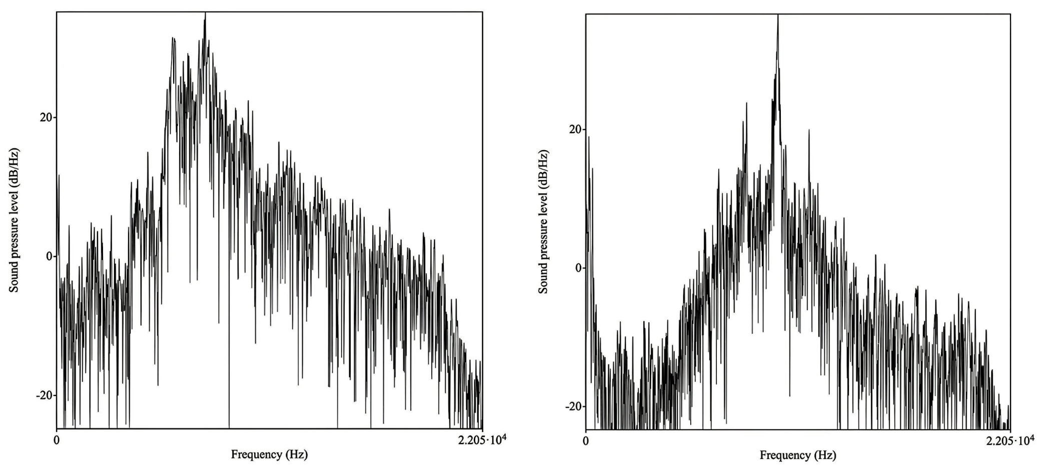

Figure 2.2: A negatively skewed /s/ (left) and a positively skewed /s/ (right) produced by two bisexual women

Duration is another measure of /s/ that has received considerable attention in studies of sexuality and the voice in English. Crist (1997) finds that /s/ is lengthened in stereotypical gay men's speech, but only in certain phonological contexts. Linville (1998) reports that gay men tend to produce /s/ with longer duration than the straight men in her sample. She also reports that /s/ duration correlates with both perceived and reported sexual orientation. A perceptual study by Rogers et al. (2000) finds that speakers rated as gayer-sounding had significantly longer normalized mean /s/ durations than those rated as straighter-sounding. However, Levon (2007) reports that digitally manipulating the duration of /s/ tokens is not enough to significantly alter listeners' perceptions of sexuality. Levon does not claim that /s/ duration (and pitch range, which he also examines) plays no part in listeners' assessments of speakers' sexuality. Rather, Levon suggests that /s/ duration may need to be combined with other features in order to alter the perception of sexuality.

The selection of these particular variables is, at least in part, inspired by ideologies of sounding lesbian or gay. The stereotype that lesbian women speak with markedly low pitch and less dynamic intonation, and conversely that gay men speak with markedly high pitch and more dynamic intonation, have motivated the numerous studies of pitch and sexuality. Likewise, the focus on /s/ is undoubtedly motivated by the widespread "gay lisp" stereotype, which claims that gay men pronounce /s/ in a way that sounds "lispy" relative to straight men. However, unlike lesbian- and gay-sounding voices, there are no salient stereotypes or ideologies about what it means to sound bisexual in American English-speaking contexts. That being the case, I choose to examine /s/ not because of an obvious ideological link between /s/ production and sounding bisexual, but rather because investigations of /s/ are relatively consistent across studies of sexuality and the voice (Zimman 2013). Given that the purpose of this study is to question whether bisexual speakers pattern similarly to lesbian and gay speakers as is typically assumed, a variable that is consistently correlated with sexuality such as /s/ is a reasonable place to begin. I specifically focus on three measures of /s/ production in my analysis: center of gravity (frontedness), skew, and duration. The findings reported in this chapter suggest that /s/ is as good of a place to start as any, as they demonstrate that the conflation of bisexual, lesbian, and gay speakers is perhaps not as well motivated as intuition might suggest.

2.3 Methodology

2.3.1 Participants

Participants were 27 graduate and undergraduate students at a California university. There were 5 speakers in each GENDER*SEXUALITY group, except for bisexual men; only two eligible bisexual men participated (Table 2.1).[9] Participants were recruited using snowball sampling and through flyers distributed electronically and posted in various physical locations on the university campus. I had met some of the participants prior to the study through my involvement in the graduate student LGBTQIA+ community, but most were strangers to me.

All participants identified as cisgender, were native English speakers, and were between 18 and 30 years old at the time of the study. Exploratory t-tests indicated that gay men were significantly older than the other groups of men in this study ($t = 1.79$, $p = 0.03$), but there was no significant difference between the bisexual women and men and the other groups of women and men. Most speakers identified as white ($n = 15$), but Latinx ($n = 5$), Asian ($n = 3$), and mixed race ($n = 4$) speakers also participated. The vast majority of participants identified one or more locations in the US as their place(s) of origin ($n = 23$). Most of the participants ($n = 17$) named California as (one of) their place(s) of origin, making the state the most frequently identified place in the dataset. Two participants reported locations in the US and Europe as their places of origin and one participant identified India as their place of origin.

2.3.2 Stimuli

Participants were recorded reading the Rainbow Passage (Fairbanks 1960), a phonetically balanced scientific passage about rainbows, as well as a list of 240 phonetically balanced sentences (Rothauser 1969). Speakers recorded the passage first and the sentences second. They were

[9] A third bisexual man who participated was excluded because he reported that he had been diagnosed with a speaking problem that affected his production of /s/.

Group	n	Mean age	Age range	Race/ethnicity	Origin
Bisexual women	5	21	18-26	1 Latinx/PI[10], 4 white	US
Bisexual men	2	19	18-19	1 Latinx, 1 white	US
Lesbian women	5	21	19-26	1 Latinx/white, 3 white, 1 white/Jewish	3 US 2 US & Europe
Gay men	5	25	18-30	1 Asian, 1 Latinx/white, 3 white	4 US 1 India
Straight women	5	19	18-20	2 Asian, 1 Latinx, 2 white	4 US 1 declined
Straight men	5	19	18-20	3 Latinx, 2 white	US

Table 2.1: Participant information

instructed to read the passage and all sentences twice only and to say the stimuli as naturally as possible. The analysis in this chapter draws only from participants' first reading of the Rainbow Passage. Restricting my interactions with (most) participants to this specific context limits the extent to which I can relate my findings to participants' everyday language use, but the method does have its advantages. Analyzing a passage recorded in a laboratory environment provides a relatively straightforward means for measuring differences in speech production, as it controls for differences in topic, context, or affect that are common in interactional contexts (Smyth et al. 2003; Zimman 2017: 1000).

2.3.3 Procedure

Participants were recorded in a private, sound-attenuated booth using Audacity (Audacity Team 2019). Recordings were sampled at a 44.1 kHz rate with 16-bit quantization using either a Blue Snowball iCE USB Condenser microphone, a Blue Yeti USB microphone, or an AKG C 3000 B microphone. I impressionistically noticed no significant difference in audio quality between the three microphones.

During the informed consent process, speakers were told that they were participating in a two-part study on LGBQ and allied voices in California. They were told that they would be further debriefed on how their data would be used after recording was completed and that they would have the opportunity to withdraw consent and destroy their data at that time. After obtaining their initial consent, participants filled out a pre-test survey asking for personal information such as age, gender, sexuality, race/ethnicity, languages spoken, and so forth. The way participants are categorized in the analyses is based on their self-assigned identity labels as indicated in the pre-test survey. When they finished recording, speakers filled out a post-test survey asking them about their gender stereotypicality and romantic partner preferences. Finally, participants were debriefed on how I intended to use their recordings and were reminded of their right to withdraw consent. None opted to do so.

2.3.4 Analysis

Participants' first readings of the Rainbow Passage (Fairbanks 1960) were transcribed in Praat (Boersma 2011) and word-initial tokens of /s/ were extracted using a Praat script (Zimman 2018). The analysis was restricted to word-initial /s/ tokens in order to reduce the likelihood of intervocalic voicing and to control for word-final lengthening effects. The Praat script generated center of gravity, skew, kurtosis, and standard deviation measurements for each /s/ token. Tokens in the word *strike* and the phrase *friends say* were discarded due to their phonological context. Specifically, /s/ retraction in /stɹ/ clusters in words such as *strike* is well documented (e.g. Shapiro 1995) and there is a tendency for the final /z/ in *friends* to blend with the initial /s/ in *say*, making the token difficult to segment consistently. Disfluencies and tokens judged to contain periodicity were also discarded, leaving a mean of 13.78 tokens per person and 372 tokens total. I also manually checked measurements that were atypical relative to previous findings, i.e. tokens in which the center of gravity was lower than 4 kHz or higher than 8.5 kHz and tokens in which the skew was less than -2 or more than 2. For tokens flagged as atypical for a given measurement, I also checked the other measurements. Duration was hand measured for every token as a part of the process of vetting tokens for periodicity.

2.3.5 Statistical methodology

Once the data was processed, I fit a linear mixed-effects regression model for each dependent variable: center of gravity, skew, and duration. Models were fit using the lmer function from the lmerTest package in R (R Core Team 2022). The maximal model included GENDER, SEXUALITY, and their interaction. SPEAKER and WORD were also included as random effects with varying intercepts in each of the models. For model selection, I used the drop1 function from the lme4 package to select for the model with the best fit. The anova function from base R was used to determine the significance of the main effects. Finally, post-hoc pairwise comparisons were calculated using the relevel function from base R in order to capture the differences between all six groups. The Bonferroni method was used to compensate for evaluating the models multiple times, up to five times in the case of the center of gravity model.

I report both Bonferroni-corrected significance values and uncorrected values in this chapter for two reasons. The first is a matter of convention: adjustments like the Bonferroni correction are not used consistently in experimental studies of sexuality and the voice, so it is not unusual to report unadjusted results. The second reason is a matter of the correction itself. The Bonferroni method is known for being intolerant of Type I errors; in other words, it can fail to detect true differences and is potentially overly conservative (Lee & Lee 2018: 237). Given these two points, I report both sets of values, with the caveat that the corrected values are held to a more rigorous standard.

2.4 Statistical results

2.4.1 Center of gravity model

Center of gravity (COG) measurements (n = 372) were subjected to a linear mixed-effects regression model. The model selection process indicated that the maximal model with GENDER, SEXUALITY and their interaction was the best fit. GENDER was a significant main effect, such that women generally produced /s/ with a higher COG than men ($\chi^2(1) = 5.9$, $p = 0.01$). The interaction GENDER*SEXUALITY was also a significant main effect ($\chi^2(1) = 9.5$, $p < 0.01$) (Figure 2.3). SEXUALITY was not a significant main effect on its own, but could not be removed from the model due to its participation in a significant interaction. The fixed effects explained about 29% of the variance ($R^2m = 0.29$) and the random effects explained about 38% of the variance ($R^2c = 0.68$). The random effect SPEAKER explained 6.6 times more variance than the random effect WORD.

Fixed effects	Estimate	Standard error	df	t-value	Pr(>|t|)
Intercept (bisexual women)	8109.01	362.54	23.26	22.37	<2e-16***
Gender:M (bisexual men)	-1578.87	658.73	20.91	-2.40	0.026*
Sexuality:Lesbian/gay (lesbian women)	-1498.13	498.14	20.94	-3.01	0.007**
Sexuality:Straight (straight women)	79.11	498.29	20.97	0.16	0.875
Gender:M, Sexuality:Lesbian/gay (gay men)	1897/97	825.87	20.92	2.30	0.032*
Gender:M, Sexuality:Straight (straight men)	-12.85	825.97	20.93	-0.02	0.988

Table 2.2: COG model coefficients table

Among the women in this study, straight women produced /s/ with the highest estimated COG (8288 Hz), followed by bisexual women (8109 Hz), and then lesbian women (6711 Hz) (Table 2.2, Figure 2.3). The straight and bisexual women in this sample produced COG estimates that fell on the higher end of the expected range for women (6.4-8.5 kHz), whereas lesbian women's COG estimates were on the lower end (Zimman 2017).

As for the men in the study, gay men produced the highest estimated COG (6930 Hz), followed by straight men (6696 Hz), and then bisexual men (6630 Hz) (Table 2.2, Figure 2.3). The estimate for gay men is on the higher end of expected COG values for men, but is still within the 4-7 kHz range that is considered typical (Zimman 2017). Bisexual men's and straight men's COG estimates are well within the expected range.

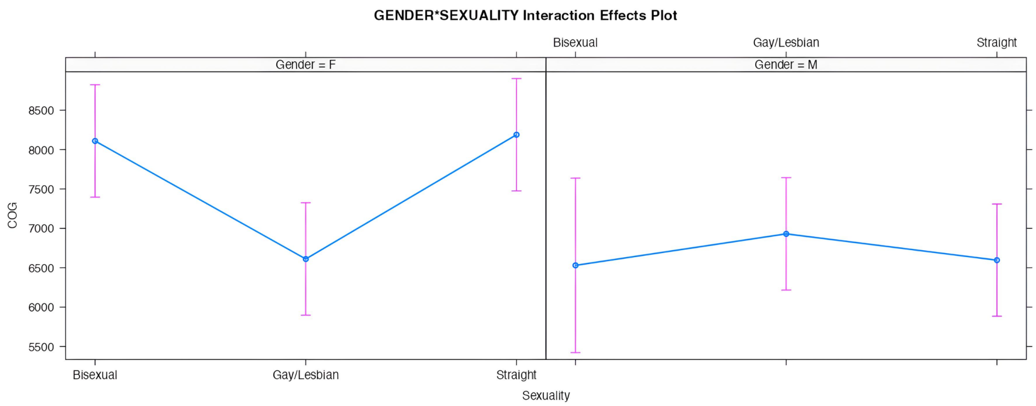

Figure 2.3: GENDER*SEXUALITY interaction effects plot for COG model

In the initial evaluation of the model (Table 2.2), a significant difference was found between bisexual women and lesbian women, such that bisexual women produced /s/ with a higher COG than lesbian women ($p < 0.01$). A significant difference between bisexual women and bisexual men was also found, such that bisexual women produced /s/ with a higher COG than bisexual men ($p < 0.05$). Similarly, a significant difference between bisexual women and gay men was found, such that bisexual women produced /s/ with a higher COG than gay men ($p < 0.05$). There were no significant differences between bisexual women and straight women or straight men.

2.4.2 Center of gravity post-hoc pairwise comparisons

The model was releveled four times in order to examine the full set of pairwise comparisons for each group (five evaluations total). Comparing within gender but across sexuality (Table 2.3), significant differences between bisexual women and lesbian women ($p < 0.01$) as well as lesbian women and straight women ($p < 0.01$) were maintained before and after the correction. No significant differences in COG were found between bisexual women and straight women nor between any of the groups of men.

Gender	**Sexuality**		**p-value**	**Unadjusted (p < 0.05)**	**Bonferroni (p < 0.01)**
Women	Bisexual	Lesbian	0.007	**	*
	Bisexual	Straight	0.875		
	Lesbian	Straight	0.005	**	*
Men	Bisexual	Gay	0.550		
	Bisexual	Straight	0.921		
	Gay	Straight	0.510		

Table 2.3: COG within-gender cross-sexuality pairwise comparisons

Comparing within sexuality but across gender (Table 2.4), the significant difference between bisexual women and men found in the initial evaluation of the model was lost after the correction. However, a significant difference between straight women and men maintained before and after the correction, such that straight women produced /s/ with a higher COG compared to straight men. This result is not surprising, given findings in previous literature on COG and gender difference (e.g. Fuchs & Toda 2010; Hazenberg 2015; Munson et al. 2006a).

Group	**p-value**	**Unadjusted (p < 0.05)**	**Bonferroni (p < 0.01)**
Bisexual	0.026	*	
Lesbian/Gay	0.529		
Straight	0.004	**	*

Table 2.4: COG within-sexuality cross-gender pairwise comparisons

2.4.3 Skew model

Skew measurements (n = 372) were subjected to a linear mixed-effects regression model. The model selection process indicated that neither gender, sexuality, nor their interaction contributed to the overall fit of the model. The random effects speaker and word accounted for about 58% of the variance in the data, with speaker explaining 43 times more variance than word. These results are surprising, given that previous studies such as the work by Munson et al. (2006b, 2006a) have found relatively strong evidence for differences in skew, especially between gay and straight men. Measures of central tendency and dispersion, however, suggest that the gay men in this sample produced /s/ with smaller skew values than straight men, which aligns with previous research (Munson et al. 2006b, 2006a) (Figure 2.4).

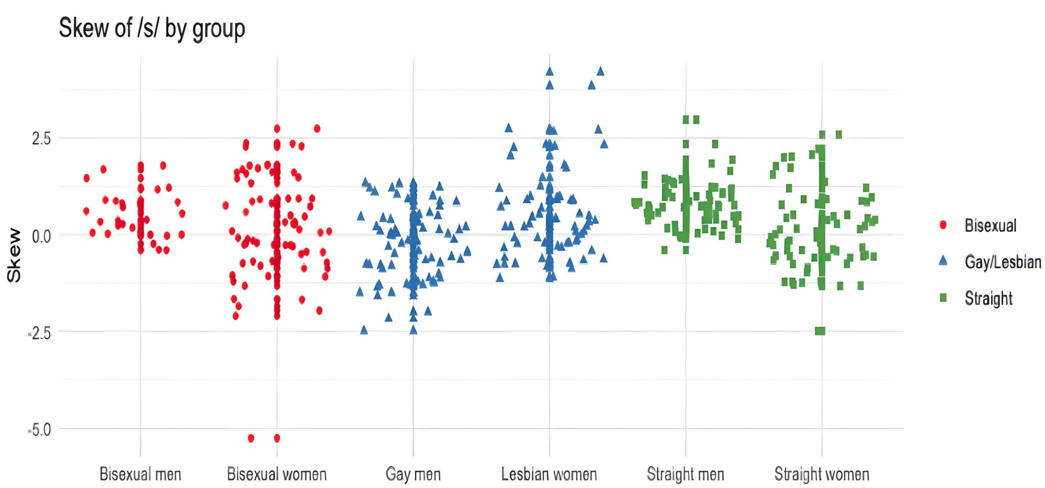

Figure 2.4: Distribution of skew by group

There are a few plausible explanations for these unexpected results. The first is concerned with methodology. The data used in sociophonetic studies of the voice are elicited in a variety of ways. Many studies—including the study at hand—extract data from readings of passages like the Rainbow Passage (e.g. Zimman 2013), whereas others analyze tokens extracted from decontextualized words (e.g. Munson et al. 2006a). Moreover, these studies make use of a wide variety of statistical modeling approaches, including MANOVA (e.g. Munson et al. 2006a; Pierrehumbert et al. 2004), ANOVA (e.g. Munson et al. 2006b; Zimman 2013), and regression models (e.g. Campbell-Kibler 2011; Zimman 2017). Perhaps, then, inconsistent results between studies are artifacts of different methodological choices. Another possibility is that different speakers use distinct combinations of features to index sexuality and this particular set of speakers happens to not employ skew (Zimman 2013). Alternatively, this group of speakers may index sexuality through skew but the sample size is too small to detect an effect. Yet another explanation is couched in how much information is packed into the random effect SPEAKER. This variable accounts for all of the ways participants may differ other than gender and sexuality: race and ethnic identity, age, regional affiliation, and languages spoken to name a few. Considering that SPEAKER is the vastly more important of the two random effects, it is possible that the influence of some other axis of identity that is embedded within SPEAKER is key. These results, considered alongside work linking /s/ variation with gender, race, and place (e.g. Calder & King 2020; Campbell-Kibler 2011; Mendoza in prep; Pharao et al. 2014; Podesva & Hofwegen 2014), point to the potential of an analysis that unpacks speaker for future research (see Chapter 3).

2.4.4 Duration model

Preliminary analyses of duration as a function of word length (in syllables), position of the word within the sentence, stress, and syllable type (e.g. open or closed) were conducted to identify outliers and confounding factors of /s/ duration. These analyses indicated that the token extracted from *superimposition* should be removed from the sample. The initial /s/ in *superimposition* was the sole token in the sample not contained in a syllable with primary stress. Moreover, *superimposition* is the only word in the sample with more than three syllables. Both stress and word length appeared to strongly affect the duration of the initial /s/ in *superimposition*, such that these tokens were outliers relative to the other tokens in the sample. As such, tokens from *superimposition* were removed, leaving 349 tokens total ($M_{speaker}$ = 12.9) for statistical analysis.

Duration measurements (n = 349) were subjected to a linear mixed-effects regression model. The model selection process indicated that a model with just SEXUALITY was the best fit and that SEXUALITY was a significant main effect ($\chi^2(1) = 29.9$, $p < 0.001$). The main effect GENDER and the GENDER*SEXUALITY interaction were not significant and thus were excluded from the final model (Table 2.5).

Fixed effects	**Estimate**	**Standard error**	**df**	**t-value**	**Pr(>|t|)**
Intercept (bisexual)	0.052	0.005	28.92	10.83	<1.09e-11***
Sexuality:Lesbian/Gay	0.013	0.005	23.95	2.78	0.0106*
Sexuality:Straight	0.035	0.005	23.95	7.35	1.40e-07***

Table 2.5: Duration model coefficients table

Bisexual speakers produced /s/ with the shortest estimated duration (52 ms), followed by lesbian and gay speakers (65 ms), with straight speakers producing the longest /s/ tokens (87 ms) (Table 2.5, Figure 2.5). The fixed effects accounted for about 33% of the variance ($R^2m = 0.33$) and the random effects speaker and word accounted for about 30% of the variance ($R^2c = 0.62$). word accounted for 1.4 times more variance than speaker.

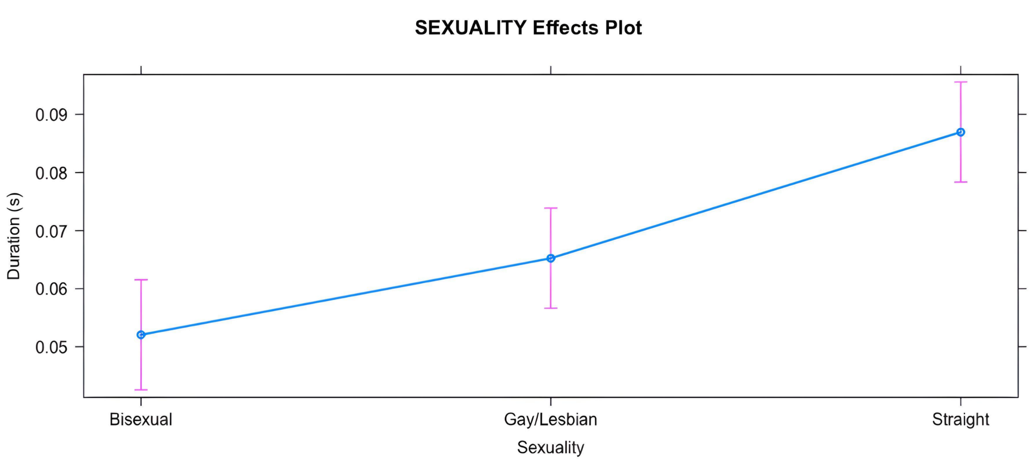

Figure 2.5: SEXUALITY effects for duration model

In addition to the exploratory analyses described previously, I also compared a model in which TOKEN was included as a random effect instead of WORD to investigate the influence of token position within the overall passage as a possible confound. TOKEN was represented as a factor with two more levels than WORD (14 levels versus 12 levels), as the words *since* and *size* were repeated twice in the passage. Model comparison with the anova function from base R indicated no significant difference between the two models.

2.4.5 Duration post-hoc pairwise comparisons

The model was releveled one time in order to examine the full set of pairwise comparisons for each group (two evaluations total). Significant differences were found between all three groups: bisexual speakers and lesbian/gay speakers ($p < 0.05$), bisexual speakers and straight speakers ($p < 0.001$), as well as lesbian/gay speakers and straight speakers ($p < 0.001$) (Table 2.6). All differences maintained significance before and after the Bonferroni correction.

Comparison	**p-value**	**Unadjusted ($p < 0.05$)**	**Bonferroni ($p < 0.01$)**
Bisexual:Lesbian/Gay	0.0106	*	*
Bisexual:Straight	1.40e-07	***	***
Lesbian/Gay:Straight	3.8e-05	***	***

Table 2.6: Duration post-hoc pairwise comparisons

2.4.6 Summary of statistical results

Before discussing the post-test survey, I briefly summarize the key statistical results (Table 2.7). GENDER and the GENDER*SEXUALITY interaction were significant main effects in the center of gravity model. Bisexual women produced /s/ with a significantly higher center of gravity relative to lesbian women. Additionally, bisexual women produced /s/ with higher center of gravity compared to bisexual men, but this finding was no longer significant after the Bonferroni correction was applied. No differences were found between bisexual women and straight women or among any of the groups of men. The skew model selection process indicated that neither GENDER, SEXUALITY, nor their interaction contributed to the overall fit of the model. The random effect speaker explained most of the variation. As for duration, SEXUALITY emerged as a significant main effect, such that bisexual speakers produced shorter /s/ tokens than both lesbian and gay speakers and straight speakers. In short, bisexual speakers do not consistently pattern with either lesbian, gay or straight speakers, or even with each other in the statistical models.

Center of gravity	**Skew**	**Duration**
Significant difference between bisexual and lesbian women before and after correction	No significant main effects	Significant differences between all three groups before and after correction
No difference between bisexual and straight women	Most variation was explained by random effect SPEAKER	Bisexual speakers' duration was the shortest, followed by lesbian/gay and then straight speakers
Significant difference between bisexual women and men before correction only		
No difference between the groups of men		

Table 2.7: Summary of key statistical results

The duration findings are particularly noteworthy in that the statistical model had the strongest evidence with respect to the Chi-squared values of the main effects and the p-values of the pairwise comparisons, yet the results contradict previous studies. Research on sexuality and duration in English has generally argued that increased duration is associated with sounding gay (e.g. Crist 1997; Linville 1998; Rogers et al. 2000, but see Levon 2007 for an exception). However, the current study finds that lesbian and gay speakers produce /s/ with shorter duration than straight speakers, and that bisexual speakers produce /s/ with the shortest duration overall.

Additional evidence is needed to explain why these results are inconsistent with earlier research and to investigate what exactly is being indexed by duration.

2.5 Qualitative analysis of post-test surveys

The post-test survey asked participants about the gender(s) to which they experience attraction as well as how they perceive themselves relative to gender stereotypes. I focus on the gender stereotypicality questions here. Participants answered the questions "how stereotypically feminine do you consider yourself?" and "how stereotypically masculine do you consider yourself?" by selecting a number on two Likert scales in which 1 indicated "not at all stereotypically feminine/masculine" and 7 indicated "very stereotypically feminine/masculine", depending on the question. All but one participant (a lesbian woman) provided ratings on both scales. I analyze these ratings qualitatively, rather than quantitatively, because the statistical models for the three measures of /s/ did not converge when the two ratings (each a seven-level ordinal variable) were added to the maximum effects structure.

Most of the straight participants rated themselves towards the more extreme ends of the scales, in line with normative expectations for their gender (Table 2.8). All of the straight women rated themselves as more feminine than masculine. Likewise, all of the straight men rated themselves as more masculine than feminine. Lesbian women and gay men, however, rated themselves more towards the middle of the scale for both measures and did not systematically rate themselves as more feminine or masculine by group.

The two bisexual men who participated provided exactly opposite ratings: one bisexual man rated himself as more masculine (5) than feminine (3), while the other bisexual man rated himself as more feminine (5) than masculine (3). Most of the bisexual women rated themselves as more feminine (5-6) than masculine (1-2). However, one bisexual woman rated herself as more masculine (5) than feminine (3).

Group	Femininity Rating		Masculinity Rating	
	mean	range	mean	range
Bisexual women	5	3-6	3	1-5
Bisexual men	4	3-5	4	3-5
Lesbian women	4	3-5	3	2-4
Gay men	3	2-5	4	3-5
Straight women	6	5-7	2	1-3
Straight men	2	1-3	6	5-7

Table 2.8: Gender stereotypicality ratings by group

When designing this study, I anticipated that including separate scales for femininfancy pagestyleity and masculinity would make interpreting the results of the post-test survey more straightforward. I did not find that to be the case. One of the fundamental issues with Likert scales is that, without additional context such as post-task interviews about the scales themselves, it is difficult to determine how participants interpret these scales or whether participants are interpreting the scales in the same way. What is stereotypically feminine or masculine may vary from person to person based on their background, life experience, or values. After the experiment had concluded, one Asian American participant expressed that she had difficulty answering these questions because “there are different stereotypes for Western and Eastern culture.” How the middle of the scale is interpreted may also vary by individual, with possible interpretations ranging from ambiguously feminine/masculine to expressing a lack of confidence in their answer. Participants may also orient differently to the task itself. For instance, the lesbian woman’s refusal to answer these particular questions indicates an orientation to the task that is distinct from the other participants. I return to this issue in the discussion section, but for now I provide a plausible interpretation of this data.

I tentatively suggest that the bisexual men who participated evaluated their gender stereotypicality in a way that is more similar to the gay men in this study. I base this conclusion on the following: (1) neither the bisexual men nor the gay men systematically rated themselves as more feminine or more masculine and (2) the bisexual and gay men’s ratings were both on the less extreme ends of the scales. In contrast, the majority of bisexual women evaluated their gender stereotypicality in a way that seems more similar to the straight women in this study. They consistently rated themselves as more feminine than masculine and their ratings were on the more extreme ends of the scales. Lesbian women, on the other hand, overall rated themselves as more feminine than masculine, but their ratings were on the less extreme ends of the scales. The one bisexual woman who judged herself as more masculine than feminine was just one of two women in the entire sample to do so (the other being a lesbian woman).

The gender stereotypicality ratings complement the results of the statistical analyses in a few key ways. Although most of the bisexual women rated themselves similarly to straight women and there are parallels between bisexual men's and gay men's ratings, bisexual women and men did not definitively pattern with lesbian, gay, or straight participants. Moreover, bisexual participants' gender stereotypicality ratings were not consistent among themselves. Most bisexual women rated themselves as more feminine than masculine, but one did not; the two bisexual men provided perfectly mirrored ratings, such that one identified as more feminine than masculine and the other identified as more masculine than feminine. Similar to the results reported in the statistical models, consistent patterns do not arise between bisexual participants and non-bisexual participants or between bisexual women and bisexual men with respect to gender stereotypicality.

2.6 Discussion

Research on lesbian- and gay-sounding speech has historically been driven, at least in part, by a desire to empirically investigate stereotypes such as lesbian women "talk like men" or sound markedly masculine and gay men "talk like women" or sound markedly feminine. Ideological connections between sexuality and gender normativity are embedded within these stereotypes about lesbian and gay speech. Namely, these claims assert that lesbian- and gay-sounding speech is characterized by an adoption of the speech norms for the ideologically "opposite" gender, thereby positioning lesbian/gay speech as marked and non-normative. In turn, these claims position the way straight people speak as the unmarked norm. Although many studies complicate these claims, arguing that lesbian- and gay-sounding speech is far more complex than a straightforward adoption of ideologically opposing gender norms (see Munson & Babel 2007 for a summary of this research), it is often the case that lesbian and gay speech is interpreted as deviance from the straight norm. Even in studies that question the simplistic way lesbian and gay speech is characterized by stereotypes, lesbian and gay speakers are often implicitly positioned as non-normative relative to straight speakers.

When bisexual speakers are included in such studies, we are typically grouped with lesbian and gay speakers and are therefore similarly positioned as non-normative vis-à-vis the ideological connection between gender typicality, sexuality, and speech production. However, the results of the post-test survey suggest that this non-normative positioning does not reflect how most of the bisexual speakers in the present study perceive their gender typicality. Most of the bisexual women positioned themselves as stereotypically feminine and one of the two bisexual men positioned himself as stereotypically masculine. The way the majority of bisexual participants in this study view their gender typicality does not align with the assumptions about gender normativity embedded in an "LGB vs. straight" framework.

The non-normative positioning of bisexual people that is common practice in language and sexuality studies also does not necessarily reflect how bisexuality is perceived in everyday life. In hegemonic discourses of bisexuality, bisexual women are often positioned as performing their sexuality for the male gaze (Alarie & Gaudet 2013; Hartman 2013; Hertlein et al. 2016). Bisexuality in women is portrayed as an appeal to heterosexuality, therefore positioning bisexual women as relatively in-line with normative femininity. Perhaps the bisexual women in my study who rated themselves as more feminine than masculine identify as bisexual—as opposed to queer, pansexual, or even lesbians who have relationships with men—in part for this reason.[11]

[11] I include *queer* in this list of potential identity labels based on my experiences with individuals who identify

For people who see themselves as stereotypically feminine and experience attraction to multiple genders, a label like bisexual may feel more appropriate compared to other possibilities, given the ideologies about bisexuality in women.

Identifying as a bisexual man engenders risk from multiple fronts. Bisexual men are often positioned as in denial or transitioning to identifying as gay (Alarie & Gaudet 2013; Brewster & Moradi 2010; Hertlein et al. 2016). As a result, they are denied access to hegemonic masculinity. They may also experience ostracization from both LGBQ+ and straight communities due to biphobia and homophobia. These factors may be related to why I experienced difficulty in recruiting bisexual men and why studies that do include bisexual participants usually only include women (e.g. Munson et al. 2006b, 2006a; Pierrehumbert et al. 2004). To be clear, I am not arguing that ideologies or stereotypes dictate how individuals identify. However, bisexual *intersubjectivity* (cf. Bucholtz & Hall 2004) is relational, such that identifying as bisexual is mediated both by the self and others. Bisexual people may experience different tensions between how they self-identify and how they are positioned by others based on their gender (miles-hercules 2022).

Tensions between *identity* and *identifications* (miles-hercules 2022) may also arise in the context of relationships. Bisexual people are often misunderstood as straight when they are in different-gender relationships or express different-gender attraction. Those same individuals may be perceived as lesbian or gay when they are in same-gender relationships or express same-gender attraction. Of course, there are bisexual people who are perceived as non-normative based on their own gender presentation or other indexes of sexuality, regardless of romance or attraction. These situations, heuristic rather than exhaustive, illustrate that the fluid, co-constructed, and context-dependent nature of sexuality is particularly salient for bisexual people. The way bisexuality is understood on the ground is not adequately captured in experimental

themselves as queer based on their attraction to multiple genders. However, I refrain from providing a specific definition of queer here because I lack information about how my participants may use or interpret this term.

language and sexuality studies, such that more theoretical attention to the gender-sexuality interface is needed within research in this area.

2.7 Conclusion

In this chapter, I presented an acoustic analysis of bisexual speakers' productions of /s/ relative to lesbian, gay, and straight speakers. The results suggest that bisexual women and men produce /s/ in a way that is distinct from lesbian, gay, and straight speakers, and even from each other in this specific sample. Therefore, the research question that defines this study—whether an a priori LGB grouping is empirically justified—has an answer: no.

Many questions still remain. The study finds differences in /s/ production, but what does this variation mean? Previous work on /s/ in English demonstrates that /s/ variation does not have a single social meaning, but rather an indexical field (Eckert 2008) of meanings, many of which are related to gender and by extension sexuality (Calder 2019), but also qualities like competence (Campbell-Kibler 2011). Some participants I continued to have relationships with outside of the research context, but most I lost access to once they stepped outside of the recording lab. Consequently, the current study lacks the ethnographic grounding necessary to make assertions about the kinds of interactional or performative goals this /s/ variation is deployed to accomplish among these speakers.

Assuming that the /s/ variation described here is available for social meaning, then we must ask not only "what does it mean" but also "for whom?" Expanding on their collaborative work on /s/ variation among Black speakers in New York and California (Calder & King 2020), Calder (2021) demonstrates that the indexical field of /s/ among Black speakers in Bakersfield, California diverges from what is expected based on previous research on white communities. In short, different communities may have distinct indexical fields for /s/. Although the participants who contributed to this study share a community—that of the university—they occupy a variety of positionalities with respect to age, place of origin, and race. Even those who share an imagined community—such as identifying as bisexual—may have been involved in nonoverlapping communities of practice with potentially divergent indexical fields for /s/ variation. As such, the question of "for whom" has many different answers in this context, none of which I can speak to without a deeper understanding of the participants and their communities.

What /s/ variation means and for whom as it relates to bisexual performance are questions this study cannot answer. Instead, the study argues that such an analysis is impossible in a research paradigm that conflates all forms of non-heterosexuality. It provides quantitative evidence that bisexual is a distinct category from lesbian and gay and thereby lays the groundwork for future research to explore how bisexuality is linguistically performed and taken up. Bisexuality is not just a number on the Kinsey Scale, nor is it a straightforward amalgam of lesbianism/gayness or straightness.[12] Bisexuality is an identity in its own right that is theoretically important in linguistic research on sexuality.

[12]Also known as the Heterosexual-Homosexual Rating Scale, the Kinsey Scale is used to describe a person's sexual orientation (Kinsey et al. 1948).

Chapter 3

Modeling the social dimensions of /s/ production: (Bi)sexuality, race, and gender in random forests

3.1 Introduction

There is a widespread belief in the United States that English-speaking gay men "talk like women" (e.g. speak with higher pitch, pronounce /s/ in a way that sounds "lispy") and that lesbian women "talk like men" (e.g. speak with lower pitch and/or monotone intonation) (Gaudio 1994). Language and sexuality researchers have invested considerable effort in attempting to prove or disprove whether these ideologies have any "real" basis in speech production. There are two key gaps in this literature. First, research in this area typically focuses on monosexual (i.e. lesbian, gay, and straight) speakers to the exclusion of bisexuality. However, Chapter 2 demonstrates that bisexual speakers do not pattern consistently with lesbian/gay or straight speakers with respect to /s/ production. Second, work in this area rarely considers the intersection of sexuality with factors outside of gender or sex, and to a lesser extent geographic location (Campbell-

Kibler 2011; Podesva & Hofwegen 2014). However, Calder and King's recent work with Black communities in California and New York demonstrates that /s/ variation is linked to both speaker race and geographic location (Calder & King 2020). The current study addresses these gaps in the literature through a quantitative analysis that (1) centers bisexual speakers and (2) attends to social factors such as race, place, age, and their intersections. The purpose of this analysis is to explore the efficacy of a particular quantitative method–random forests–for research on bisexuality and /s/ and, by extension, similar sociolinguistic questions.

To that end, we iterate on previous work by analyzing the data presented in Chapter 2 using random forests. Chapter 2 presents three linear mixed-effects regression models of /s/ center of gravity, skew, and duration to analyze the production tendencies of bisexual English speakers relative to lesbian, gay, and straight speakers. That analysis contains valuable insights and adheres to methodological standards established in similar work in the area. However, the scope of the analysis is constrained by a number of factors. Specifically, the models include only a limited number of predictors due to issues of sample size and structure. For instance, Chapter 2 analyzes participants' answers to questions about their gender stereotypicality qualitatively, as these facts could not be included in the regression models due to issues with the ratio of predictors and their levels to sample size. Indeed, these problems are quite common among datasets used in linguistics, which are typically assembled through observation or experiments. These data collection methods tend to engender certain characteristics, such as:

1. non-random sampling
2. small sample size
3. repeated-measures structure
4. an imbalanced dependent variable
5. too many predictors relative to sample size

6. low cell counts

It is often the case that linguistic datasets are compiled using non-random sampling methods, such as convenience or snowball sampling. These methods potentially introduce selection bias into the sample, such that the sample may not be representative of the population intended to be analyzed. This matter of representation may be problematic for statistical tests that assume normality at some point(s) in the model pipeline. For example, significance testing in linear and logistic regression models assumes that the averages of the factors of interest are normally distributed across the different samples of the underlying population. Additionally, linear regression models assume that the residuals after applying the model are normally distributed. Residuals are the difference between the observed and predicted value (i.e. $y - \hat{y}$) of a response variable for any given datapoint. They represent the variation that a model cannot explain. Random forests, as a type of nonparametric test, do not assume a particular distribution at any point, effectively bypassing these potential roadblocks. Furthermore, a small sample size may also lead to cases in which there are few data points relative to the number of predictors and interactions that may be of interest. Even less extreme versions of these so-called "small N large P" cases can lead to cell counts that are too low for reliable parameter estimation when using standard parametric models (Strobl et al. 2009). Regardless of sampling method or sample size, linguistic datasets usually involve repeated measures. That is, participants contribute more than one datapoint in most linguistic studies. Statistical tests that assume observations are independent (e.g. analysis of variance or ANOVA) are not suitable for analyzing data with a repeated-measures structure. For sociolinguists who rely on observational or experimental data, it is crucial to adopt a modeling technique that overcomes the inherently imbalanced nature of such datasets for hypothesis generation, testing, and modeling. We argue that random forests are better equipped for this task compared to more "traditional" modeling techniques (see also Tagliamonte & Baayen 2012, who were early proponents of tree-based methods in linguistics).

A random forest (Breiman 2001) is a tree-based supervised machine learning algorithm used to identify structure in the relations between a response variable and multiple predictors for regression and classification tasks, among others. Random forests contain a predetermined number of trees. For each tree in the forest, the algorithm repeatedly divides the data into successively smaller groups based on the values of the dependent variable. Each division or split within a tree leads to the largest improvement possible in terms of classification or prediction accuracy for the response variable (Ben Youssef & Gries forthcoming; Gries 2021). Once all of the individual trees are grown, a random forest reports a final prediction that is the mean or mode of the predictions from all of the trees (Figure 3.1).

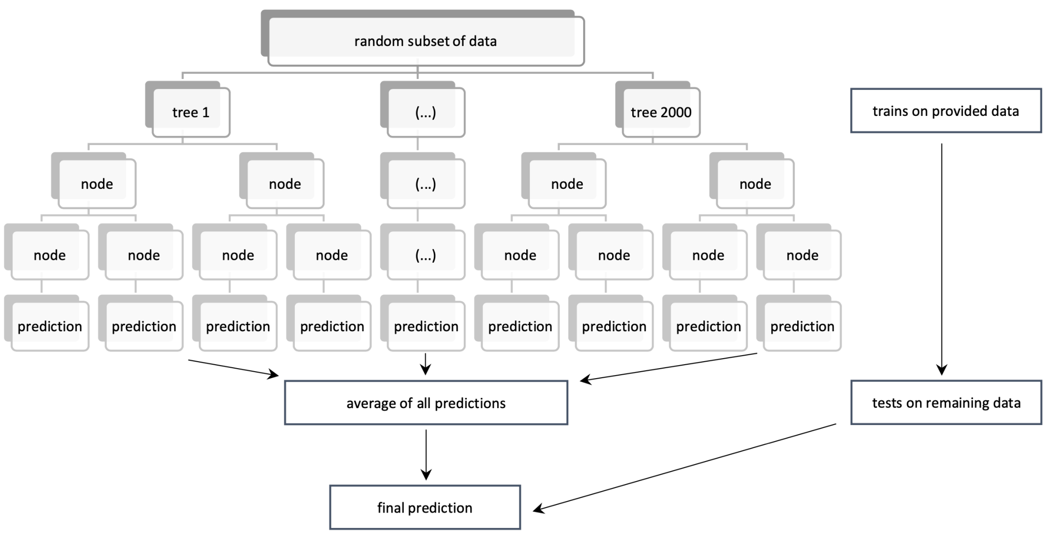

Figure 3.1: Random forest example

Let us make random forests more concrete using a hypothetical example generated from our data. In this example, we use random forests to model differences in /s/ center of gravity (COG) (see Section 3.2). First, we feed the model our data and indicate what is the dependent variable (COG) and what factors to include (for this example: GENDER, SEXUALITY, RACE). We also tell the model how many trees to grow (n = 2000) and how many factors to consider at each split (n = 2). The algorithm then separates the COG data into two groups: the training data that it uses to make predictions, and the testing data that it uses to evaluate its predictions. Once the data has been separated, the model begins growing individual trees. For each tree (e.g. tree 1 in Figure 3.1), the algorithm randomly selects two of the three factors to consider when making the first split. From those two factors, the algorithm selects the factor that, broadly speaking, improves model performance the most to make the split.[1] The example tree in Figure 3.2 splits first on RACE.[2] Each subsequent split follows the same process: the algorithm randomly selects two of the three factors to consider, and from those two chooses the factor that improves the model the most. The terminal nodes contain the models' COG predictions based on the previous splits. The model then averages these predictions and evaluates its work based on the set-aside testing data. It then generates a final prediction. At this point, we can examine the model evaluation metrics (e.g. R^2), variable importance scores (similar to univariate results, see Section 3.4), and interactions.

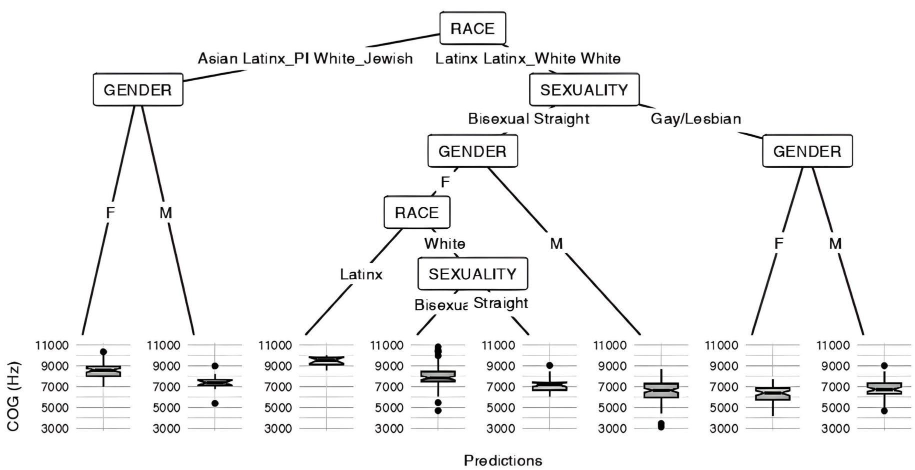

Figure 3.2: A hypothetical tree in a random forest analyzing center of gravity (COG)

Why are random forests suitable for data with characteristics (1-6) (and indeed, our data)? Random forests are a type of nonparametric test and therefore do not assume a particular data

[1]Our dependent variable in this example (and in all our reported models) is continuous, such that our random forest is a regression one. As such, this model splits based on weighted mean-squared error.

[2]The trees shown in Figure 3.2 is for heuristic purposes only. It is not extracted from the COG random forest presented in Section 3.4.1.

distribution, normal or otherwise. Therefore, they are appropriate for many cases a linguist might find themselves in: when the sampling method used is not random (e.g. snowball and convenience sampling), when the data are not normally distributed, or when the sample size is too small to make an educated guess about the true shape of the distribution (1, 2). Additionally, random forests are relatively resistant to the effects of empty or low cell counts because they randomly select observations, making them suitable for "small *N* large *P*" cases and/or datasets that are unbalanced due to non-random sampling (1, 2, 5, 6).[3] Next, random forests do not assume independence of the observations and thus are effective at dealing with repeated measures (3). Finally, random forests are often superior when predicting an imbalanced response variable, i.e. one that is characterized by an uneven distribution of its levels (4) (Muchlinski et al. 2016).

In addition, random forests add two layers of randomness that produce a number of advantages compared to more "traditional" modeling methods in linguistics. Randomness is desirable because it guards against bias in the sample and, in ensemble models like random forests, protects against errors in individual trees. The first layer of randomness in a random forest happens at the tree level (e.g. tree 1 in Figure 3.1). Individual trees are grown using bootstrapping or random sampling with replacement. In other words, only a random subset of the data is considered when growing a tree (as opposed to the entire dataset). The data set aside at this stage is later used to test the accuracy of–or to validate–the results of the tree (e.g. "tests on remaining data" in Figure 3.1). Thus, unlike other modeling approaches commonly used in linguistics, such as linear regression or ANOVA, model validation is built into the process of growing a random forest. The second layer of randomness occurs at the level of the split (the transition from tree to node in Figure 3.1). For each split in a tree, a random forest considers a randomly selected subset of predictors. Together with bootstrapping, limiting splits "decorrelates trees, helps identify the

[3] A caveat is that any method based on resampling cannot extrapolate beyond the data given as input. Random forests may still miss patterns that do not have strong evidence in the provided sample due to data sparsity issues, and there is no guarantee that the patterns they identify generalize beyond the provided sample. In other words, there are still risks associated with particularly small effective sample sizes (i.e. both the overall number of samples, but also the number of samples per cell).

importance of predictors and their interactions to the predictions, avoids collinearity problems, and protects against overfitting" (Ben Youssef & Gries forthcoming: 9).[4][5] To summarize, the data and predictors each tree "sees" are randomized, which makes separating the wheat from the chaff and generalizing beyond a given dataset more straightforward. These layers of randomness are not included by default in other methods like linear regression or ANOVA. Model validation through bootstrapping is certainly possible with other methods, but it is not automatically included as a part of the fitting process. Recreating the randomness introduced at the level of the split would be incredibly tedious and computationally demanding in a linear regression or ANOVA model. All of this is not to say that other modeling techniques are wholly inappropriate for (quantitative socio-) linguistic questions. Indeed, the linear mixed-effects regression model presented in Chapter 2 of this dissertation identifies patterns that engender powerful qualitative insights (and, as shown below, this chapter reinforces and expands upon those insights). Rather, random forests are a viable option when the idiosyncrasies of the data make other modeling approaches problematic or limiting. In this chapter, we demonstrate the application of random forests in the context of quantitative sociolinguistic research, namely, the relationship between social factors and /s/ production.

3.2 /s/

This chapter uses random forests to identify correlations between particular groups or axes of identity and patterns of variation in /s/ production. Given the methods used to collect our data (that is, lab-based experiments with participants unfamiliar to us), we lack the ethnographic understanding necessary to make assertions about how this variation is used and taken up in

[4]We take "collinearity problems" to refer to, for example, multicollinearity or a phenomenon in which one predictor variable in a multivariate regression model is linearly predicted by another predictor variable at above-chance levels. Multicollinearity is problematic when determining what variables are important for the analysis.

[5]*Overfitting* refers to a situation in which an analysis corresponds too closely to a particular data set, which is problematic when the goal is to make predictions that apply beyond the training data.

everyday life. In turn, our ability to theorize what these correlations mean for theorization of /s/ and its indexical field is limited. Therefore, we focus on the implications of our findings for research practice, rather than theory, in the discussion. Regardless, we overview the sociophonetic literature on /s/ and its indexical field in this section because it provides important context for what measures and factors we examine and why.

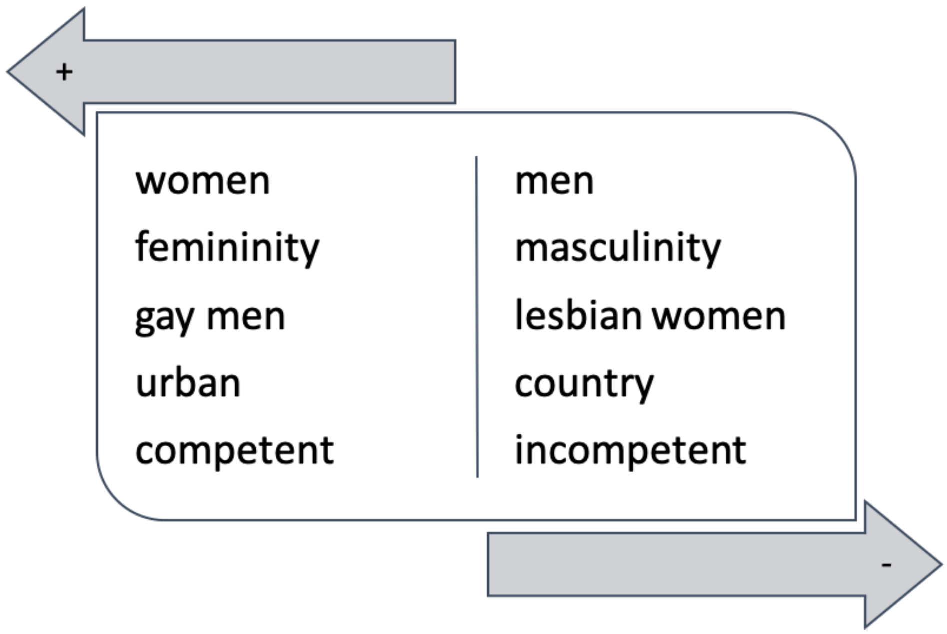

Figure 3.3: Indexical fields of fronted [s+] and retracted [s-] based on previous sociolinguistic research (adapted from Calder (2021))

Variation in /s/ production has been linked to a number of identity categories and social characteristics in North-American, English-speaking contexts (Figure 3.3). From gender, to sexuality, region, and class, the /s/ sound has a complex web of interrelated social or indexical meanings (e.g. Calder 2021; Campbell-Kibler 2011; Munson 2007; Podesva & Hofwegen 2014). This voiceless anterior fricative is produced by creating a narrow constriction between the tongue and the alveolar ridge, such that the airstream passes over the back of the top teeth. Differences in /s/ articulation are often discussed in terms of a cline of frontness [+] to backness [-]. On the front end of the spectrum, the tongue is placed relatively close to the teeth (i.e. in the front of the mouth), which creates a high frequency hissing sound. On the back end of the spectrum, the tongue is retracted away from the teeth (i.e. towards the back in the mouth), creating a lower frequency hissing sound (Calder 2021; Fuchs & Toda 2010).

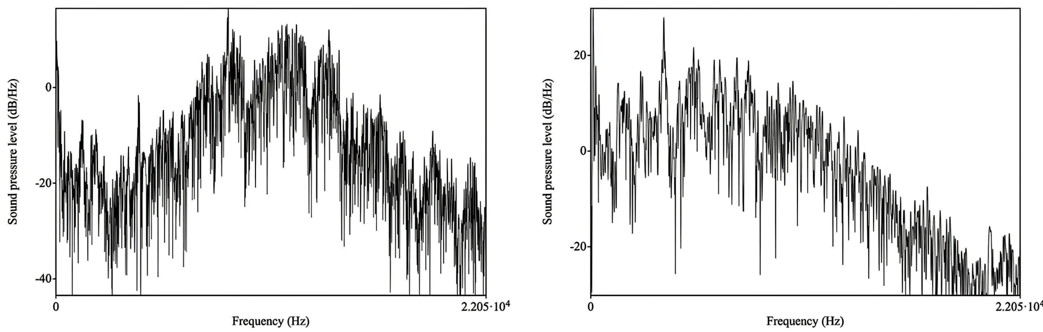

Figure 3.4: A high COG /s/ produced by a bisexual woman (left) and a low COG /s/ produced by a bisexual man (right) (from Willis forthcoming, Chapter 2)

Differences in tongue position are connected to two key acoustic correlates in the literature: center of gravity and skewness. Both of these measures contribute to how "hissy" an /s/ sounds. The first factor, center of gravity (COG), refers to the weighted mean of frequencies in the spectrum (Figure 2.1).[6] Center of gravity typically ranges from 6.4 kHz to 8.5 kHz for English-speaking women and from 4 kHz to 7 kHz for English-speaking men (Zimman 2017). Studies of English speakers suggest that gay men produce /s/ with higher COG than straight men (e.g. Linville 1998; Munson et al. 2006a; Podesva & Hofwegen 2014). Perceptually, higher COG /s/ tokens are consistently evaluated as more gay-sounding and less masculine-sounding relative to low COG tokens (Campbell-Kibler 2011; Linville 1998; Rogers & Smyth 2003; Zimman 2013). Studies of women's speech are less coherent, with some research finding that lesbian women produce lower COG /s/ than straight women (Hazenberg 2015; Munson et al. 2006a) and others finding no significant difference (Barron-Lutzross 2015). Regardless, there appears to be relatively reliable distinctions in /s/ vis-à-vis gender and sexuality, whether they be described in articulatory (e.g. front vs. back) or acoustic (e.g. high vs. low COG) terms.

The relationship between /s/ center of gravity, gender, and sexuality is further inflected by

[6]Center of gravity is sometimes also known as the centroid.

axes of social difference such as competency, regional affiliation, age, and race. Campbell-Kibler (2011) reports a complex web of indexical associations between /s/ frontedness (as measured by COG), masculinity, sexual orientation and competency. Specifically, she reports that listeners tend to cluster higher COG /s/ with “intelligent, effeminate, gay men” on the one hand and lower COG /s/ with “unintelligent, masculine, straight men” on the other. She also finds a relationship between /s/-fronting and regional variation, such that lower COG /s/ tokens are associated with southern U.S. dialects and sounding “country”. Likewise, Podesva & Hofwegen (2014) study of English speakers in Redding, California reports that /s/ production varies relative to gender, sexuality, age, and orientation to “country” (i.e. a socially conservative political ideology defined in opposition to “city” or urban life). Gay men in Redding produce higher COG /s/ than straight men in their community, albeit within the limits of local gender norms. /s/-fronting also inversely correlates with age in this population, such that older speakers produce lower COG /s/, and this effect is driven by country-oriented participants. Finally, work by Calder & King (2020) discusses the relationship between /s/, gender, and region in Black communities. They report gender-based distinctions in /s/ frontedness among African American speakers in urban Rochester, New York, but do not find such differences among African American speakers in non-urban Bakersfield, California. In short, /s/ variation as measured by COG has a complex indexical field (Eckert 2008) of interrelated social meanings related to gender, sexuality, region, age, and race.

In addition to center of gravity, /s/ frontedness is also associated with an acoustic measure referred to as skew. Skew or skewness, the third spectral moment, measures spectral tilt, or how fast energy falls off as frequency increases (Thomas 2017). Positive skew indicates a concentration of energy in the lower frequency ranges, whereas negative skew indicates concentration in the higher frequency ranges (Figure 2.2). Zero skew indicates a symmetrical distribution of energy around the mean (Jongman et al. 2000). Fronted /s/ tend to have lower (often negative) skewness. Conversely, retracted /s/ tend to have higher (often positive) skewness

(Podesva & Hofwegen 2014). A study by Munson and colleagues (2006a) reports that English-speaking women produce /s/ with a more negative skew than men. Additionally, they find that gay men produce /s/ with a more negative skew than straight men. Regarding perception, the same study reports that men who produced /s/ with a more negative skew were more often rated as gay-sounding. Overall, evidence suggests that skew varies as a function of gender and of sexuality. To our knowledge, skewness has not been discussed relative to other axes of social difference in the literature on /s/ variation in U.S. Englishes.[7]

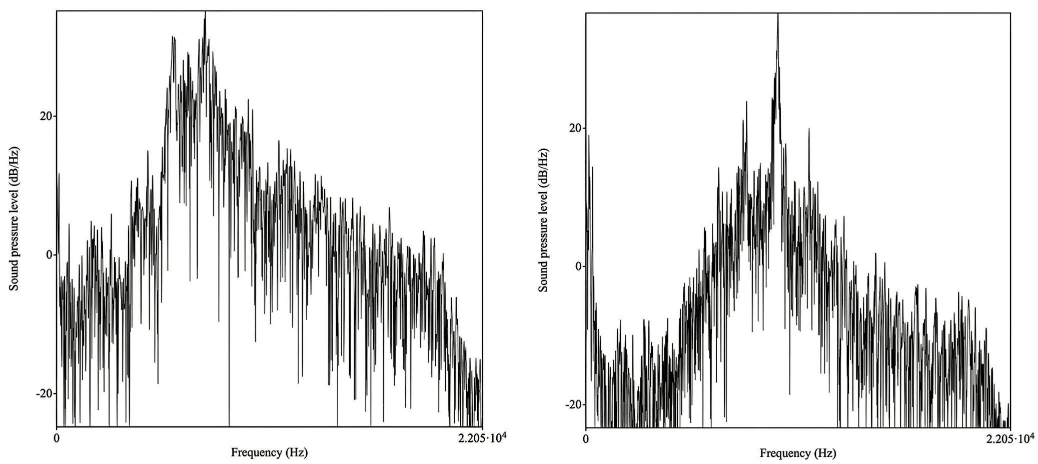

Figure 3.5: A negatively skewed /s/ (left) and a positively skewed /s/ (right) produced by two bisexual women (from Willis forthcoming, Chapter 2)

Aside from frontness and backness (as measured by COG and/or skew), duration is also a relatively well-investigated dimension of difference for this fricative. There are several studies demonstrating a relationship between /s/ duration and sexuality. Crist (1997) reports that /s/ is lengthened in stereotypical gay men's speech, albeit in limited phonological contexts. Linville (1998) finds that gay men tend to produce longer /s/ than the straight men in her sample. Her analysis also finds that duration correlates with both perceived and reported sexual orientation. Rogers et al. (2000) report that speakers rated as gayer-sounding produced significantly longer /s/ than those rated as straight-sounding. To summarize, /s/ duration appears to have a relatively

[7]Podesva & Hofwegen (2014) analyzed skew but did not report the results.

straightforward relationship with sexuality among English-speaking men. Specifically, longer /s/ are associated with gay men from both a production and perception standpoint. Similar to skew, duration has not been discussed relative to other axes of social difference in the literature on /s/ variation in U.S. Englishes to our knowledge.

We continue in the tradition of sociophonetic research on English /s/ and apply random forests to center of gravity, skew, and duration. First, however, we discuss the details of our data collection, stimuli, and study design.

3.3 Data

3.3.1 Speakers

Twenty-seven cisgender native English speakers from varying ethnoracial backgrounds and places of origin were recruited from a California university (Table 1). Speakers ranged from 18 to 30 years old at the time of recording and self-identified as bisexual (n = 7), lesbian (n = 5), gay (n = 5), or straight (n = 10). Participants were recruited using snowball sampling and through flyers distributed electronically and posted in various physical locations on the university campus.

3.3.2 Stimuli

Participants were recorded reading the Rainbow Passage (Fairbanks 1960), a phonetically balanced scientific passage about rainbows, in a private, sound-attenuated booth using Audacity (Audacity Team 2019). They were instructed to read the passage twice only and as naturally as possible. Next, /s/ tokens were extracted from participants' first readings of the passage using a Praat script (Zimman 2018). The script generated measurements for center of gravity and skew based on long-term average spectra. Token start and end points, as well as duration, were

Group	n	Mean age	Age range	Race/ethnicity	Origin
Bisexual women	5	21	18-26	1 Latinx/PI[8], 4 white	US
Bisexual men	2	19	18-19	1 Latinx, 1 white	US
Lesbian women	5	21	19-26	1 Latinx/white, 3 white, 1 white/Jewish	3 US 2 US & Europe
Gay men	5	25	18-30	1 Asian, 1 Latinx/white, 3 white	4 US 1 India
Straight women	5	19	18-20	2 Asian, 1 Latinx, 2 white	4 US 1 declined
Straight men	5	19	18-20	3 Latinx, 2 white	US

Table 3.1: Speaker information

manually checked by the first author.

3.3.3 Design

We created three random forests using the randomForestSRC package (Ishwaran & Kogalur 2022) and the ggRandomForests package (Ehrlinger 2016) in R (R Core Team 2022). For each random forest, the predictors were as follows: GENDER, RACE, FEMININITY RATING, MASCULINITY RATING, SEXUALITY, REGION, AGE, WORD, SPEAKER. Self-reported gender stereotypicality ratings were coded as ordinal variables with 7 levels (1 = not at all feminine/masculine, 7 = very feminine/masculine).[9] RACE was coded such that mixed identities were conflated into a single level (e.g. Latinx/white), a choice that is explored in the discussion section. Participants' places of origin were also conflated into regions for interpretability: East coast, West coast, Southwest, Midwest, Rocky Mountains, India, multiple, and declined. SPEAKER and WORD function similarly

[9]The post-test survey asked participants "how stereotypically feminine/masculine do you consider yourself" (as opposed to normatively feminine/masculine) in order to elicit a comparison between themselves and idealized (rather than average or prototypical) femininities and masculinities.

to random effects in mixed-effects models in that they theoretically capture any individual differences between participants or structural differences between words that are not covered by the other named predictors in the forest.

Although we maintain the same predictors in all three models, we fit each random forest using different hyperparameters that performed optimally during the development stage: the number of trees to grow in each forest (ntree) and the number of randomly sampled predictors to use in each tree (mtry). Table 3.2 summarizes the hyperparameters as well as the resulting error rates and R^2 (variation in the dependent variable that is predictable from the predictors) for each model. Model error is a measure of how often the model is wrong, such that a low error rate (relative to the baseline) is desirable. Our baseline models are fit as a single tree with only the most important predictor in each forest (for the notion of importance, see below).[10] R^2, on the other hand, is a measure of the percentage of variance in the outcome that is explained by the predictor variables. A perfect R^2 of 1.00 indicates that the predictor variables explain 100% of the variance in the outcome that the model aims to predict. Thus, while a high R^2 is generally desirable, a very high R^2 value may be indicative of overfitting.

	Hyperparameters		**Model statistics**		
Model	***ntree***	***mtry***	***Baseline error rate***	***RF model error rate***	***R^2***
COG	900	6	1382926	693492.6	0.56377
Skew	800	5	0.913773	0.528616	0.531209
Duration	1100	8	0.000548	0.000315	0.573064

Table 3.2: Hyperparameters and statistics for each model

When reporting and interpreting the results of a random forest, it is crucial to examine variable importance scores as well as partial dependence plots (Gries 2021). Simply put, variable importance scores (VIMP) estimate the importance of a variable by comparing the performance

[10]The difference in scale for the baseline error rate and RF model error rate of the three models is related to differences in the response variable. These values are calculated as mean-squared error. That is, they represent the average of the squared differences between the observed and predicted value for each datapoint not used to train the model.

of the estimated model without the variable in it. More specifically, VIMP scores reflect the absolute size of the effect of a predictor on the response variable. In random forests used for regression tasks (such as the ones reported here), VIMP scores represent the equivalent of how far regression coefficients of (z-standardized) predictors are from zero in whatever direction. Figure 3.6, Figure 3.12, and Figure 3.15 in the following subsections plot VIMP values computed by randomly permuting each variable's values and comparing the prediction error to that of the observed values. A large VIMP score suggests that the variable is important for obtaining accurate predictions, whereas a VIMP score closer to zero suggests that the variable contributes little to the accuracy of predictions. In Figure 3.6, Figure 3.12, and Figure 3.15, factors with high VIMP scores relative to other factors in the model are indicated in purple, and factors with relatively low VIMP scores are indicated in green.

Crucially, VIMP scores are not the same as the monofactorial results of linear regression models. In a linear regression model, the monofactorial results indicate the impact of each predictor on the dependent variable. It measures the relationship between a given predictor and the predictor independently, assuming all other variables remain constant. In contrast, VIMP scores capture the collective importance of variables in making accurate predictions. Instead of analyzing each predictor variable independently, the algorithm considers combinations of variables and evaluates how much the model's predictions would suffer if a particular variable is perturbed or removed. Another way of thinking about VIMP scores is that they are similar to the Chi-squared statistics reported when dropping a particular term, as shown in Chapter 2 (as opposed to the coefficient estimates of a linear regression model). Thus, our discussion of VIMP scores in the following sections does not resemble what one might expect based on how monofactorial results in linear regression models are typically presented in the literature.

Combinations of predictor variables are considered when calculating VIMP scores, but they are not the same as interactions. VIMP scores quantify the importance of individual variables for prediction accuracy. Interactions, on the other hand, quantify the combined effects of two or more predictors that cannot be explained by their individual effects alone. We employ a joint-variable approach (Ishwaran 2007) to calculate the interactions between predictors. In this approach, the paired importance of each pair of variables is calculated and then subtracted from the sum of both variables' VIMP scores. A large association value indicates that the interaction is worth exploring (but not necessarily significant) when the univariate VIMP score for both variables is also relatively large. What constitutes a "high" or "low" association value is decided by the researcher, similar to the process of setting a significance threshold (Ben Youssef & Gries forthcoming). We report interactions based on three criteria: (1) both factors in the interaction have high VIMP values relative to other factors, (2) the interaction has a high association value relative to other possible interactions, and (3) the interaction is theoretically important for this particular analysis. For each model, we select up to four interactions that fit two or more of these criteria for in-depth discussion for the sake of brevity.

3.4 Results

3.4.1 Center of gravity (COG)

Figure 3.6 features the VIMP scores for COG. GENDER, RACE, SEXUALITY, and REGION all have relatively large effects on the forest's predictions and make substantial contributions to prediction accuracy. The low scores for MASCULINITY RATING, and FEMININITY RATING, SPEAKER, WORD, and AGE indicate that there is little variation in COG across these factors in this sample. However, the gender stereotypicality ratings contribute to the discussed interactions hereafter.

GENDER:RACE is the most important interaction in the COG random forest (Figure 3.7). Note that not all GENDER:RACE comparisons are represented, as no Latinx/PI or white/Jewish men participated in the study. Women produced higher COG estimates than men within each ethnoracial category in the sample. The gap between predicted COG is largest among monoracial

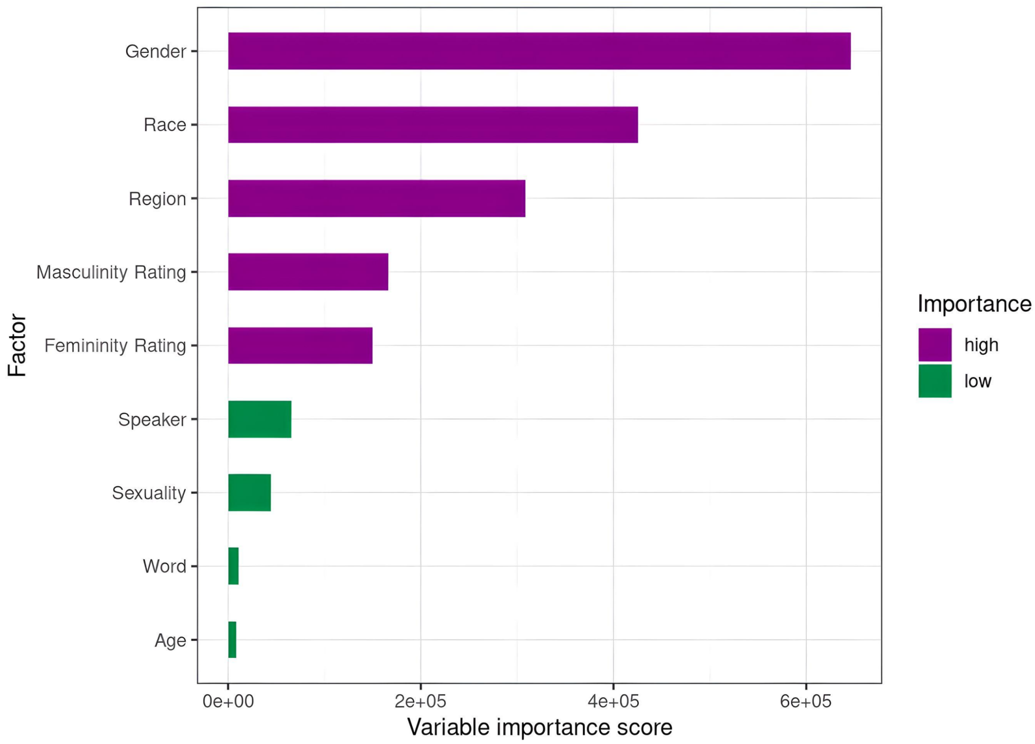

Figure 3.6: COG VIMP scores

Latinx women (n = 1) and men (n = 4) at approximately 2.5 kHz. Gender-based differences in predicted COG are more moderate between monoracial Asian (n = 3) and white speakers (n = 15) and rather minimal among mixed race Latinx/white speakers (n = 2).

GENDER:MASCULINITY RATING was the second most important interaction in the COG random forest. Predictions generated from this sample situate the estimates for women around 7.45 kHz - 7.9 kHz and between 6.45 kHz - 6.7 kHz for men (Figure 3.8). Among women, estimated COG generally increases as masculinity rating increases. In other words, the more masculine a woman rates herself, the higher her COG predictions are. As for men, estimated COG generally decreases as masculinity rating increases. Put differently, the more masculine a man rates himself, the lower his COG predictions are. The former finding is unexpected given the literature on /s/ COG among English-speaking women, whereas the latter finding aligns with previous research on COG among English-speaking men.

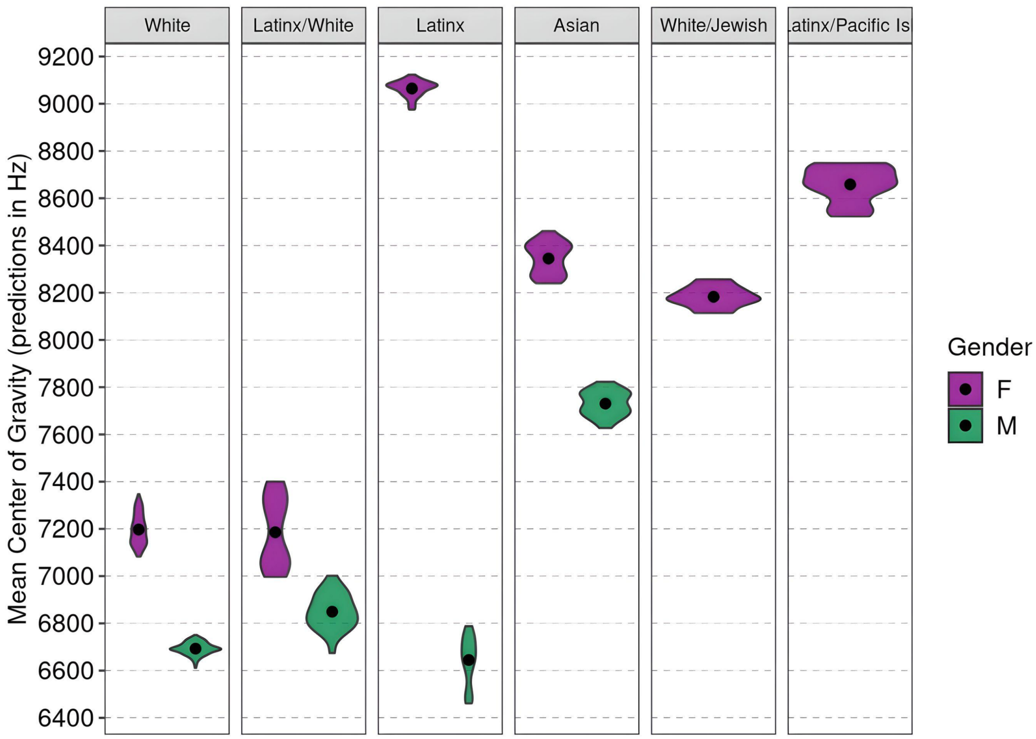

Figure 3.7: GENDER:RACE partial dependence co-plot for COG

GENDER:FEMININITY RATING is the next most important interaction in the COG random forest (Figure 3.9). Viewing women's femininity ratings holistically, COG estimates generally increase as femininity ratings increase. That is, the more feminine a woman rates herself, the fronter her predictions are. However, COG estimates peak at a femininity rating of 5 out of 7 and then decreases slightly. Among men, there is a dramatic increase in estimated COG between men who rated themselves as not at all feminine (1) to a little feminine (2), and then another slight increase between a little (2) and moderately (3-4) feminine.

The differences in the intra-gender COG predictions featured in Figure 3.8 and Figure 3.9 are very small, sometimes less than 100 Hertz. This raises the question of whether these differences are perceptible. Applying Weber's Law to fricative spectra, fricatives in which energy is skewed towards the higher frequencies like /s/ require larger changes in Hertz for differences between

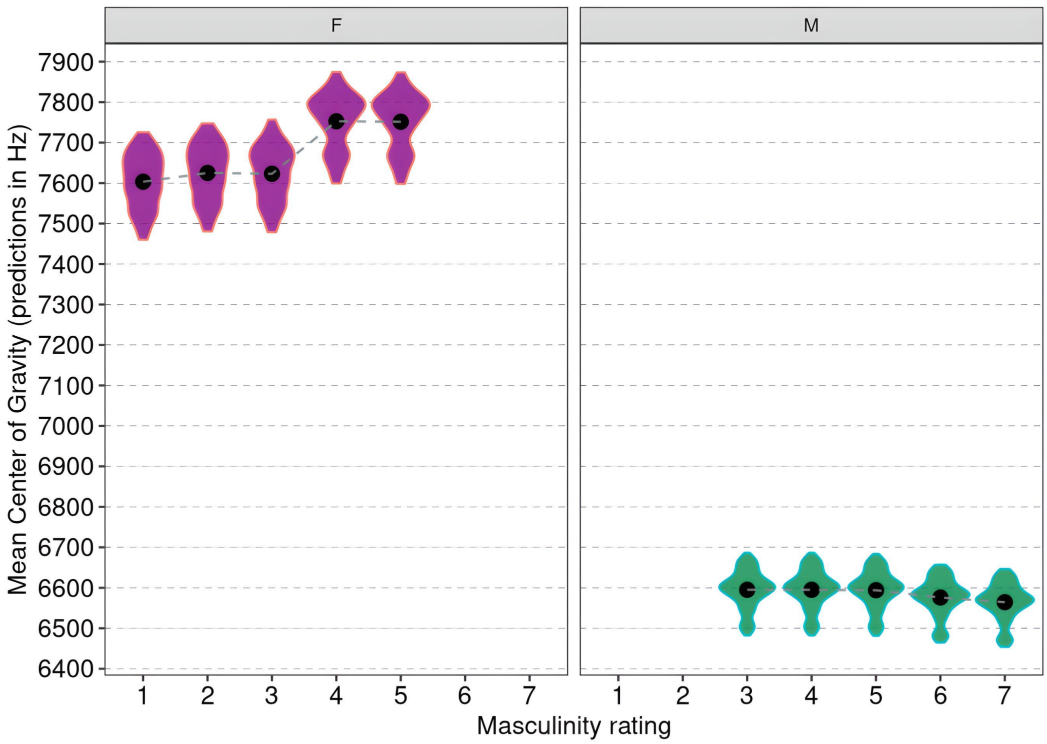

Figure 3.8: GENDER:MASCULINITY RATING partial dependence co-plot for COG

tokens to be salient.[11] By extension, /s/ tokens with a particularly high COG ostensibly require a rather substantial change for differences to be perceptible. Anecdotally, we are able to hear the differences between tokens (7-9), extracted from our data:

7. /s/ produced by a lesbian woman (6619 Hz COG)

8. /s/ produced by a bisexual woman (9286 Hz COG)

9. /s/ produced by a straight woman (9151 Hz COG)

Anecdotes aside, previous literature suggests that the Just Noticeable Difference (JND) between sounds sharply increases after 4 kHz (Figure 3.10) (e.g. Forinash & Christian 2012). Still, even a difference of less than 100 Hz is theoretically perceptible for sounds in the 9 kHz

[11]Weber's Law refers to how humans are less sensitive to higher frequency sounds, such that the actual change must be greater to perceive differences between higher frequency stimuli compared to lower frequency stimuli.

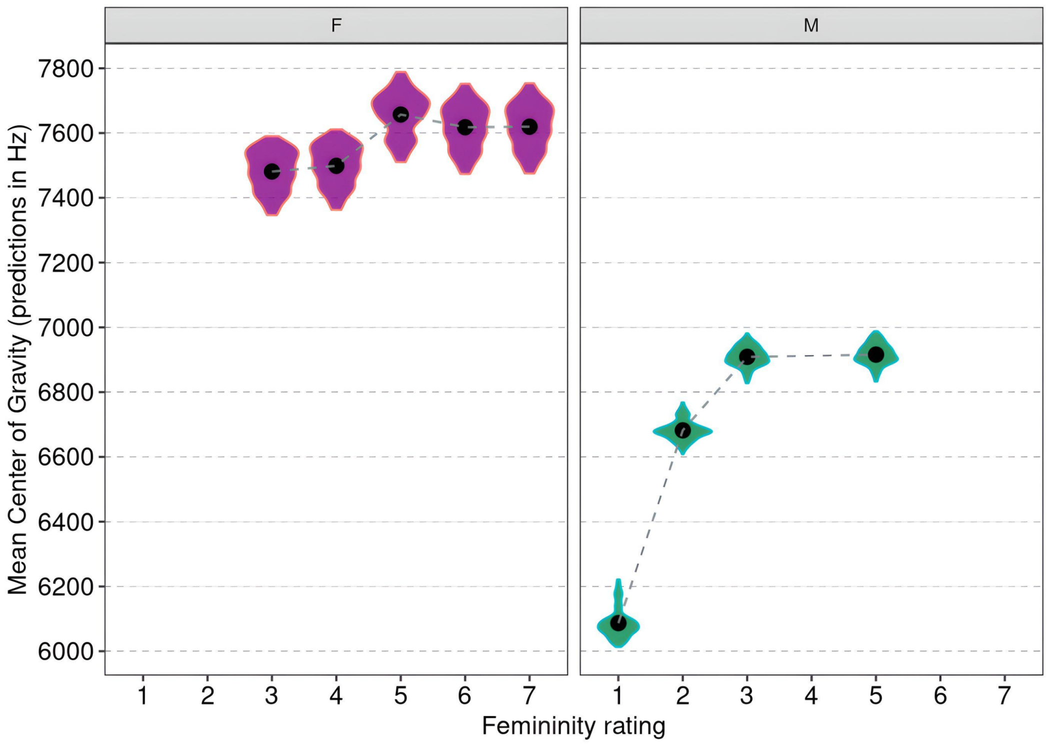

Figure 3.9: GENDER:FEMININITY RATING partial dependence co-plot for COG

range. However, the literature on JND is based on pure tones, whereas the energy in fricatives like /s/ covers a broad range of frequencies. It is not entirely clear how JND maps onto the contributions of noise at different frequencies, but there is some evidence to suggest that listeners are able to hear the small differences in COG reported here. Regardless, what is certain is that the acoustic differences between men with respect to femininity rating are much more likely to be salient than the other intra-gender differences reported in Figure 3.8 and Figure 3.9.

Finally, SEXUALITY:GENDER is the last interaction above the importance threshold for COG Figure 3.11. Note that SEXUALITY as a single predictor is not important according to the VIMP values in Figure 3.6, but it participates in this important interaction so we discuss it here. Overall, straight women produce the highest COG estimates, followed by bisexual women, bisexual men, straight men, gay men, and finally lesbian women. Considering the within-sexuality, cross-gender comparisons, the difference between women and men's predicted COG values is greatest among

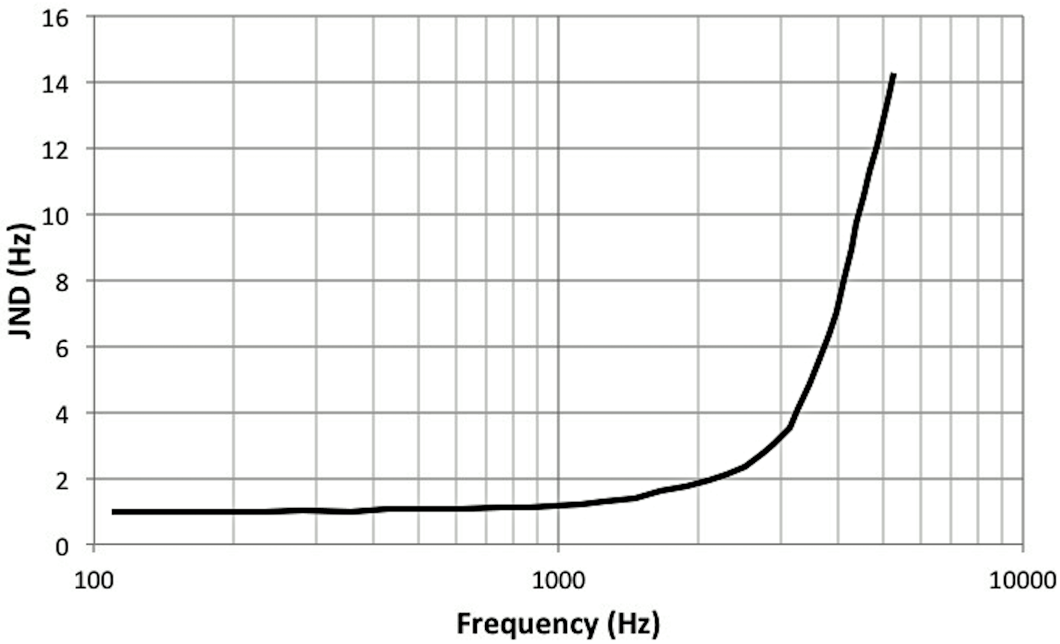

Figure 3.10: fig:Just Noticeable Differences in Frequency (from Forinash & Christian 2012

straight participants, followed by bisexual participants, and finally lesbian/gay participants. In the former two groups, women produce higher COG estimates than men. Unlike straight and bisexual speakers, however, lesbian women produce lower COG estimates than gay men.

Turning to the within-gender, cross-sexuality comparisons, the differences between the three groups of men are rather minimal. In contrast, there is a visually salient difference between straight and bisexual women's COG estimates compared to lesbian women's, such that lesbian women produce substantially lower COG estimates than the other two groups of women. The difference between straight and bisexual women is also more extreme than the differences between the groups of men. In short, there seems to be much wider variation in predicted /s/ COG between the three groups of women than the groups of men in this sample.

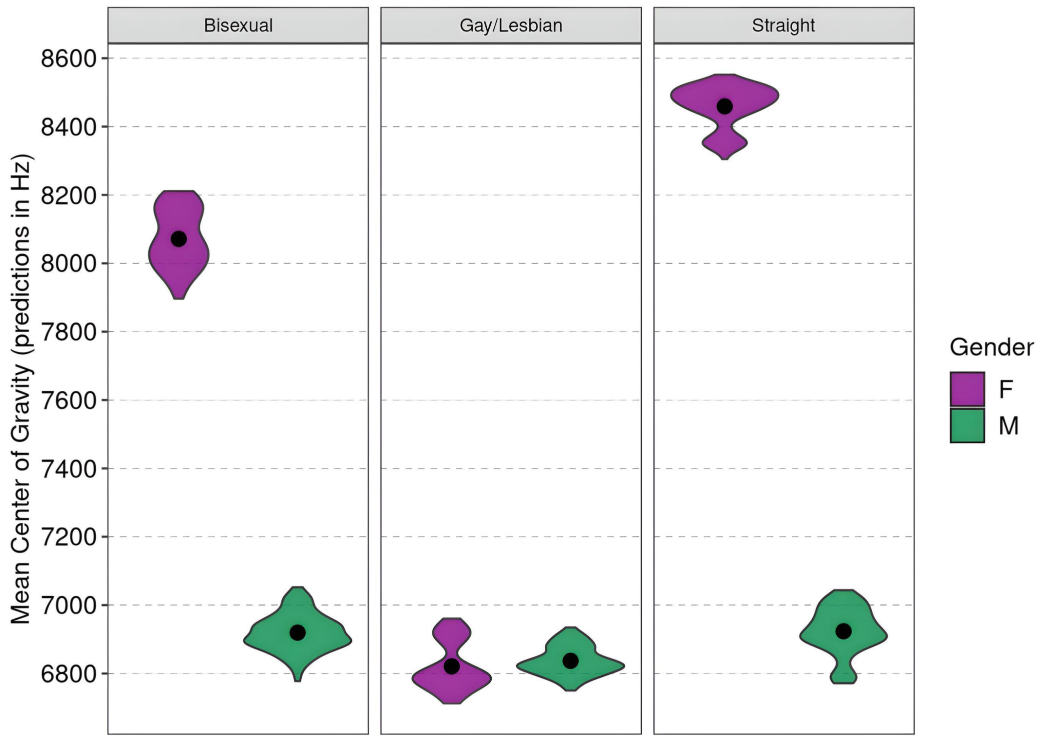

Figure 3.11: GENDER:SEXUALITY partial dependence co-plot for COG

3.4.2 Skew

For skew, VIMP scores indicate that SPEAKER and RACE have a relatively large effect on the forest's predictions and make substantial contributions for prediction accuracy (Figure 3.12). The low scores for GENDER, REGION, the gender stereotypicality ratings, SEXUALITY, AGE, and WORD suggest that these factors contribute little to predicting skew within this sample.

Figure 3.13 features the SPEAKER partial dependence plot for skew. Speaker 1, a white bisexual woman, and speaker 2, a mixed-race Latina/white lesbian woman, are distinct outliers compared to the majority of speakers. Both of these speakers' skew predictions are substantially higher relative to the rest of the speakers in the sample. Speaker 3, a white lesbian woman, and speaker 27, the gay man who was born in India, are also both outliers, albeit less extreme ones. The former's estimates are moderately higher than the rest of the sample and the latter's

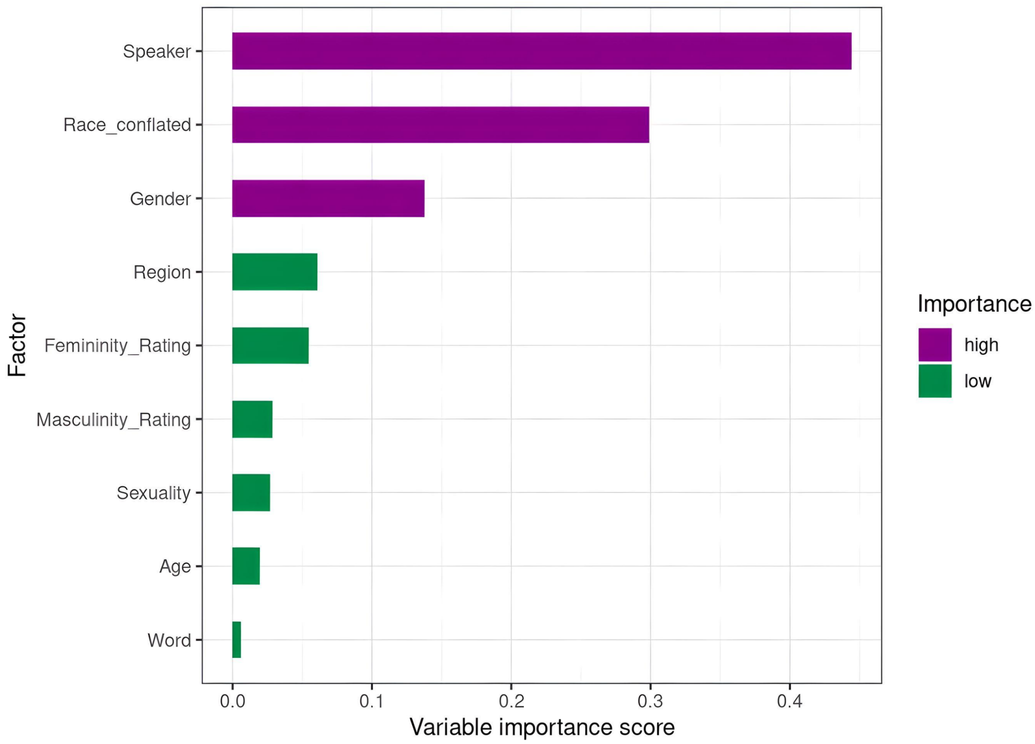

Figure 3.12: Skew VIMP scores

estimates are moderately lower.

We elected not to eliminate outliers for two reasons. First, maintaining sample size across models is ideal for model comparison. Removing potential outliers from the skew model but not the COG or duration models makes qualitative interpretation of the differences between models more difficult. Second, removing speakers identified as potential outliers in terms of one dependent variable from all three models undesirably reduces our already small sample size. Therefore, we chose to keep these three speakers in all three random forests.

SPEAKER is often included in experimental linguistic studies as a "catch-all" factor. Ideally, SPEAKER captures differences between participants not otherwise accounted for by other factors included in the model. For example, SPEAKER is often included as a random effect in mixed-effects modeling to capture random "unlabeled" variation that occurs naturally between people. SPEAKER serves more-or-less the same function in our random forests–it is the bin that catches

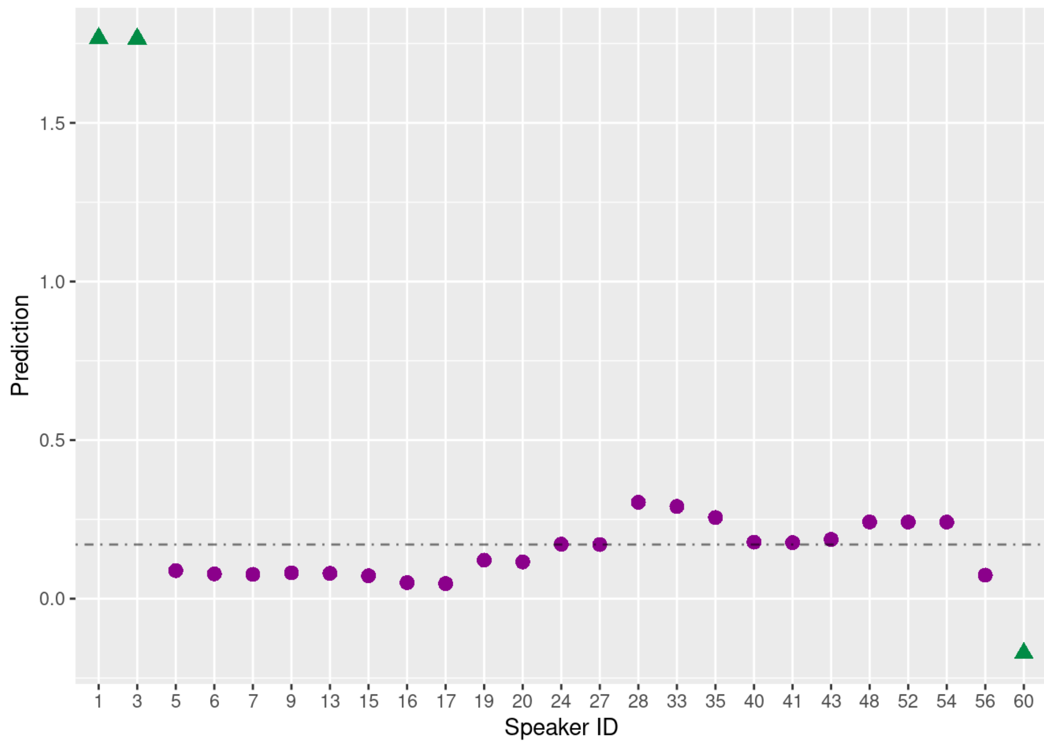

Figure 3.13: SPEAKER partial dependence plot for skew

everything not covered by GENDER, RACE, SEXUALITY, AGE, REGION, the stereotypicality ratings, and WORD. As such, the importance of SPEAKER indicates that unspecified differences are more informative for predicting skew than the social or structural factors named in our random forest. We offer two plausible interpretations of the importance of SPEAKER. The first is that speakers may manipulate /s/ frontedness in distinct ways, such that some speakers do not employ skew. Instead, they may front /s/ in such a way that COG is the stronger acoustic correlate. The second explanation is that there may be some social factor collapsed into SPEAKER (e.g. socioeconomic class, multilingualism) that drives the contribution of this variable. To fully tease apart SPEAKER is outside the scope of this chapter, but is certainly an area of interest for future work in this area.

Figure 3.14 features the partial dependence co-plot of the interaction between GENDER:RACE, the only important interaction for skew. Similar to Figure 3.7, not all gender and race combinations are represented in this sample. Across ethnoracial groups, there is an overall increase

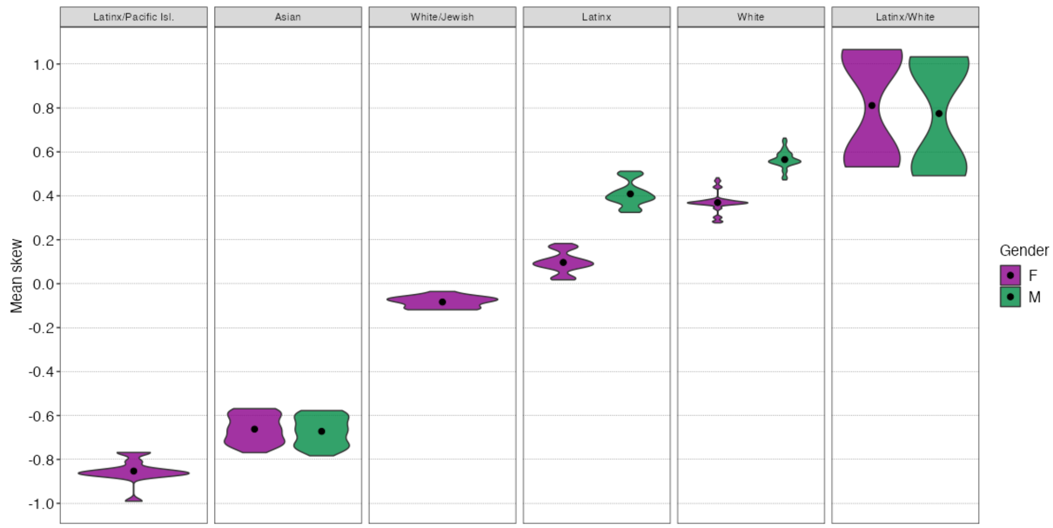

Figure 3.14: GENDER:RACE partial dependence co-plot for skew

in skew estimates when viewing Figure 3.14 from left to right. The mixed-race Latina/Pacific Islander woman produced the most negative skew estimates, followed by the Asian speakers, the white/Jewish speaker, the monoracial Latinx speakers, the monoracial white speakers, and finally the mixed-race Latinx/white speakers with the most positive skew estimates. Next, consider the general direction of difference between women and men in a given ethnoracial category. Overall, the difference between women and men's skew estimates is not consistent across ethnoracial groups. Monoracial Asian women and mixed-race Latinx/white women produce more positive skew estimates than men in the same group, whereas monoracial Latinx women and monoracial white women produce more negative skew estimates compared to their masculine counterparts. Finally, consider the degree of difference between women and men in a given ethnoracial category. The largest distinction is between monoracial Latinx women and men in this sample. Latina women produce substantially more negative skew estimates than Latino men. We also find moderate differences in predicted skew between monoracial white women and men, such that white women produce lower estimated skew than white men. We find minimal differences in estimated skew between monoracial Asian and mixed race Latinx/white women and men. Women produced /s/ with higher skew than men in both of these groups. To summarize, skewness varies considerably relative to gender and race/ethnicity across and within different groups.

3.4.3 Duration

Figure 3.15 indicates that SEXUALITY and WORD have a relatively large effect on the forest's predictions and make substantial contributions to prediction accuracy. The remaining factors received comparatively low VIMP scores, such that these predictors contribute little to predicting /s/ duration.

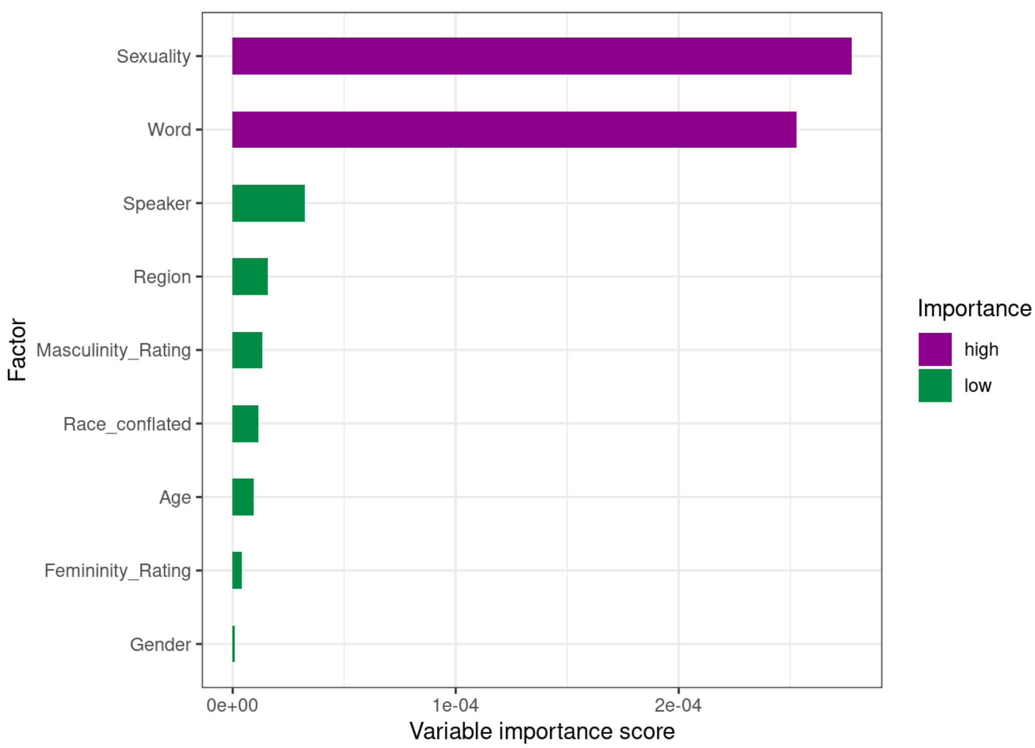

Figure 3.15: Duration VIMP scores

Results indicate that SEXUALITY was the most important predictor in the forest (Figure 3.16). Straight speakers produced the longest estimated duration (approximately 90 ms), followed by lesbian/gay speakers (approximately 66 ms), and finally bisexual speakers with the shortest estimated duration (approximately 58 ms). There is a relatively large difference between the duration estimates of straight speakers and the other groups. The difference between estimated duration is more moderate between lesbian/gay and bisexual speakers.

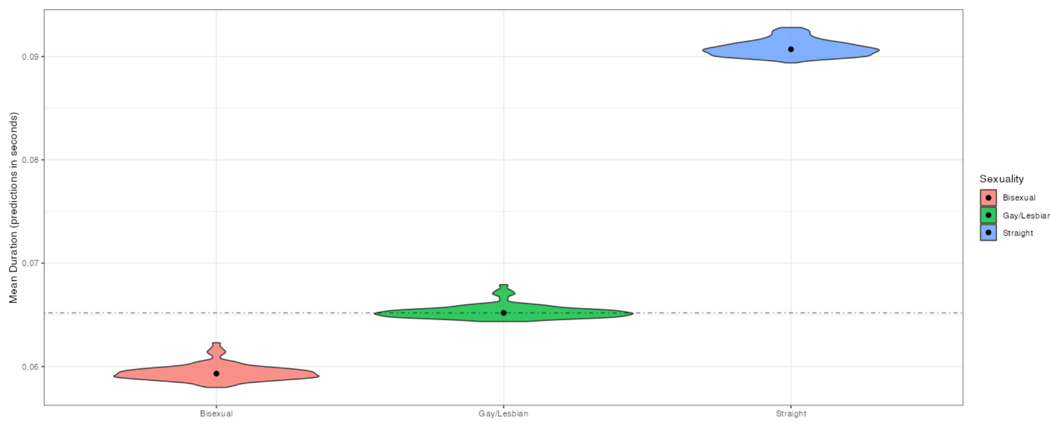

Figure 3.16: SEXUALITY partial dependence plot for duration

WORD was the second most important predictor of /s/ duration in the model. Most words are relatively balanced around the mean duration, with *superimposition* as a salient outlier (Figure 3.17). *Superimposition* likely behaves dissimilarly to the other words in the dataset for a few reasons. First, *superimposition* is the only six-syllable word in the data (most words are mono- or disyllabic). It is also the longest word in terms of number of segments. Second, the /s/ in *superimposition* is the only /s/ that receives secondary stress (as opposed to primary stress) in the sample. Nonetheless, we again chose not to remove outliers post-hoc to preserve the comparability of the three random forests, as discussed previously.

There are a few more things worth mentioning about *superimposition*. First, English [s] tend to be shorter in multisyllabic words (Klatt 1974), yet the initial [s] in *superimposition* here has substantially longer estimates than all other words in the sample. Second, English [s] tend to be longest before stressed vowels (Klatt 1974), but the initial [s] in *superimposition* only receives secondary stress, such that one might expect it to be shorter than the other /s/ tokens on the basis

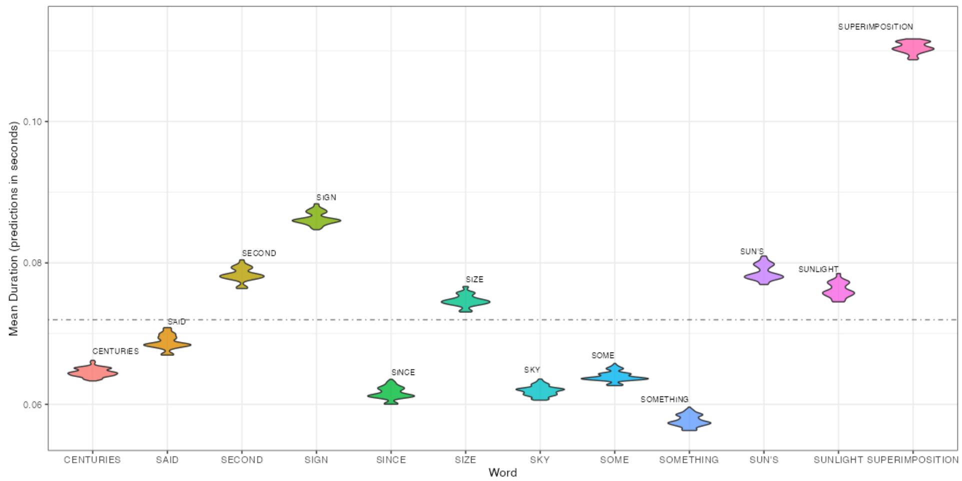

Figure 3.17: WORD partial dependence plot for duration

of stress as well. These anomalies are possibly attributed to word frequency. *Superimposition* is less frequent and used in more specialized contexts than the other words in the sample. Speakers may have produced this word more carefully than the others because of lack of familiarity, which may have affected duration. *Superimposition* also triggered more disfluencies than all the other words combined.

3.4.4 Summary of results

Prior to the discussion, we briefly summarize the results presented in the previous section in Table 3.3. For COG, GENDER was the most important predictor in the random forest, followed by RACE, SEXUALITY, and REGION, in that order. GENDER participated in every significant interaction in the COG random forest. Despite their lack of importance vis-à-vis VIMP scores, the gender stereotypicality ratings are included in important interactions in the COG model. As for skew, the most important predictors in terms of VIMP scores were SPEAKER and RACE. Only the interaction between GENDER and RACE passed the importance threshold in the skew random forest. Finally, the duration random forest included no interactions that passed the importance threshold. The two most important factors in the duration model were SEXUALITY and WORD.

Measure	Important predictors (most to least)	Important interactions (most to least)
COG	GENDER RACE SEXUALITY REGION	GENDER:RACE GENDER: MASCULINITY RATING GENDER:FEMININITY RATING GENDER:SEXUALITY
Skew	SPEAKER RACE	GENDER:RACE
Duration	SEXUALITY WORD	none

Table 3.3: Summary of key results

3.5 Discussion

The results for both duration and COG indicate that bisexual speakers do not reliably pattern with lesbian/gay speakers or straight speakers. These findings call into question approaches that exclude bisexuality or group bisexual speakers with other participants. Conflating bisexual speakers with lesbian/gay speakers a priori is a relatively common practice in experimental production studies (e.g. Munson & Babel 2007; Munson et al. 2006b, 2006a; though see Pierrehumbert et al. 2004 who couch their conflations in data exploration). Bisexual erasure or exclusion is also prevalent in perception studies that elicit sexuality judgements. Two of the most common elicitation methods, the Likert scale paradigm (e.g. Munson et al. 2006a) and the forced choice paradigm (e.g. Smyth et al. 2003), render bisexuality either ambiguous or unavailable, respectively (Willis forthcoming; Chapter 2). The widespread and uncritical engagement with these research practices is indicative of the pervasiveness of hegemonic monosexist ideologies in experimental studies of sexuality and the voice in English. We recommend that future research operationalize bisexuality as a distinct category unless data exploration, local epistemologies of sexuality, and/or theorization justify otherwise.

The salience of GENDER (in the COG random forest in particular) is in keeping with previous research on gender-based differences in /s/ production (e.g. Fuchs & Toda 2010; Munson 2007).

What is noteworthy, however, is the interactions between GENDER and the gender stereotypicality ratings. The interplay between these factors in this dataset is elaborated upon elsewhere (Willis forthcoming; Chapter 2), but we briefly discuss a few key points here. First, the importance of both gender stereotypicality ratings in the COG random forest suggests that the ways these participants engage with /s/-fronting is not solely correlated with gender assignment or identity, but also with gender expression and orientations to gender normativity. Second, the direction of the effect is not what one might expect based on the associations between /s/-fronting and femininity/masculinity reported elsewhere (Figure 3.8, Figure 3.9). Fronted /s/ is typically associated with femininity whereas retracted /s/ is typically associated with masculinity. However, we find that the more masculine our women participants rated themselves, the more fronted their predictions were. We also found that the more feminine our women participants rated themselves, the more fronted the predictions were. That is, the more gendered a woman rated herself, the more fronted her predictions were. We did not find this pattern among men in our sample; men who rated themselves as more masculine produced more retracted estimates, and men who rated themselves as more feminine produced more fronted estimates. However, /s/-frontedness appears to correlate with larger distinctions in self-reported femininity than masculinity among these men. In other words, differences in /s/-frontedness are correlated with more salient distinctions in femininity than masculinity, suggesting that the former is more marked for men in this study. Taken together, these results suggest that the relationship between /s/-frontedness and gender expression is not a straightforward, dualistic contrast between front/feminine and back/masculine for all speakers (see Chapter 2 for an in-depth qualitative analysis of this point). The implications of these findings for the indexical field of /s/ variation is beyond the scope of this chapter, but the complex ways in which speakers engage with these meanings is discussed at length elsewhere (Calder 2019; Campbell-Kibler 2011; Zimman 2017).

Next, we turn our attention to race and its intersections with gender. First, however, we discuss two important caveats. First, the way ethnoracial categories were elicited likely primed a particular way of thinking about race. Participants were presented a list of terms (e.g. Asian, Latinx, etc.) as well as a free response section. They were instructed to circle one or more of the provided categories, write their own answer, or elaborate in the free response section, or to combine multiple options. This operationalization of distinct, discrete racial categories and their combinations may not reflect the way participants understand race or experience racialization in everyday life. Given that many of the participants are unknown to us outside the research context, we are unable to assess the extent to which the presented operationalization captures participants' understandings and experiences of race and racialization.

The second caveat ultimately boils down to our small sample size. Above, we argue that random forests are particularly well-suited for handling the proverbial "small *N* large *P*" problem, that is analyzing datasets with a small number of participants relative to the number of predictors, as well as low cell counts. Our data is an extreme version of these two issues; there are only three times as many participants as there are predictors, and there are some cells with no counts. That being the case, it is possible that (some of) the variation reported here reflects individual differences. For example, it is impossible to determine whether the differences vis-à-vis race and gender for the lone Latinx/Pacific Islander participant are idiosyncratic to that individual or are indicative of a broader pattern within that community. Alternatively, the variation we find among ethnoracial groups could be related to some other correlated patterned difference, such as multilingualism or class. To tease apart multilingualism and/or class from ethnoracial identity, those factors need to be included as predictors in the models. Adding these predictors would further exacerbate the "small *N* large *P*" issue with our data, and we lack the information to incorporate these factors regardless. All of this to say, we urge readers to interpret the discussion with caution. Random forests are powerful, and reveal nuances that the regression analysis in the previous chapter could not, but they are not so powerful that these potential confounds can be straightforwardly dismissed.

With those caveats in mind, what can be tentatively said is that the dimension labeled as self-reported ethnoracial distinctions appears to be highly relevant for predicting variation in /s/ skew and COG among this group of participants. The GENDER:RACE interaction for both variables (Figure 3.7, Figure 3.14) suggest two broad patterns. First, members of the same gender do not have similar estimates across ethnoracial categories. That is, the mean and range of COG and skew estimates for a particular gender is not consistent across ethnoracial groups. For example, Latinx women produce much higher COG estimates than Asian women, who produce much higher estimates than white women, etc. Second, ethnoracial groups vary in terms of the direction of difference between women and men. It is typically assumed that women produce /s/ with a more negative skew than men (Fox & Nissen 2005; Jongman et al. 2000). The results for most ethnoracial groups in our sample are consistent with this generalization. However, Asian and multiracial Latinx/white women produced /s/ with slightly more positive skew than their male counterparts. Moreover, ethnoracial groups vary substantially in terms of the degree of difference between women and men. For example, the estimated mean COG differs almost 2.5 kHz between monoracial Latinx women and men, but only about 500 Hz between white women and men. Wider variability between women and men for some measures but not others suggests that ethnoracial groups may achieve gendered distinctions in /s/ production through different means. Why one measure of /s/ is used over another is not yet clear, but in any case these results demonstrate that previous research on white or racially unspecified speakers (which we elaborate on below) may not be easily generalizable across ethnoracial communities. Indeed, this conclusion is not a novel one. Calder (2021) demonstrates that Black speakers in Bakersfield, California do not employ /s/ variation the way that is expected by the white canon. In fact, the indexical field of gender-based variation is inextricable from whiteness and is not incorporated into performances of local Black identity. In short, generalizations about gender differences in /s/ production are not consistent across ethnoracial groups in our (small) sample for reasons that are inaccessible without deeper ethnographic understanding. All of this to say, our intention with this analysis is not to provide strong or generalizable empirical claims about the relationship between /s/, race, and gender. Indeed, to do so with such a small participant group is inadvisable even when taking the power of random forests into account. Rather, we aim to provide preliminary evidence that supports future research accounting for the ways /s/ variation is used to produce bisexual identities in different queer communities across race and gender.

Finally, our results demonstrate the degree of impact the measure of /s/ has on the outcome of the analysis. While some studies of /s/ consider multiple measures (Linville 1998; Munson et al. 2006b, 2006a), many focus only on COG (Hazenberg 2015; Podesva & Hofwegen 2014; Zimman 2013). Our findings across the three measures we examined–COG, skew, and duration–are radically different from one another. Yet, one might expect an interaction between these measures. COG and skew are both acoustic correlates of /s/ frontedness, and duration conditions COG, such that longer /s/ tokens typically have higher COG among some groups (Calder & King 2020). A potential direction for future research is to explore how these measures of /s/ interact with each other (if at all) in terms of articulation, and how these interactions affect the perception of socially meaningful /s/ variation.

3.6 Conclusion

Throughout this chapter, we have demonstrated the efficacy of random forests as a possible methodological intervention to grapple with various difficulties many sociolinguists face in quantitative research (see also Tagliamonte & Baayen 2012). The application of random forests to this particular dataset replicates and enriches previous findings on bisexuality and /s/ production (Willis forthcoming, Chapter 2) by incorporating a broader set of social factors as predictors. Such an analysis is possible because random forests are well-equipped to analyze small, unbalanced, and non-random data structures with many dimensions of interest. Extending the toolbox of quantitative sociolinguists (and perhaps linguists in general) to include random forests and similar modeling techniques may direct analysts towards important factors that might otherwise be missed or neglected by more traditional hypothesis-testing models. In short, we encourage researchers to consider random forests when embarking on quantitative studies of language variation and change.

The methodological considerations of our analyses extend beyond random forests, however. They also highlight critical issues related to research practice. Although random forests allow for more nuanced representations of identity by incorporating more predictors, how these factors are quantified remains an issue. Eliciting the complex ways people experience identity and translating that complexity into discrete categories is no simple task. The way we elicited and operationalized ethnicity and race in this study, for example, shaped the extent to which we could reasonably interpret our results. A one-size-fits-all solution to this issue is unlikely to exist. Instead, we argue that transparency in how and to what end participants' intersectional identities are elicited and operationalized is necessary. Previous work is not always forthcoming about aspects of their participants' identities that are not specifically under investigation. Studies typically do not report the ethnic or racial identities of their participants, much less include these categories as factors in any ensuing statistical modeling. When ethnicity/race is reported, it is usually to say that a small number of participants identify as non-white (Gaudio 1994), the implication being that speakers identify as white unless otherwise noted. The results regarding race in particular raise questions about the generalizability of previous work, which is largely based on white or racially unidentified speakers. Indeed, comparison between studies is tenuous at best when there is limited information about the similarity of distinct groups of participants. Going forward, we urge researchers to be transparent about the various positionalities their participants occupy as well as how those identities were elicited and operationalized, even when those categories are not the focus of the analysis. Transparency and attention to research practice are vital steps towards combating the influence of monosexism and white privilege in studies of sexuality and the voice.

Chapter 4

Happy, *#horny*, *valid*: A keyness analysis of bisexual discourses on Twitter

4.1 Introduction

What makes a Tweet #bisexual? A #bisexual Tweet might be one that talks about bisexual-identified people and their experiences. It could potentially discuss biphobia, or monosexism[1], or where bisexual people fit in the broader queer and trans community. Alternatively, a #bisexual Tweet might invoke bisexuality using the hashtag to discuss something else entirely. In this chapter, we examine bisexual discourse on Twitter at a big-data scale. We refine the question "What makes a Tweet #bisexual?" into two specific research questions, namely:

- What are bisexual discourses on Twitter about?
- What is the communicative or interactional function of invoking bisexuality in a Tweet?

The primary goal of this work is to understand the concepts and ideologies that characterize discourses related to bisexuality on a broad scale. In doing so, we create an open access codebase

[1] *Monosexism* refers to the ideology or assumption that people are, or should be, attracted only to one gender (Eisner 2016).

for discursive research that uses data from Twitter or similar websites (Todd & Willis 2023).[2] Regardless, answering our research questions at a broad scale requires a quantitative approach. Therefore, we need a way of measuring or quantifying aboutness. For that we turn to *keyness*.

What is keyness? Imagine you are trying to find posts about mushroom foraging in your local area on Twitter. You might use query terms like “mushroom(s)”, “foraging”, “mushroom identification”, “mycology”, and the name of your local area in your search. You do not pick these terms at random; you select these terms because you believe that they characterize the target of your search. Keyness (also known as keywords analysis) reverses this scenario. This statistical method analyzes the frequency of words in a text to reveal what words best characterize that text. The resulting keywords are indicative of what the target text is about.

What makes a word key? A keyword is a word that occurs significantly more often than expected in the collection of target or study texts. The expected frequency of a word is calculated based on its frequency in a collection of reference texts of the same genre/register that represent language usage in general (see Section 4.3.1). Returning to our earlier Twitter search example, it is likely that words like “mushroom(s)”, “foraging”, “mushroom identification”, and “mycology” occur far more often in Tweets about mushroom foraging than they do in Tweets in general. On the other hand, it is unlikely that words with a largely grammatical function such as “the” or “and” occur more often in Tweets about mushroom foraging than Tweets in general. This example illustrates three important facts about keyness: (1) it is a comparative method, (2) it is based on word frequency, and (3) there are expectations about what types of words will be key. In this chapter, we use this method to reveal what Tweets related to bisexuality are about.

[2] Parts of the codebase we constructed for this project are specific to collecting and preparing Twitter data collected through the summer 2022 version of Twitter API academic access. Adjustments can be made to translate these parts of the codebase for interfacing with other websites and data structures. The remainder of the codebase is widely applicable to quantitative discourse analysis on large datasets.

4.2 Data

4.2.1 Collection

Given that keyness analysis is an inherently comparative method, we have a specific task regarding data collection. That is, we must collect two pools of data: a collection of Tweets related to bisexuality that constitutes our study corpus (the #Bi Twitter Corpus), and another collection of Tweets not related to bisexuality that constitute our reference corpus, to which we compare the study corpus (the General Twitter Corpus). To that end, we harvested Tweets from the Twitter API v2 using the twarc2 Python package. Both corpora contain only Tweets written in English that were posted in the United States between January 1, 2020 and June 28, 2022. Retweets and quotes were excluded to avoid artificially inflating the counts of a repeatedly reposted or quoted Tweet.

The #Bi Twitter Corpus is a collection of Tweets related to bisexuality. Relatedness is operationalized as a given Tweet containing one or more of a predetermined set of hashtags referring to bisexual lived experiences, communities, or issues (1-7). These hashtags were selected from a larger list of potential hashtags compiled based on the authors' experiences navigating Twitter.[3] The hashtags in (1-7) were selected from the larger list due to their productivity; a search with each hashtag produced at least 1000 Tweets per year between 2020 and 2022.[4]

1. #bi
2. #bi pride
3. #bisexuality

[3]Willis identifies as bisexual and is an active member of multiple online bisexual communities. Much of her research is concerned with the implications of bisexuality for linguistic theories and research practice (Willis 2021; Willis & Ben Youssef 2023). Todd also identifies as bisexual and acts as a bisexual student mentor at UC Santa Barbara.

[4]Hashtags are represented with spaces in between words for readability.

4. #bisexual pride

5. #bisexuals

6. #bi visibility day

7. #bi week

Manual inspection of test pulls identified two types of irrelevant Tweets (i.e. Tweets unrelated to bisexuality) included in the data harvested using #bi. The first type was Tweets related to business intelligence (frequently abbreviated as BI). The second type was Tweets related to the Korean rapper Hanbin, who uses the moniker b.i. These Tweets were filtered out of the final study corpus harvest using the manually-identified hashtags, keywords, and usernames in Table 4.1 as exclusion criteria.

Hashtags	**Keywords**	**Users**
#analysis	analytics	@shxx131bi131
#analytics	analysis	
#business analytics	hanbin	
#business intelligence	Microsoft	
#hanbin	query	
#microsoft		
#power bi		

Table 4.1: Exclusion terms based on manual inspection of preliminary data

The General Twitter Corpus is a collection of Tweets that “match” the #Bi Twitter Corpus. That is, these Tweets fulfill the same language, location, and post-type criteria (i.e. not quotes or retweets), and they have the same distribution over the target time period. To replicate the distribution of the #Bi Twitter Corpus Tweets across time, all Tweets in the #Bi Twitter Corpus were binned according to the hour they were posted, and then an equivalent number of tweets not mentioning the query terms (1-7) were harvested from that 1-hour time period. For example, if the study corpus had 1000 Tweets posted on January 1, 2020 from 11-12pm, we harvested

1000 reference Tweets posted on January 1, 2020 from 11-12pm. Since Twitter API v2 harvests the most recent Tweets first, Tweets in each time bin in the reference corpus were usually posted towards the end of the hour (e.g. 11:59:59). The result is that each study Tweet has a reference Tweet that "matches" the time when it was posted down to the hour, but not to the minute. Pairing Tweets in this way ensures that the study and reference corpora are maximally similar to each other in terms of the distribution of Tweets over time. Long-term (e.g. days, months, years) temporal matching ensures that both corpora capture the same large-scale events (e.g. the COVID-19 pandemic, the January 6 2021 insurrection, etc.). Short-term (e.g. hours) temporal matching ensures that both corpora capture the same user groups (e.g. people who post at the morning versus at night, people in the same time zone, etc.).

Tweets harvested with the Twitter API contain rich metadata. We use metadata about the poster's username and any embedded media (images and videos) in our data filtering pipeline, as described in Section 4.2.2.

4.2.2 Cleaning

Harvested Tweets were normalized for data exploration and analysis. Our normalization process included: lowercasing all text to regularize capitalization; removing links and media (which are separately captured through metadata); splitting up sequences of emoji; separating words based on whitespace and punctuation (except for word-internal punctuation such as apostrophes and hyphens); and stripping word-external punctuation. Normalization ensures that small differences in how the same word is written do not affect its count, an essential step for keyness analysis which relies heavily on such counts.

4.2.3 Filtering

Manual inspection of the harvested Tweets identified the need for two layers of filtering. The first layer of filtering was done at the user level to remove bots or users engaging in excessive bot-like behavior. We define bot-like behavior as the overuse of formulaic or template-like Tweets. Such Tweets may be the result of repeatedly posting the same Tweet verbatim or by repeatedly posting Tweets that fit a template. Many non-bot (i.e. human) users who fit this criteria were in fact using automation. For example, many of these users were using automation to generate posts about items for sale on another website (e.g. Etsy) or to generate posts from other social media websites (e.g. Instagram).

To that end, we first used the bot classifier from the R package Tweetbotornot2 (Kearney 2020) as a first pass to determine which of the users in both corpora were bots. This classifier categorizes a given user as a bot (or not) by evaluating the latest (up to) 200 posts on their Twitter timeline based on certain criteria, including whether the user states they are a bot in their profile, number of followers, retweets, quotes, etc.[5] The classifier then generates a probability score between 0 (definitely not a bot) and 1 (definitely a bot). Based on manual inspection of the probability scores, we elected to exclude users with a Tweetbotornot2 score of 0.9 or higher from the dataset (about 2% of all users).

Next, we used the timeline data harvested to generate Tweetbotornot2 scores to calculate type:token ratios (TTR) for all users. TTR is calculated as the number of unique words (types) appearing in the (up to) 200 Tweets on a user's timeline, divided by the total number of (potentially non-unique) words (tokens) on the timeline; it represents the extent to which the words (tokens) posted by a user are all distinct from each other (i.e., tokens of different types). Users who repeatedly post the same Tweet verbatim, or who post highly formulaic Tweets, have a low TTR.

[5]The expectation regarding number of followers is that bots are likely to be followed by many users but unlikely to follow other users, whereas human users are more likely to have a relatively more even balance of followers versus following accounts.

Using TTR as an exclusion criteria caught bot-like users who were not flagged by Tweetbotornot2. When calculating the ratio, links to media files such as photos, videos, and gifs were conflated into a single type, "MEDIA", and links to websites were similarly conflated into a single type, "LINKS". These Tweets were otherwise normalized in the same way as the Tweets harvested for our corpora, as described previously. Descriptive statistics of the TTR results are in Table 4.2.

	Types	**Tokens**	**Ratio**	**Timeline Tweets**
mean	111	3220	0.37	199
sd	428	1451	0.10	1.1
min	1	200	0.0003	196
25%	818	2141	0.032	199
50%	1076	2927	0.38	200
75%	1383	4039	0.43	200
max	4025	21233	1.00	200

Table 4.2: Descriptive statistics for type:token ratio (TTR) for all users

Based on exploration of the ratio results, we elected to exclude users with a TTR less than or equal to 0.2 (about 6.6% of all users in both corpora). The overall distribution of ratios is skewed high among our users Figure (4.1) compared to other studies, likely due to the influence of text length on the ratio. The longer a text is, the more likely it is that each word has already occurred (Covington & McFall 2010). Tweets are very short, limited to 280 characters, so it is not unexpected for their TTR to be skewed higher than longer texts. Manual inspection of the results confirmed that 0.2 is an appropriate cutoff point for our use case, and a 0.2 TTR has also been the discrimination point to distinguish bots from humans in previous work (Ikae et al. 2019).

The process of calculating TTRs also provided us with data on user activity. We found that users in the upper three quartiles had 196 Tweets or more on their timelines. We chose to exclude users in the lowest quartile (i.e. users who had posted less than 196 Tweets during their entire time using Twitter). Users who post so little may engage with discourses on Twitter in other ways (for example, by liking others' Tweets or by retweeting), but they are not necessarily a part

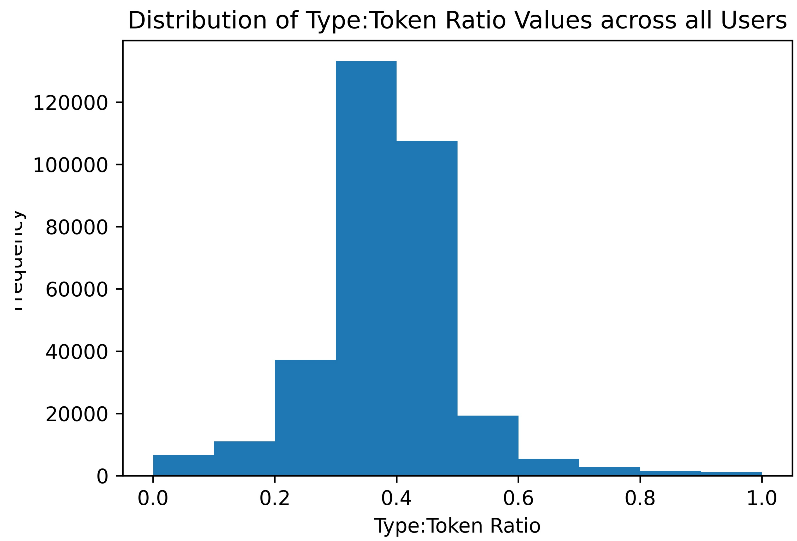

Figure 4.1: Distribution of TTR values across users

of the community of users who actively use these terms and therefore might use or understand these terms in distinct ways.

The second layer of filtering was done at the Tweet level. In manually inspecting Tweets collected for the study corpus, we found that #bi, #bisexuals, #bisexuality, and #bisexualpride produced many "not safe for work" (nsfw) Tweets; i.e., Tweets that were pornographic or sexually explicit. In fact, Tweets of a sexually explicit nature vastly outnumber those that are not; the most frequent word in the #bi Tweets is *#horny*. We decided that we wanted to do two analyses: one that included all of the posts regardless of their sexually explicit nature and another to examine "safe for work" (sfw) or non-explicit discourses about bisexuality on Twitter. To that end, we needed ways to separate posts that were nsfw from sfw ones.

Manual inspection of the data indicated that many nsfw Tweets include media files, such as sexually explicit photos, gifs, or videos. We used the NSFW Detection Machine Learning Model to identify posts with sexually explicit images (Laborde 2021). For each image, the model produces a probability distribution over five classes: (1) sfw drawings, (2) hentai (sexually explicit or pornographic art drawn in the style of Japanese animation) or a pornographic drawing,

(3) neutral images, porn, and (4) "sexy" images (i.e. images that are sexually suggestive but not pornographic). The probability distribution indicates which of the five classes best fits the image. Based on manual inspection and quantitative exploration of the results, we elected to exclude posts containing images that had a probability of 0.6 or higher in any one of the following categories: pornographic, hentai, or sexy.

In addition, we used a list of stopwords to filter out nsfw posts. We considered as potential stopwords all 5567 words and emoji that occurred at least 100 times in the nsfw subset of the study corpus (i.e. Tweets harvested from the #bi, #bisexuals, #bisexuality, and #bisexual pride queries). Restricting attention to these frequenct words limits the influence of sampling error due to data sparsity.[6] To determine which of these potential stopwords were highly indicative of nsfw content, we calculated a nsfw:sfw ratio for each word, by dividing the raw frequency of the word in the nsfw subset by the raw frequency of the word in the sfw subset. Any word for which the nsfw:sfw ratio was higher than that of #bi (the query term with the highest ratio), or for which the ratio was undefined (meaning that the word occurred at least 100 times in the nsfw subset, but never occurred in the sfw subset), was treated as highly indicative of nsfw content and consequently added to a list of stopwords. The cutoff ratio defined in this way was 224; thus, the stopword list comprised 1087 words that occurred in the nsfw subset more than 224 times more often than they occurred in the sfw subset. Any Tweet which included any of these words was excluded from the dataset during the second layer of filtering.

We applied both layers of filtering to both the study and reference corpora. In doing so, we paired the Tweets in each corpus by time period (the hour they were posted) and by filtering outcome (which filter layer, if any, they first failed). We took two steps to ensure that the number

[6] Sampling error refers to the discrepancy between a sample and the larger population it represents. Data sparsity refers to situations in which the available data is limited. The likelihood of sampling error potentially increases when the data is sparse. There is a greater chance that the sample is not truly representative of the broader population when there are fewer data points in the sample. Therefore, we set a baseline frequency for potential stopwords in order to reduce the chance that a word is included as a stopword (because of sampling error, data sparsity, or both) when it should not be.

of Tweets in each time period would remain matched after each layer of filtering. If there were more reference Tweets than study Tweets in a given time period that failed a particular layer of filtering, we replaced the excess reference Tweets with newly-harvested Tweets from the same time period that would not fail that layer of filtering. Conversely, if there were fewer reference Tweets than study Tweets in a given time period that failed a particular layer of filtering, we excluded additional reference Tweets that would otherwise not be filtered out. In this way, our pairing of study and reference Tweets precludes a situation in which a study Tweet that does not need to be removed at a particular layer of filtering is paired with a reference Tweet that needs to be removed by that layer of filtering.

The first filter reduced the original harvested data by about half, from 751,246 Tweets to 326,164 Tweets. That is, the user-filtered study corpus contains 163,082 Tweets and the corresponding reference corpus also contains 163,082 Tweets, paired based on the time of posting. The second layer of filtering reduced the corpora by more than half, from 326,164 Tweets to 117,566 Tweets. The Tweet-filtered study corpus contains 58,783 Tweets and the corresponding reference corpus also contains 58,783 Tweets, paired based on the time of posting. In the following sections, we analyze the two layers of filtering independently, resulting in two separate analyses that we bring together in the Discussion section.

4.3 Analysis

4.3.1 Calculating keyness: frequency

We calculated keyness using the “traditional” method, i.e. based on frequency lists and the log-likelihood ratio or G^2 statistic (Gries 2021; Rayson & Potts 2020). First, we compiled a frequency list, i.e. a list of each unique word and its frequency in each corpus (the study / #Bi Twitter Corpus, and the reference / General Twitter Corpus). For each word observed in at least

	Study count	**Reference count**	**Sum**
target word	a	b	a+b
other words	c	d	c+d
sum	a+c	b+d	N

Table 4.3: Schematic for calculating keyness for one word, adapted from Gries (2021)

one of the two corpora, we constructed the following frequency table:

In Table 4.3, a is the frequency of the target word in the study (#Bi Twitter) corpus and b is the frequency of the target word in the reference (General Twitter) corpus. The size of the study corpus (in number of words) is represented by $a + c$; likewise the size of the reference corpus is represented by $b + d$. We then calculated the expected frequencies for each cell from a to d based on the marginal frequencies (the row and column sums); Equation (4.1) illustrates the calculation of the expected value for a.

$$\hat{a} = \frac{(a + b) \times (a + c)}{N} \tag{4.1}$$

Finally, we calculated G^2 using Equation (4.2) (Dunning 1993; Gries 2021). We set the resultant keyness scores to be positive when the observed count in the study corpus was larger than the expected count ($a \gg \hat{a}$), and negative when the observed count was smaller than the expected count ($a \ll \hat{a}$). The absolute value of the keyness score indicates association strength; for example, a large positive keyness score indicates a strong association with the study corpus (relative to the reference corpus).

$$G^2 = 2\left(a \log \frac{a}{\hat{a}} + b \log \frac{b}{\hat{b}} + c \log \frac{c}{\hat{c}} + d \log \frac{d}{\hat{d}}\right) \tag{4.2}$$

This G^2 statistic is highly correlated with frequency, such that words with a high absolute keyness value tend to be highly frequent. However, lower frequency words can also be highly key if they are distributed unevenly across the two corpora.

This method produces a list of isolated words paired with a positive or negative keyness value. Visualizations of the quantitative keyness analysis are presented and discussed in Section 4.4.1 (user-filtered results) and Section 4.4.3 (Tweet-filtered results).

4.3.2 Qualitative interpretation

We use quantitative keyness analysis as a means to identify abstract patterns and to select Tweets that exemplify these patterns. The qualitative interpretation of the quantitative analysis addresses the two broad research questions discussed in the introduction, namely:

- What exactly are people talking about when they are discussing bisexuality?
- What is the interactional or communicative function of invoking bisexuality?

To that end, we examine the use of individual keywords and as well as commonly co-occurring sets of keywords in context. For individual words, we focus our analysis on the top 50 keywords at each layer of filtering. For sets of words, we use a custom function to identify the top 10 sets of two or more co-occurring keywords. We select keyword co-occurrence sets based on their heuristic value, focusing on sets that are representative of different parts of the data. Analyzing sets of keywords in addition to individual keywords reveals how these lexical items (1) interact in characterizing what discourses related to bisexuality are about and (2) work in concert to fulfill a particular interactional function. Regardless, we do not analyze any particular Tweet(s) too closely, as we lack the necessary ethnographic understanding of their posters for a close qualitative analysis of individual Tweets.

4.4 Results

4.4.1 User-filtered results

First, we present the results of a G^2 keyness analysis of the Tweets remaining after the first layer of filtering (see Section 4.2.3), i.e. the user-filtered data.

Figure 4.2 plots all words observed in one or more of the user-filtered corpora by their overall frequency and keyness value.[7] Words on the left side of the figure are key in the reference corpus (e.g. *the*, 😂, 🤣, 😭) and words on the right side of the figure are key in the study corpus (e.g. *#gay*, *#horny*, *happy*, *bi*, *me*). It is clear from this figure that there is a correlation between keyness and frequency, but that this correlation is not absolute.

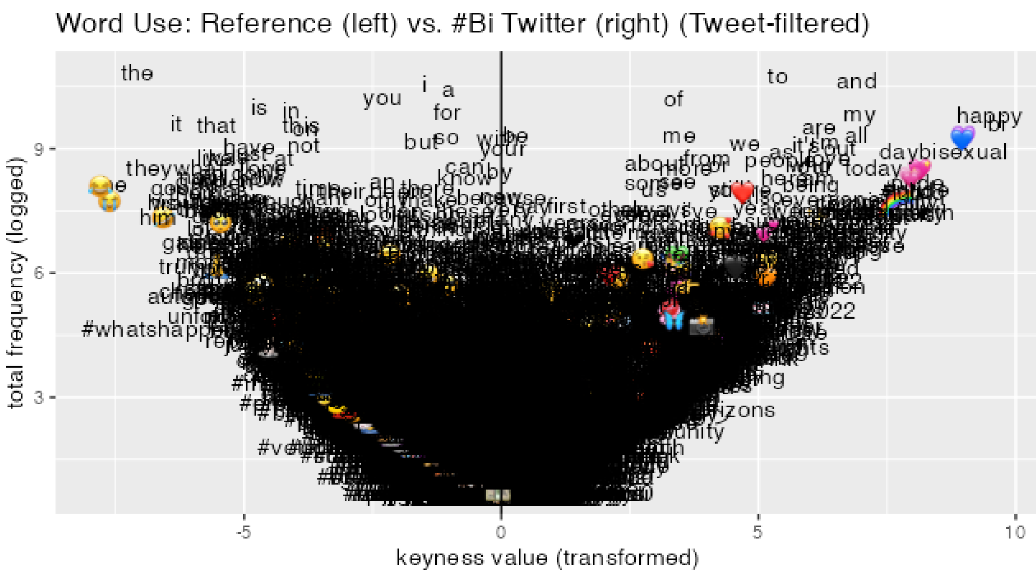

Figure 4.2: Relation between the total frequency and keyness of words in the user-filtered reference corpus (left) and study corpus (right)

Figure 4.3 is a word cloud of the top 50 keywords in the user-filtered study corpus.[8] In Figure 4.3 and all following word clouds, the size of the word represents its keyness value, such that

[7]Keyness scores in Figure 4.2 and Figure 4.5 have been altered for visualization with the following transformation: $sign(G) * log(G + 1)$.

[8]The Korean hashtag #비아이 ('biai') under the "h" in *#horny* refers to the Korean rapper b.i, demonstrating that the exclusion terms in Table 4.1 did not catch all posts about this unrelated topic.

large words are more key than small words. These keywords are consistent with our manual inspection of the preliminary data; nsfw or sexually explicit discourses dominate posts organized around bisexuality-related hashtags. There are several keywords related to genitalia (e.g. *#dick*, *#cock*), other sexualized body parts (e.g. *#ass*, *#tits*), and arousal (e.g. *#horny*). There are also many identity terms (e.g. *#straight*, *#trans*) and community terms (e.g. *#lgbt*).

Figure 4.3: User-filtered top 50 keywords (bi-related words included)

Figure 4.4 is a word cloud of the top 50 keywords in the user-filtered study corpus with terms that refer directly to bisexuality (e.g. *bi*, *bisexual*, *#bisexual*) removed. This word cloud excludes *bi*, *biexual*, and *#bisexual* and adds *#boobs*, *#dmme*, and *#sexting*.

4.4.2 Interim discussion: User-filtered results

Figures 4.2, 4.3, and 4.4 indicate that the user-filtered study corpus is dominated by sexually explicit discourses. Figures 4.3 and 4.4 clearly feature lexical items related to sex and sexual arousal (e.g. *#horny*, *#cum*), sexualized anatomy (e.g. *#cock*, *#pussy*, *#ass*), erotic personae (e.g. *#daddy*, *#slut*), and porngraphy or sexwork (e.g. *#porn*, *#onlyfans*) as salient keywords. There are also several keywords that reference identity labels (e.g. *#bisexual*, *#lesbian*, *#trans*, *#straight*, *#gay*). These identity labels are also often used as search categories or terms on

Figure 4.4: User-filtered top 50 keywords (bi-related words excluded)

pornographic websites, such that context is necessary to determine in what way these items are used in this dataset.

To that end, the sets of keywords that we selected to discuss based on the insight they provide into broader patterns of language in use are {*#gay*, *#straight*} and {*happy*, 💜, 💙}. Patterns in Tweets including the set {*#gay*, *#straight*} and individual keywords such as *#horny*, *#onlyfans*, and *#cumtribute*, reveal the salience of particular topics and associated interactional functions displayed in Table 4.4.

Table 4.4: Topics and interactional functions associated with {*#gay*, *#straight*} and other nsfw keywords

Beginning of Table 4.4

Topic(s)	**Interactional functions**	**Example**
Sex and desire	Soliciting sexual encounters	Any tops in CITY? #PLACENAME #PLACE **#gay** #bi #dl #top #bottom #bottomfortop #PLACE #bitop #bibottom #lowkey **#straight** #straightcuriousboys #Straightgay

Continuing Table 4.4		
Topic(s)	**Interactional functions**	**Example**
	Soliciting explicit content	I am the number 1 cock rater on Earth send dick in my dms #cockrate #dickrate #cawks #cockrates #hard #cock #dick #horny #nudes #dm #cocktribute #nsfw #nsfwtwt #hornyteen #rates #porn #cumtribute #dmme #cockdm #bi #cocks #bigdicks #sendplease **#gay** #bottom #BBC 😈
	Soliciting engagement with explicit content	YO WE ARE ABOUT TO HIT 30 SUBSCRIBERS, LIKE FOLLOW AND RETWEET FOR A CHANCE TO WIN. THE WINNER CAN CHOOSE TO EITHER CATCH THESE HANDS OR A DICK RATING (pussy is fine too) YOU DONT WANT TO MISS THIS CHANCE! #furry #furryfandom **#gay** #bi **#straight**
Sexwork and pornography	Promoting content related to pornography or sexwork	needing cash so anyone want to pay me to be a bad boy plz **#gay** #gayboy #boy #sugardaddylegit #gayabs #abs #muscle #gaymuscle #gaytrade **#straight** #gayPLACE #bi #big
Goods and services	Selling products and services	We specialize on Men grooming #wax #waxing #sugaring #sugarhairremoval #men #man **#gay** #PLACEgay #PLACE #PLACEgay **#straight** #dude #bi #brazilianwax #BrozilianWaxing
End of Table 4.4		

Tweets that are about sex and desire generally fall into three categories in terms of interactional goals: soliciting sexual encounters, soliciting explicit content, and soliciting engagement with explicit content. By “sexual encounters”, we refer to digitally-mediated interactions such as sexting or video chatting (e.g. *#sexting*, *dm*, *#dm*, *#dmme*) as well as in-person meetups.Tweets in this category are aimed at finding (typically short-term) sexual partners. The second category of Tweets, those that solicit explicit content, are characterized by the user asking other users to send them nudes, dick pics, pornography, or other media that is sexual in nature. For example, a user might post a Tweet asking others to send them pictures of their penises to rate them (e.g. *#dickrate*, *#cockrate*). The final function of Tweets in this topic is to solicit engagement with explicit content. This function differs from the previous in that, rather than the user asking for others to send them content, the user is instead asking others to engage with content they have posted. Again using dick pics as an example, a user might Tweet a picture of their penis and ask other users to rate it.

The next two groups of Tweets move from the personal to the professional sphere. Tweets organized around sexwork and pornography specifically are most often aimed at promoting the user's work on sites like *OnlyFans* (e.g. *#onlyfans*, *#porn*). These Tweets often include a titillating photograph, gif, or video as well as an invitation to explore the user's content on an external website. The final topic "goods and services" is very broad, but these Tweets too are aimed at promoting the user's work, ranging from apparel, to sex toys, to grooming services, to art commissions (*#art*).

Why is bisexuality invoked in these Tweets? A close analysis of {*#gay*, *#straight*} in context reveals that this pair frequently co-occur with *#bi*, one of our query terms, as well as the individual keywords mentioned previously. Invoking *#bi* alongside *#gay* and *#straight* is a search engine optimization strategy. A post that is simultaneously *#bi*, *#gay*, and *#straight* casts a wide net over possible search terms a Twitter user might draw on when searching for sexually explicit content. In this way, Twitter users can increase the possibility that their Tweet reaches its desired target(s), whether that target be a *#straight* or *#straight*curious (i.e. identified as straight but curious about same-gender sex) partner for a *#dl* ('down low', i.e. discrete) same-gender sexual encounter or a person looking for pornographic content featuring a variety of gender dynamics. The same strategy is applicable for advertising goods and services that are not necessarily related to sexwork or pornography. For example, a Tweet advertising hair removal services for men might include *#bi*, *#gay*, and *#straight* alongside hashtags like *#men* and *#BrozilianWaxing* (a portmanteau of the slang term "bro" and Brazilian waxing) to capture all potential customers.

In all of these cases, the construction and deployment of bisexuality is instrumental. Labeling a post as *#bi* is a means to an end, a way to find an intimate partner, sexting buddy, or potential customer. That is, the topics and interactional functions associated with bisexuality are not so much about bisexual lived experience or identity as they are about negotiating sexual availability or advertising. The association between these discourses and hashtags like *#bi*, *#gay*, and *#straight* is revealed by an interrogation of their combinatorial use in context; a surface-level investigation of these terms in isolation might mistakenly lead one to think that they are being used in the sense of identity terms, for example.

A very different set of discourses lies beneath those revealed by the set {*#gay*, *#straight*} and the higher ranked discrete keywords like *#horny* and *#nsfw*. The most common keyword co-occurrence set in the user-filtered data is {*happy*, '💜', '💙'}. Organized around temporally-situated events related to bisexuality (e.g. Bi Visibility Day, Bi Week, Pride Month, etc.), discourses associated with this set serve a variety of interactional functions. The most straight-forward of these functions is to express a greeting related to the event (e.g. Happy Bi Visibility Day). The targets of this greeting range from a specific person (e.g. ACCOUNT), a specific group of people (e.g. my bisexual friends), other bisexual people (e.g. my fellow bisexuals), or the bisexual community in general.

Many Tweets include only a greeting and hashtags, but others expand upon the greeting. In doing so, the function of the greeting changes from a simple well-wishing to a means to fulfill some other interactional function. One of these purposes is validation or authorization of bisexual existence and experience. Tweets serving this function often make connections between visibility and authorization (Bucholtz & Hall 2004) by linking validation with cultural intelligibility or "being seen" (e.g. we are valid and deserve to be seen). Other Tweets follow the greeting with an explicit contestation of bisexual stereotypes. Some frequently addressed stereotypes include the misconception that bisexual people are confused, that bisexuality is a phase, that bisexuality is transphobic or trans-exclusive, or that bisexuality does not exist. Contestations range from outright affirmation (e.g. bisexuals DO exist) to implicit problematization (e.g. POV: you assumed I'm straight 😡). Yet other posters take the opportunity to educate readers on issues related to bisexuality. Some common discussion items include the difference between sexuality and gender, how to be a good ally, and the idea that relationship status does not determine sexuality. The majority of these posts are ostensibly targeted at people who do not identify as bisexual, but may also be applicable to bisexual people who have different understandings of bisexual experience. These interactional functions are not mutually exclusive; many posts that address stereotypes are also educational and are categorized as such because they are explicitly framed as explanatory (e.g. "a friendly reminder"), for example. We categorize Tweets in this way for ease of explanation.

In these Tweets, bisexuality is not a tool or instrument but an identity under active negotiation. These Tweets point to a number of ways bisexuality can be experienced and practiced. For example, many Tweets emphasize that bisexual people vary in how they experience desire and attraction. Some bisexual people might be more attracted to one gender over others, whereas other bisexual people might be equally attracted to multiple genders. Many Tweets also specifically *authenticate* (Bucholtz & Hall 2004) or verify bisexual people who are not out, i.e. not publicly open about their bisexuality. Indeed, the one stance that is the most salient in its consistency is that all of the diverse ways of being bisexual are equally valid. Regardless, the high frequency of these identity-negotiating Tweets is instructive of bisexuality's status as an identity in constant need of explanation, correction, and specification. "Bisexual", with all its tenuousness and polysemy, then becomes a nexus around which Twitter users can find community and educate others on bisexual issues.

To summarize the findings related to the user-filtered data, we find three main topics: (1)

Topic(s)	**Interactional function(s)**	**Example**
Temporally-situated events related to bisexuality (e.g. bi visibility day)	Greetings or congratulations related to the event	Happy #BiVisbilityDay 💕💜💙
	Validation	#HappyBiVisibilityDay my sexuality is valid and i deserve to be seen! 💗💜💙
	Combating stereotypes	Happy #BiVisibilityDay! Not confused, not a phase
	Education	Happy #BiVisibilityDay! To be a good ally, it's important to call out biphobia and harmful stereotypes, don't assume people's sexual orientation and be open to hearing about others' experiences! Happy #BiWeek folks!

Table 4.5: Topics and interactional functions associated with {*happy*, 💜, 💙}

sex, desire, sexwork, and pornography or nsfw discourses, (2) temporally-situated events related to bisexuality, and (3) goods and services. These topics are related to a variety of interactional functions. In (1) and (3), bisexuality is a tool in the (sexual) marketplace, a means to locate consumers of dick pics and apparel alike. In contrast, bisexuality in (2) is constructed as an identity label that is deployed for community organization and educational purposes.

4.4.3 Tweet-filtered results

Here we present the results of a G^2 keyness analysis of the Tweets remaining after the both layers of filtering (see Section 4.2.3), i.e. the Tweet-filtered data.

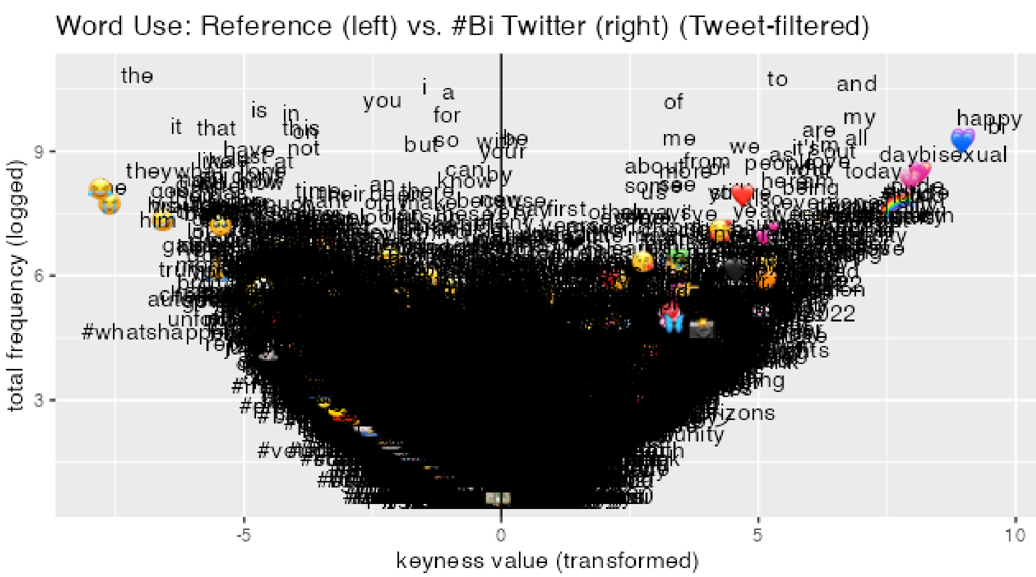

Figure 4.5: Relation between the total frequency (logged) and keyness of words in the Tweet-filtered reference corpus (left) and study corpus (right)

Similar to Figure 4.2, Figure 4.5 plots all words observed in one or more of the Tweet-filtered corpora by their overall frequency (logged) and keyness value (logged). Words on the left side of the figure are key in the reference corpus (e.g. *the*, *it*, *they*, 😂, 😭) and words on the right side of the figure are key in the study corpus (e.g. *happy*, *bi*, *bisexual*, *my*, *and*).

Figure 4.6 is a word cloud of the top 50 keywords in the Tweet-filtered study corpus. The keywords include those that were key in the user-filtered dataset (e.g. *happy*, 💜, 💙) alongside many new additions.

Figure 4.6: Tweet-filtered top 50 keywords (bi-related words included)

Figure 4.7 is a word cloud of the top 50 keywords in the Tweet-filtered study corpus with terms that refer directly to bisexuality (e.g. *bi*, *bisexual*, *#bisexual*) removed. The Figure 4.7 word cloud excludes *bi*, *bisexual*, *#bisexual*, *bisexuality*, *bisexuals*, and *#bivisibility* from Figure 4.6. The next keyword after these exclusions was *biphobia*, which we also chose to exclude as it is directly related to bisexuality, and instead include the following keyword reminder. This process also added keywords *am*, *are*, *loved*, ✨, and *pan* to the word cloud.

4.4.4 Interim discussion: Tweet-filtered results

Next, we discuss the Tweet-filtered data presented in Section 4.4.3 and in Figures 4.5, 4.6, and 4.7. Recall that this layer of filtering removes the sexually explicit nsfw discourses that dominate the user-filtered data in order to further investigate non-explicit bisexual discourses. Again, we use co-occurring keywords alongside individual keywords to guide this analysis.

Figures 4.6 and 4.7 suggest that the discourses about temporally-situated events associated with the co-occurrence set {*happy*, 💜, 💙} in the user-filtered data come to the forefront

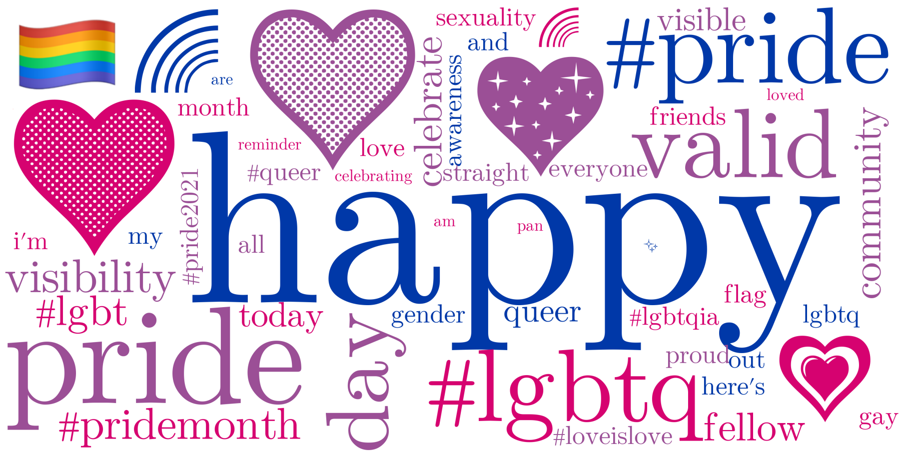

Figure 4.7: Tweet-filtered top 50 keywords (bi-related words excluded)

when further filtering is applied. "Happy" jettisons from 16th place to the top of the keywords list and is joined by other keywords related to temporally-situated events like *#pridemonth*, *#pride*, *pride*, *#pride2021*, *celebrate*, and *month*. Keywords such as *community*, *fellow*, *visibility*, and *awareness* also align with the previous analysis of the interactional functions of these well-wishing Tweets.

The Tweet-filtered keywords we find most intriguing, however, are the ones that are the most banal. Consider the keywords *i'm*, *my*, *all*, *are*, *am*, and *and*. These are all high frequency words in English that are used in a wide variety of contexts, and as such words of this type are not typically expected to be keywords. It is therefore surprising that there is a large difference in how often these words appear in the #Bi Twitter Corpus versus the reference corpus.

Co-occurrence sets are particularly useful for investigating *i'm*, *my*, *all*, *are*, *am*, and *and* due to the spread of their use throughout the study corpus. Examining co-occurrence sets containing two or more of these keywords, or alternatively their co-occurrence with keywords like *out*, reveals the ways these words collaborate in establishing the topics and interactional functions in Table 4.6.

Table 4.6: Topics and interactional functions associated with *i'm, my, all, are, am, and* and *out* and their co-occurring combinations, with keywords in bold

Beginning of Table 4.6		
Topic(s)	**Interactional functions**	**Example**
Coming out	Seeking advice on how to come out	Hey how can I come **out** To my parents without being to forward? #LGBTQIA2021 #Bi #comingout2021
	Using a Tweet to come out	#bipride This is **my** official coming **out** tweet!
	Discussing life after coming out	it's **my** first #BiVisibilityDay after being publicly **out** so... it's lit yo
Visibility	Calling attention to one's identity	It's #BiVisibilityDay, do your job **and** observe me 👁️👁️
	Acknowledging support	Thanks ACCOUNT for **all** you do to help men embrace #bisexuality **and** not feel shame over it! #Bisexualmenexist #biawarenessweek #bitwitter
Bisexual culture	Enregistering signs as bi culture	it's #BiVisibilityDay **and** no one told me!? gonna go cuff **my** pants real quick hold on
Non-monogamy	Sharing experiences	**W**ell my #DnD session was canceled but I just picked up a date with **my** bf **and** gf so I see it as a win! #bi #polyamorous
	Contesting naturalized relationship between bisexuality and polyamory and/or non-monogamy	What, wait? Oh nothing, just the @RELIGION saying #bisexuality is always wrong because it undermines monogamy (less than 30 years ago). **And** we wonder so many #bi people **my** age are closeted.
Denaturalizing stereotypes	Humor	I know it's #BiVisibilityDay **and all** but being actually invisible would be kinda fucking sweet

Continuing Table 4.6		
Topic(s)	**Interactional functions**	**Example**
	Education	@ACCOUNT "#bisexuality includes people of any/**all**/no genders **and** isn't binary or only liking men and women! it also doesn't mean someone likes different genders 50/50" – 💯THIS Please remember: We are the "B" in #LGBTQProud #BiPride #BisexualMenExist
Struggling with bi identity	Sharing experiences to encourage acceptance	For years, I didn't know I was allowed to "like both." I hated myself for even HAVING those feelings. Sometimes I was like "If I just WEREN'T ALIVE to feel these feelings it would **all** be better!" But I was wrong. It's so much better to be alive **and** to feel. #BiVisibilityDay
Goods and services	Selling products and services	#BiVisibilityDay Wassup chucklefucks check **my** art **out!** [images]
End of Table 4.6		

In addition to the discourses about temporally-situated events found in the user-filtered data, we find a number of topics related to bisexual identity and lived experience. First there are a group of Tweets that are all about coming out. These Tweets can be subdivided based on their interactional function: seeking advice about coming out, using a Tweet to come out, or discussing life after coming out. Tweets fulfilling the first function generally involve the user asking for guidance about coming out to specific people or about anxieties surrounding coming out. Those in the second category are essentially performing the speech act of coming out publicly. Tweets in the last category discuss life experiences that are demarcated as being after coming out. These Tweets all highlight the importance of out status among members in this community, even though being out is not necessary for one's bisexual identity to be authenticated by others per the preceding discussion.

The next topic is concerned with bisexual visibility. Many of these Tweets name events like Bi Visibility Day or other temporally-specific events, similar to the Tweets identified with {*happy*, 💜, 💙} in the user-filtered data, but are distinct in their interactional goals. The first function is to explicitly call attention to one's bisexual identity, typically by calling on others to "observe", "witness", or "see" the poster. Again, there is a connection between "being seen" and authorization. What distinguishes these Tweets from those like "we are valid and deserve to be seen" is that these Tweets often include an imperative. Rather than asserting that bisexual identity has the right to cultural intelligibility, these users instead demand it: "do your job and observe me". The use of imperatives places the writer in a position of power from which they exercise agency over authorization by others. The second function draws on a slightly different sense of visibility. These Tweets also tend to be associated with Bi Visibility Day or some other celebratory time period and use the event as an opportunity to thank or acknowledge a particular person or group, often bi content creators, for their contributions to the bi community. The acknowledged contributions range from representation (as in, the thanked person/people identify as bi or include bi people in their content) to advocacy work (as in, that of an official organization).

Bisexual internet culture is another salient topic in this dataset. Using *#BiVisibilityDay* or a similar hashtag to introduce the topic, Tweets in this category *enregister* (Agha 2003) or indexically associate specific signs as bisexual cultural artifacts. Signs range from the material (e.g. cuffed jeans) to media (e.g. The Mummy (Sommers 1999), Atlantis: The Lost Empire (Trousdale & Wise 2001)). This thread of bisexual discourses on Twitter invokes largely the same signs and strategies described by Willis and Fine (in progress) in their work on bisexual discourses on Tumblr. In short, these Tweets contribute to discourses of bisexual visibility through their overt efforts to establish a culturally intelligible bisexual identity.

The next group of Tweets is concerned with non-monogamy. We use the term non-monogamy to refer to a variety of practices in romantic/sexual relationships, ranging from polyamory to swinging. The specific non-monogamous practice invoked is dependent on the Tweet. The first function of Tweets in this topic is to share bi-identified people's experiences or desires relating to non-monogamy. For example, users may use hashtags like *#bi* and *#polyamorous* when expressing satisfaction with their romantic relationships, whereas others may use hashtags like *#bi* and *#swinging* when discussing a desire to add a third person to their bedroom. Expressing the joy that non-monogamous relationship dynamics brings to these posters (and their partners) is radical, given the negative way in which bisexuality and non-monogamy are portrayed in many contexts. The idea that bisexual people struggle with or are incapable of monogamy or are more likely to commit infidelity is a common negative stereotype. The other interactional function of Tweets in this category is to condemn this stereotype. Specifically, users reject the naturalized relationship between bisexuality and non-monogamy, but do not necessarily disapprove of non-monogamous practices.

In a similar vein, Tweets in the next category topicalize a variety of bisexual stereotypes, including but also beyond those related to non-monogamy. Given that the purpose of many of these Tweets is to dispel misconceptions about bisexuality, many of them are similar to the educational Tweets described in the discussion of the user-filtered data. What the additional layer of filtering surfaces is a group of Tweets deployed for the sake of humor. These Tweets draw on bisexual stereotypes for a humorous effect; for example, a Tweet like "bi, shy, and ready to cry" is funny based on an implied contrast between an emotionally vulnerable bisexual persona with the idea that bisexual people are suave and hypersexual (Willis & Fine, in progress). Tweets fulfilling this function reclaim bisexual stereotypes through humor as a form of resistance.

Emotional vulnerability is not relegated entirely to humor, however. Many Tweets in the #Bi Twitter Corpus share bisexual people's struggles related to their sexual identity. These Tweets often describe users overcoming feelings of anxiety, rejection, or fear or coming to terms with biphobic experiences. Rather than giving advice, these Tweets share these experiences to model an emotional journey with bisexual acceptance and joy as the conclusion.

Finally, there remains a large contingent of Tweets that invoke bisexuality to sell goods or services. The second layer of filtering brings to the forefront goods and services that are more transparently related to bisexual identity. These Tweets differ from the similar Tweets in the user-filtered corpus in that they tend to be targeted at bisexual-identified people specifically, rather than invoking bisexuality alongside a number of other identity labels to maximize visibility. Many of the products promoted in this category use bisexual representation to appeal to their target base; a user may promote their book on the basis that it includes bisexual characters, for example. Alternatively, users may promote their product or services using identity marketing strategies. That is, content creators signal their affiliation with the bisexual community as a part of the appeal for their product or service.

Removing the nsfw discourses from the data brings to the forefront a more detailed picture of how bisexuality is constructed relative to individual identity. Tweets that construct bisexuality as an identity are much more prominent and diverse in the types of topics and functions with which they engage. Across these variable topics and functions, bisexuality is constructed relative to highly individualized and specific experiences and practices. Tweets that enregister signs as bi culture or discuss struggling with bi identity both draw upon this understanding, for example. Again, Tweets concerned with bisexual visibility and validation are prevalent, underscoring bisexuality's status as an identity in need of explanation and correction. However, there is also a contingent of Tweets in which bisexuality is doing identity work but its meaning is not necessarily contested or under negotiation, particularly among Tweets that share the poster's positive experiences with non-monogamy.

The second layer of filtering similarly brings forth more detail on how bisexuality is constructed as a community. Tweets seeking advice about coming out are the most transparent instances of community members seeking connection, but Tweets featuring phrases like my fellow bisexuals (all of which are keywords) also presuppose the existence of a bisexual community on Twitter. Given that Twitter is a global social media platform, it is likely that the communities most often invoked in this understanding are closer to an idealized or imagined community, as opposed to communities of practice. However, that is certainly not to say that (digitally) local bisexual communities of practice (Eckert 2006) do not exist on Twitter. The analysis presented here does not include information about how Twitter users interact with one another (through retweeting, likes, quotes, comments, mentions, etc.), but it is certainly possible to examine how users form smaller communities within the broader #Bi Twitter sphere. As such, attention to (digitally) localized communities or practice is outside the scope of this chapter, but is certainly an interesting direction for future research.

4.5 Discussion

Figure 4.8 summarizes our findings for both the user- and Tweet-filtered datasets. The user-filtered data is the outermost layer, in which bisexual discourses are related to a variety of different topics, from desire and sexual availability to sexwork and stereotypes. What unites these topics is the construction of bisexuality as instrumental. Whether the target is a sexting partner or consumer goods customer, in these Tweets bisexuality is invoked as a means to an end.

Within the user-filtered data lies the Tweet-filtered data. In this layer, bisexual discourses that are sfw become more salient; topics range from support and struggle, to internet culture, to relationships and coming out. Bisexuality is constructed in multiple ways within these Tweets. The first construction is at the level of the individual; bisexuality is understood and defined relative to individual actors' (hyper specific) affective experiences, practices, and sensibilities. The second construction is at the level of the community. Here "bisexual" is a label that encapsulates a range of gender dynamics, attraction styles, and relationship dynamics; this label is deployed for education and community organizing.

This diagram also visualizes the relationship between the two layers of analysis. The topics,

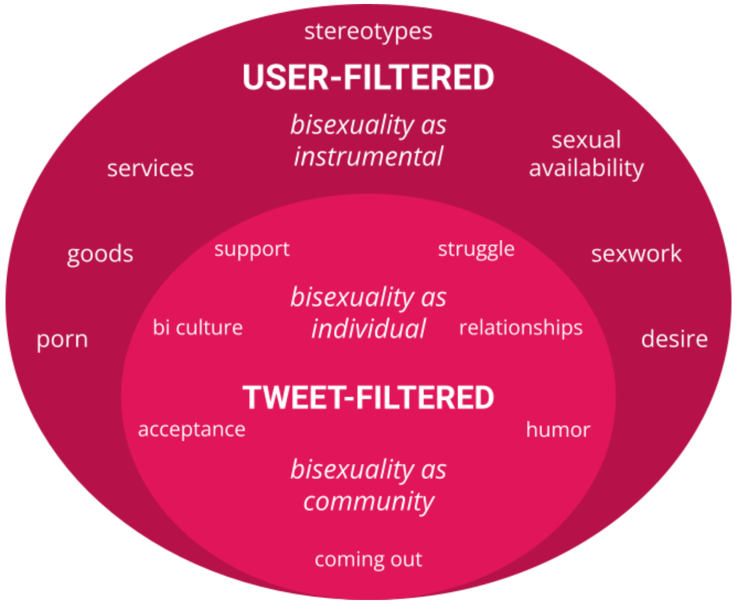

Figure 4.8: Diagram of the topics/functions (plain text) and constructions of bisexuality (italicized) in the user- and Tweet-filtered datasets

interactional functions, and constructions of bisexuality found in the Tweet-filtered data are also present within the user-filtered data (as the Tweet-filtered data is a subset of the user-filtered data). However, the patterns found in the Tweet-filtered data are obfuscated by the nsfw discourses that dominate the user-filtered data. This obfuscation is driven by the relationship between the "traditional" measure of keyness (i.e. G^2) and raw frequency. We expand more upon this connection in the next section.

Our analysis, particularly of the Tweet-filtered data, is compatible with the ways many sociolinguists and linguistic anthropologists theorize sexuality and identity. The construction of bisexuality relative to lived experiences, practices, and communities is not radically different from recent work on queer and trans identities in these fields. This perspective is also what a Twitter user sees when using the default settings. Twitter's default search settings block "sensitive content". Consequently, a user who searches "#bisexual" using the default settings sees a curated, sanitized view of how bisexuality is talked about on Twitter. The vast majority of discourses related to bisexuality are, due to their explicit nature, hidden from view, despite the fact that they are far more numerous (and therefore often more key). However, it is within these discourses that the most theoretically challenging work is happening because of the extent to which they highlight the link between bisexuality and sexual availability.

Indeed, it is a common negative stereotype that bisexual people are hypersexual and therefore more likely to participate in either ethical non-monogamy or unethical non-monogamy (i.e. cheating). Many Tweets that include hashtags like *#bi* when marketing pornography or other sexually explicit content capitalize on these stereotypes. Although there are many cases when it is unclear how the poster identifies, there are also many cases when posters who engage with these stereotypes explicitly identify as bisexual. There are a number of bisexual sexworkers and content makers who market their sexual orientation as a cornerstone of their appeal. These creators use hashtags like *#bi* as semiotic markers to signal the assets they bring to the sexual marketplace (Gershon 2011): content that features authentic same- and different-gender encounters, engages with hypersexual erotic personae (e.g. sluts), and/or caters to cheating, hotwife, or cuckhold/cuckqueen fetishes. While this behavior is arguably troubling because of how it engages with problematic stereotypes and self-commodification, it also ostensibly empowers these creators by affording them the material means to exist in a capitalist society that commodifies their labor regardless (Hall 1995).

Rather than using bisexuality as a means to connect with consumers or display market value, many Tweets in the nsfw dataset instead invoke bisexuality to find sexual partners. There are a number of gender dynamics engaged in these Tweets, but there are a disproportionate number of Tweets posted by men seeking men for clandestine in-person or digitally-mediated sexual encounters. Such Tweets typically use one or more hashtags containing the abbreviation dl (“down low”), a term originating from African American Language in the 1990s used to describe information or activities that should be kept secret (e.g. *#dlbros*, *#dltrade*, *#dltop*). The association between masculinity, bisexuality, and being dl in this dataset is unsurprising given the particular kinds of discrimination that bisexual men face. Bisexual men are often positioned as in denial or actually gay, and in turn are denied access to hegemonic masculinity (Alarie & Gaudet 2013; Brewster & Moradi 2010; Hertlein et al. 2016). As a result, many men who engage in both same- and different-gender relationships may opt to label themselves as dl rather than bisexual (or another plurisexual identity). Unpacking the complex relationship between masculinity, bisexuality, and being dl versus out warrants a closer examination of individual Tweets and their posters than we are able to provide in this chapter. What can be said is that there is a contingent of men on Twitter who use hashtags like *#bi* and *#dl* to navigate sexual availability without explicitly constructing themselves as bisexual.

The portion of the data that is "behind the curtain" of default content viewing settings suggests that the construction or invocation of bisexuality has a complex relationship with systems of power, marginalization, and commodification. The analysis presented here presents only the tip of the proverbial iceberg, and there is clearly much more important work to be done. What this chapter does contribute, then, is a foundation upon which future research can grow.

4.6 Next steps: Dispersion and alternative measures

Calculating keyness based on raw frequency, as we have done in this chapter, is the most common quantitative approach to keywords analysis. However, in recent years researchers have recognized that this measure is not ideal due to the correlation between G^2 and overall token frequency (Gries 2021). Improvements generally supplement raw frequency information by adding dispersion to the analysis (Baker 2004; Egbert & Biber 2019; Gries 2021). Given the salience of temporally-situated events in our data, we present a preliminary analysis of the keyness of keywords over time in this section.

Figure 4.9 visualizes the keyness of the top 50 keywords in the user-filtered dataset over time.[9] There are three particularly salient patterns in this analysis. The first pattern, exemplified

[9] In Figure 4.9 and Figure 4.10, the y-axis of each graph is re-scaled between its own maximum and minimum

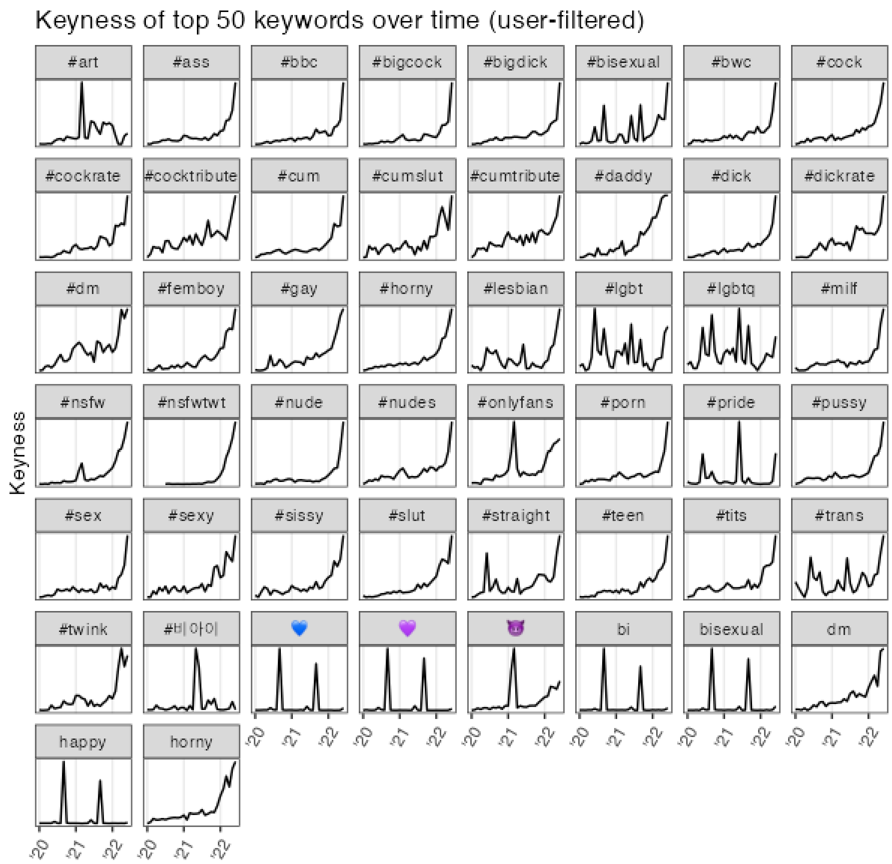

Figure 4.9: Keyness value (y-axis) over time (x-axis) of top 50 keywords (user-filtered)

by the nsfw keywords such as *#cock* and *#nsfw* is characterized by a general increase in keyness over time. The second pattern is characterized by spikes in September 2020 and September 2021; this is seen in keywords such as bi and bisexual. Finally, other queer and trans community terms such as *#lgbt* and *#lgbtq* see increases in keyness in June and September of 2020 and 2021.

Figure 4.10 plots the keyness of the top 50 keywords in the Tweet-filtered dataset over time. Many keywords see spikes in September 2020 and September 2021, i.e. during Bi Visibility Month. Similar to the user-filtered data, the keyness of queer and trans community umbrella terms like *#lgbtqia* increases in both June and September. Figures 4.9 and 4.10 again make clear the relationship between the user- and Tweet-filtered datasets; the trends found in the Tweet-filtered data are also present in the user-filtered data, but are less obvious due to the prevalence of nsfw discourses.

At first glance, one might interpret these patterns demonstrating a relationship between particular keywords and specific times. It seems logical that keywords related to bisexual pride events are more key in September than other time periods. The general increase in the keyness of nsfw keywords might then suggest that Twitter users are getting #hornier over time. However, the interpretation of these patterns is not as straightforward as it seems because of the relationship between time and corpus size. That is, what these patterns actually demonstrate is the impact of changes in corpus size over time on changes in keyness over time.

Figure 4.11 plots the size of both the user- and Tweet-filtered corpora over time. Notice the similarity to the patterns in Figures 4.9 and 4.10; there are small peaks in corpus size in June, large peaks in September, and a general increase in number of tokens per month over time. This demonstrates a clear relationship between corpus size, raw frequency, and association that needs to be disentangled. Willis & Todd (in progress) address these correlations by comparing and contrasting the G^2 analysis presented in this chapter with an analysis using Kullback-Leibler

keyness value.

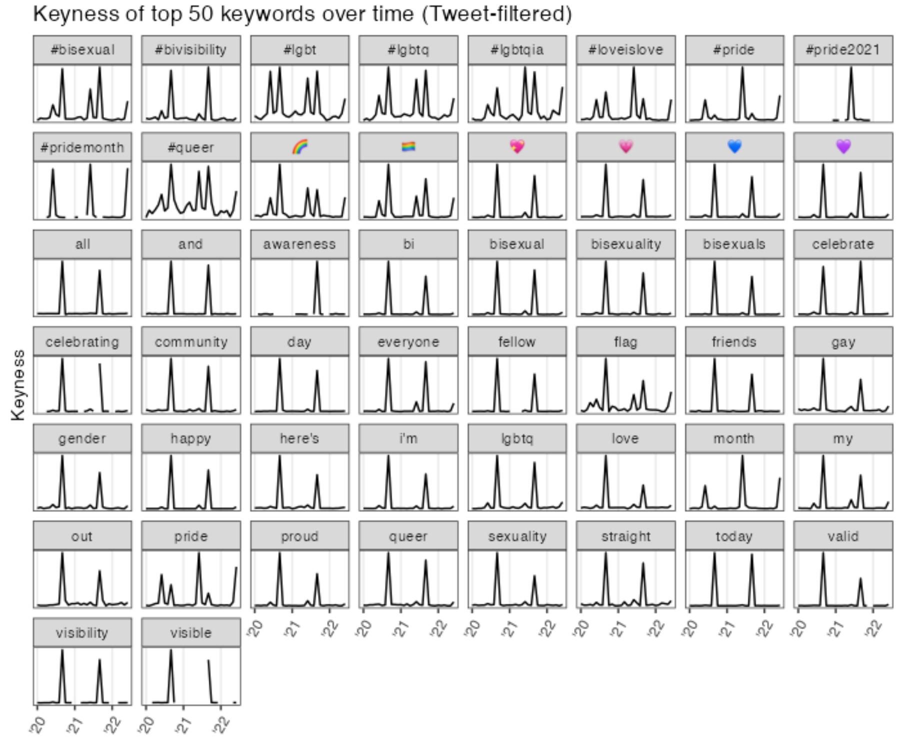

Figure 4.10: Keyness value (y-axis) over time (x-axis) of top 50 keywords (Tweet-filtered)

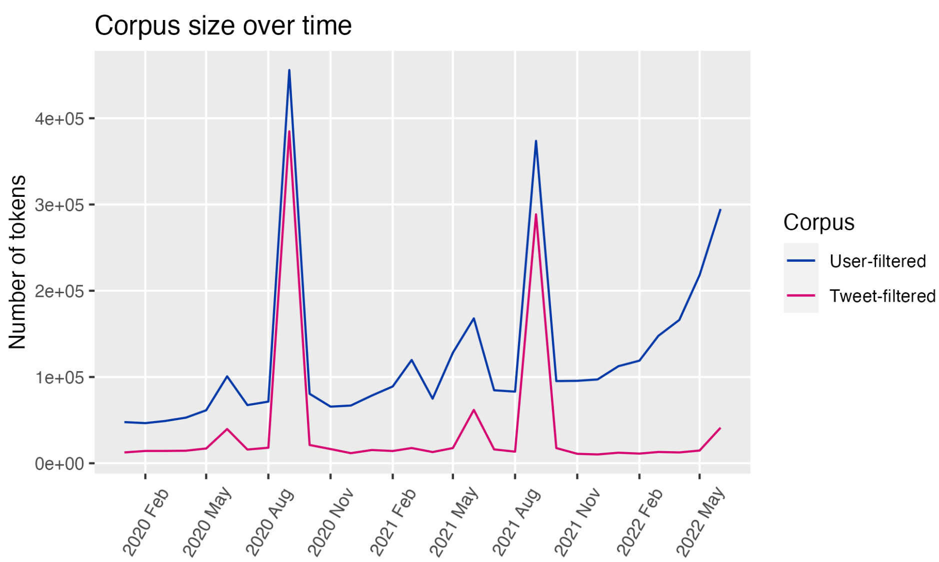

Figure 4.11: Size of corpora over time

(KL) Divergence or DKL, an alternative measure of keyness that is not as strongly correlated with raw frequency (Gries 2021).

4.7 Conclusion

The goal of this chapter was to uncover the ideologies and concepts that characterize bisexual discourses on Twitter. To that end, we conducted a keyness analysis of over 300,000 Tweets and in doing so developed a codebase for analyzing language-based Twitter data (Todd & Willis 2023). Sexually explicit discourses dominate the analysis of Tweets filtered at the user level. Bisexuality is primarily invoked as a means to connect with potential consumers or sexual partners in these Tweets (bisexuality as *instrumental*). In order to better understand bisexual discourses beyond those related to sex and desire, we further filtered our dataset at the Tweet level and conducted a second analysis. In the Tweet-filtered dataset, we found a number of topics and interactional functions related to bisexuality, ranging from support and struggle, to internet culture, to relationships and coming out. These Tweets bring to the forefront constructions of bisexuality relative to individual affective experiences and practices (bisexuality as *individual*) and relative to community and belonging (bisexuality as *community*). Our analysis highlights two main perspectives on bisexual Twitter discourses. The first engages with the link between bisexuality and sexual availability in ways that play with the relationship between commodification, marginalization, and power. The second amplifies bi-identified voices to create culturally intelligible bisexual identities. Both perspectives highlight the complex ways bisexual people strategically embrace or reject hegemonic understandings of bisexuality.

Chapter 5

Conclusion

This dissertation has explored the theoretical and methodological implications of bisexuality in language and sexuality studies. The conclusion first summarizes the key findings and arguments of each content chapter and then integrates all three content chapters. It also incorporates some discussion of directions for future research, including brief previews of some work in progress.

The first content chapter presents a linear regression analysis of acoustic data, comparing how bisexual people produce /s/ relative to lesbian, gay, and straight speakers. Results suggest that bisexual speakers produce /s/ differently compared to the other groups of speakers, albeit in complex ways. The chapter contextualizes these quantitative results with a qualitative analysis of post-test survey data on participants' gender stereotypicality. The argument focuses on the theorization of sexuality and gender, demonstrating the ways in which bisexuality disrupts the often unspoken ideological connections between queerness and gender normativity. It also discusses how bisexuality complicates the uncritical way methods like Likert scales and forced-choice tasks are used to collect information about participants, such as their sexual identity. The second content chapter presents an analysis of the acoustic data from content chapter one using random forests, a supervised machine learning algorithm. Results suggest that gender, sexuality, and race all play an important role in /s/ production. Through this analysis, the chapter demonstrates the efficacy of random forests for addressing sociolinguistic questions, particularly for those studies in which the participant group is recruited in a non-random fashion. It also draws attention to the complexity of collecting and operationalizing data about identity for quantitative studies, and makes recommendations for how to address these complexities.

The third content chapter presents a statistical keyness analysis of how people talk about bisexuality on Twitter. The analysis identifies a number of topics and interactional functions, some of which are transparently related to bisexuality as a self-identification term (e.g. coming out, bi culture) and others for which the relationship to bisexuality is more opaque (e.g. goods and services). Within these topics and functions, three salient constructions of bisexuality emerge: bisexuality as *instrumental*, bisexuality as *individual*, and bisexuality as *community*. Ultimately, the chapter provides insights into the concepts and ideologies that characterize discussions of bisexuality online. Quantitative studies often identify what merits further qualitative analysis, and this study is no different; as the chapter is necessarily broad, a closer qualitative analysis of individual Tweets and their posters or sub-communities within the broader #Bi Twitter sphere are potentially fruitful directions for future research.

As the title suggests, the primary goal of this dissertation was to demonstrate the theoretical and methodological efficacy of a bisexual framework in language and sexuality studies. To first address the former, this dissertation aimed to develop a theory of sexuality that is disentangled from ideology. I situated this effort within trans linguistics (Zimman 2020), and specifically problematized the "common sense" cultural understanding of sexuality that is binary, white, and mapped onto gender expression which permeates many studies of sexuality and the voice. The dissertation as whole takes as given that sexuality is not a dualistic contrast between lesbian/gay and straight and that bisexuality should be considered as distinct. The first content chapter expanded upon this theorization by examining the complex ways cisgender bisexual women and men orient themselves towards gender stereotypicality. The findings disrupt the assumption that people who are not straight are also not gender normative. That is, a theory of sexuality that accounts for bisexuality must simultaneously recognize the relationship between gender and sexuality (indeed, within the Western context in which this dissertation is written, sexuality is understood primarily through gender) without extending that relationship to naturalize mappings between sexuality and gender expression. The second content chapter enriches this theorization by identifying correlations and patterns that merit further exploration with more ethnographically-informed methods. Instead, the second core chapter focuses on methodologies, their limitations, and their implications for research practice. In that sense, the second core chapter addressed the methodological component of that first goal. Specifically, it argued for the value of random forests and reporting transparency in a bisexual framework. Put differently, it is not that random forests or transparent reporting are bisexual per se, but rather that their value is particularly salient when conducting research about bisexuality in different groups of people.

The secondary goal of this dissertation was to create a body of linguistic work about bisexuality. To the extent that this dissertation represents the largest collection of studies about bisexuality within linguistics (to my knowledge), this goal has been met. That is not to say, however, that the work is done. The work presented here suggests a number of directions for future research. Indeed, all three content chapters highlight the need for qualitative, ethnographically-informed analyses of people construct and negotiate bisexuality on the ground. For example, the extent to which the /s/ variation reported in the first two content chapters is taken up, if at all, is worth investigating using community-based approaches. Both chapters on /s/ repeatedly point out that it is impossible to know what this variation indexes without a better understanding of the communities who use it (or not). To address these questions (among many others), Willis (in progress) examines how bisexual participants construct their sexuality in semi-structured ethnographic interviews. The analysis attends to how participants' understandings of their bisexuality take shape in relation to other aspects of their identities, as well as how participants orient to discourses about bisexuality that circulate in their local communities and online.

Relatedly, the results of the third content chapter raise questions about the extent to which the discourses found on Twitter are found elsewhere. What is idiosyncratic to the Twitterverse and what is more widespread? In their study of bisexual discourses on the microblogging platform Tumblr, Willis and Fine (in progress) reveal that there is significant overlap between Twitter and Tumblr discourses about bisexuality. For example, there are many similarities between the salience of bisexual stereotypes (and contestation of these stereotypes) on both platforms. The analysis by Willis and Fine addresses the need for a close, qualitative analysis of how bisexuality is constructed on the ground, albeit on a related but distinct platform.

The final goal of this dissertation was to use this visibility as a platform to raise awareness about bisexual issues. The long-term impact of this work remains to be seen, but at the time of writing I am confident in saying that the fulfillment of this goal is ongoing. Through the numerous formal and informal presentations and conversations about the studies included in this dissertation, my collaborators and I have brought a bisexual perspective into spaces where it was lacking. As academic researchers, we play a critical role in leading efforts to increase cultural competency around bisexuality. Scholars can facilitate a deeper understanding of the unique challenges and experiences of bisexual people by actively engaging in research on bisexuality in linguistics (Murphy 1997; Thorne 2013; Turai 2021; Wilkinson 2019), but also elsewhere (e.g. Alarie & Gaudet 2013; Choi & Israel 2019; Feinstein & Galupo 2020; Galupo et al. 2019; Israel 2018). Engagement may range from conducting original research on bisexuality in their respective fields, to including research on bisexuality as assigned readings in coursework, to developing and disseminating educational materials for people outside the academy. Through this work, academic researchers can challenge stereotypes, promote visibility, and foster a more inclusive society for bisexual people. This work is far from done, however. In closing, I challenge the readers of this dissertation (academic researchers and otherwise) to translate their experiences with this document into action. Use the language, findings, concepts, and imaginings of the future in this dissertation to promote equality and social justice for bisexual people wherever and whenever you are.

Bibliography

Agha, Asif. 2003. The social life of cultural value. *Language & Communication* 23(3-4). 231–273. URL https://linkinghub.elsevier.com/retrieve/pii/S0271530903000120.

Alarie, Milaine & Stéphanie Gaudet. 2013. "I Don't Know If She Is Bisexual or If She Just Wants to Get Attention": Analyzing the Various Mechanisms Through Which Emerging Adults Invisibilize Bisexuality. *Journal of Bisexuality* 13(2). 191–214. URL https://doi.org/10.1080/15299716.2013.780004.

Audacity Team, 2019. *Audacity(R): Free Audio Editor and Recorder*. URL www.audacityteam.org.

Baker, Paul. 2004. Querying keywords: Questions in difference, frequency, and sense in keywords analysis. *Journal of English Linguistics* 32(4). 346–359.

Barron-Lutzross, Auburn. 2015. The Production and Perception of a Lesbian Speech Style. *UC Berkeley Phonology Lab Report* 11(11). URL https://escholarship.org/uc/item/7f6332bh.

Ben Youssef, Chadi & Stefan Th Gries. Forthcoming. Code-Switching in Tunisian Arabic. *Corpora*. 28.

Boersma, Paul. 2011. *Praat: doing phonetics by computer*. URL http://www.praat.org/.

Breiman, Leo. 2001. Random Forests. *Machine Learning* 45(1). 5–32. URL https://doi.org/10.1023/A:1010933404324.

Brewster, Melanie E. & Bonnie Moradi. 2010. Personal, Relational and Community Aspects of Bisexual Identity in Emerging, Early and Middle Adult Cohorts. *Journal of Bisexuality* 10(4). 404–428. URL https://doi.org/10.1080/15299716.2010.521056.

Bucholtz, Mary & Kira Hall. 1995. Introduction: Twenty Years after Language and Woman's Place. In *Gender Articulated: Language and the Socially Constructed Self*, Mary Bucholtz & Kira Hall (eds.), pp. 2–22. New York & London: Routledge.

— 2004. Theorizing identity in language and sexuality research. *Language in Society* 33(4). URL http://www.journals.cambridge.org/abstract_S0047404504334020.

— 2005. Identity and interaction: a sociocultural linguistic approach. *Discourse Studies* 7(4). 585–614. URL http://journals.sagepub.com/doi/10.1177/1461445605054407.

— 2016. Embodied sociolinguistics. In *Sociolinguistics*, Nikolas Coupland (ed.), pp. 173–198. Cambridge: Cambridge University Press. ISBN 978-1-107-44978-7. URL https://www.cambridge.org/core/product/identifier/CBO9781107449787A016/type/book_part.

Calder, Jeremy. 2019. From *Sissy* to *Sickening:* The Indexical Landscape of /s/ in SoMa, San Francisco. *Journal of Linguistic Anthropology* 29(3). 332–358. URL https://doi.org/10.1111/jola.12218.

— 2021. Whose indexical field is it?: The role of community epistemology in indexing social meaning. *Texas Linguistics Society* 39. 39–55.

Calder, Jeremy & Sharese King. 2020. Intersections between Race, Place, and Gender in the Production of /s/. *University of Pennsylvania Working Papers in Linguistics* 26(2). 10. URL https://repository.upenn.edu/pwpl/vol26/iss2/5.

Campbell-Kibler, Katherine. 2007. Accent, (ING), and the social logic of listener perceptions. *American Speech* 82(1). 32–64. URL https://read.dukeupress.edu/american-speech/article/82/1/32-64/5791.

— 2011. Intersecting Variables and Perceived Sexual Orientation in Men. *American Speech* 86(1). 52–68. URL https://read.dukeupress.edu/american-speech/article/86/1/52-68/5899.

Choi, Andrew Young & Tania Israel. 2019. Affirmative mental health practice with bisexual clients: Evidence-based strategies. In *Handbook of evidence-based mental health practice with sexual and gender minorities*, pp. 149–171. New York, NY, US: Oxford University Press. ISBN 978-0-19-066930-0.

Covington, Michael A. & Joe D. McFall. 2010. Cutting the Gordian Knot: The Moving-Average Type–Token Ratio (MATTR). *Journal of Quantitative Linguistics* 17(2). 94–100. URL https://doi.org/10.1080/09296171003643098.

Crenshaw, Kimberle. 1989. Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *University of Chicago Legal Forum*. 139.

Crist, Sean. 1997. Duration of Onset Consonants in Gay Male Stereotyped Speech. *University of Pennsylvania Working Papers in Linguistics* 4(3). 20.

Dunning, Ted. 1993. Accurate methods for the statistics of surprised and coincidence. *Computational Linguistics* 19(1). 61–74.

Eckert, Penelope. 2008. Variation and the indexical field. *Journal of Sociolinguistics* 12(4). 453–476.

Egbert, Jesse & Douglas Biber. 2019. Incorporating text dispersion into keywords analysis. *Corpora* 14(1). 77–104.

Ehrlinger, John. 2016. *ggRandomForests: Exploring Random Forest Survival*. URL https://arxiv.org/abs/1612.08974.

Eisner, Shiri. 2016. Monosexism. In *The SAGE Encyclopedia of LGBTQ Studies*, Abbie E. Goldberg (ed.), pp. 793–796. Thousand Oaks, SAGE Publications, Inc. URL https://sk.sagepub.com/reference/the-sage-encyclopedia-of-lgbtq-studies/i8278.xml.

Fairbanks, Grant. 1960. The Rainbow Passage. In *Voice and articulation drillbook*. New York: Harper & Row. URL https://doi.org/10.1288/00005537-194112000-00007..

Feinstein, Brian A. & M. Paz Galupo. 2020. Bisexual orientation cannot be reduced to arousal patterns. *Proceedings of the National Academy of Sciences* 117(50). 31575–31576. URL https://www.pnas.org/doi/full/10.1073/pnas.2016612117.

Forinash, Kyle & Wolfgang Christian. 2012. *Sound: An Interactive Book*. URL https://www.compadre.org/books/SoundBook.

Forrest, Karen, Gary Weismer, Paul Milenkovic, & Ronald N. Dougall. 1988. Statistical analysis of word-initial voiceless obstruents: Preliminary data. *The Journal of the Acoustical Society of America* 84(1). 115–123. URL http://asa.scitation.org/doi/10.1121/1.396977.

Fox, Robert Allen & Shawn L. Nissen. 2005. Sex-Related Acoustic Changes in Voiceless English Fricatives. *Journal of Speech, Language, and Hearing Research* 48(4). 753–765. URL https://pubs.asha.org/doi/full/10.1044/1092-4388%282005/052%29.

Fuchs, Susanne & Martine Toda. 2010. Do differences in male versus female /s/ reflect biological or sociophonetic factors? In *Turbulent Sounds: An Interdisciplinary Guide*, Susanne Fuchs, Martine Toda, & Marzena Žygis (eds.), pp. 281–302. New York: De Gruyter Mouton. URL https://doi.org/10.1515/9783110226584.281.

Galupo, M. Paz, Johanna L. Ramirez, & Lex Pulice-Farrow. 2017. "Regardless of Their Gender": Descriptions of Sexual Identity among Bisexual, Pansexual, and Queer Identified Individuals. *Journal of Bisexuality* 17(1). 108–124. URL https://doi.org/10.1080/15299716.2016.1228491.

Galupo, M. Paz, Son M. Taylor, & David Cole. 2019. "I Am Double The Bi": Positive Aspects of Being Both Bisexual and Biracial. *Journal of Bisexuality* 19(2). 152–168. URL https://doi.org/10.1080/15299716.2019.1619066.

Gaudio, Rudolf P. 1994. Sounding Gay: Pitch Properties in the Speech of Gay and Straight Men. *American Speech* 69(1). 30. URL http://www.jstor.org/stable/455948?origin=crossref.

Gershon, Ilana. 2011. "Neoliberal Agency". *Current Anthropology* 52(4). 537–555. URL https://www.journals.uchicago.edu/doi/10.1086/660866.

Gries, Stefan Th. 2021. A new approach to (key) keywords analysis: Using frequency, and now also dispersion. *Research in Corpus Linguistics* 9(2). 1–33. URL https://ricl.aelinco.es/index.php/ricl/article/view/150.

Hall, Kira. 1995. Lipservice on the fantasy lines. In *Gender Articulated*, Kira Hall & Mary Bucholtz (eds.), pp. 183–216. 1st Edition. Routledge.

Hartman, Julie E. 2013. Creating a Bisexual Display: Making Bisexuality Visible. *Journal of Bisexuality* 13(1). 39–62.

Hazenberg, Evan. 2015. Walking the straight and narrow: linguistic choice and gendered presentation. *Gender and Language* 10(2). 270–294. URL https://www.equinoxpub.com/journals/index.php/GL/article/view/19812.

Hertlein, Katherine M., Erica E. Hartwell, & Mashara E. Munns. 2016. Attitudes Toward Bisexuality According to Sexual Orientation and Gender. *Journal of Bisexuality* 16(3). 339–360. URL https://doi.org/10.1080/15299716.2016.1200510.

Ikae, Catherine, Sukanya Nath, & Jacques Savoy. 2019. UniNE at PAN-CLEF 2019: Bots and Gender Task. In *Conference and Labs of the Evaluation Forum*.

Ishwaran, Hemant. 2007. Variable importance in binary regression trees and forests. *Electronic Journal of Statistics* 1(none). 519–537. URL `https://doi.org/10.1214/07-EJS039`.

Ishwaran, Hemant & Udaya B. Kogalur. 2022. *randomForestSRC*. URL `https://cran.r-project.org/web/packages/randomForestSRC/randomForestSRC.pdf`.

Israel, Tania. 2018. Bisexuality: From margin to center. *Psychology of Sexual Orientation and Gender Diversity* 5(2). 233–242.

Jacobs, Greg, Ron Smyth, & Henry Rogers. 2000. Language and sexuality: Searching for the phonetic correlates of gay- and straight-sounding male voices.

Jones, Jeffrey M. 2022. *LGBT Identification in U.S. Ticks Up to 7.1%*. Section: Politics. URL `https://news.gallup.com/poll/389792/lgbt-identification-ticks-up.aspx`.

Jongman, Allard, Ratree Wayland, & Serena Wong. 2000. Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America* 108(3). 1252. URL `http://scitation.aip.org/content/asa/journal/jasa/108/3/10.1121/1.1288413`.

Kearney, Michael W. 2020. *tweetbotornot2*. Version 0.0.1. URL `https://github.com/mkearney/tweetbotornot2`.

Kinsey, Alfred C., Wardell Pomeroy, & Clyde Martin. 1948. *Sexual behavior in the human male*. Philadelphia: W.B. Saunders; Bloomington: Indiana University Press. URL `https://opus4.kobv.de/opus4-Fromm/frontdoor/index/index/docId/13535`.

Klatt, Dennis. 1974. The Duration of [s] in English Words. *Journal of Speech and Hearing Research* 17(1). 51–63. URL `https://pubs.asha.org/doi/abs/10.1044/jshr.1701.51`.

Laborde, Gant. 2021. *NSFW Detection Machine Learning Model*. Version 1.1. URL `https://libraries.io/pypi/nsfw-detector`.

Lakoff, Robin. 1973. Language and a woman's place. *Language in Society* 2(1). 45–79.

Lee, Sangseok & Dong Kyu Lee. 2018. What is the proper way to apply the multiple comparison test? *Korean Journal of Anesthesiology* 71(5). 353–360. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6193594/`.

Levon, Erez. 2006. HEARING "GAY": PROSODY, INTERPRETATION, AND THE AFFECTIVE JUDGMENTS OF MEN'S SPEECH. *American Speech* 81(1). 56–78. URL `https://read.dukeupress.edu/american-speech/article/81/1/56-78/5768`.

— 2007. Sexuality in context: Variation and the sociolinguistic perception of identity. *Language in Society* 36(04). 533. URL `http://www.journals.cambridge.org/abstract_S0047404507070431`.

— 2011. TEASING APART TO BRING TOGETHER: GENDER AND SEXUALITY IN VARIATIONIST RESEARCH. *American Speech* 86(1). 69–84. URL `https://read.dukeupress.edu/american-speech/article/86/1/69-84/5901`.

LGBTQIA+ GLOSSARY (2021). URL `https://rcsgd.sa.ucsb.edu/education/lgbtqia-glossary`.

Linville, Sue Ellen. 1998. Acoustic correlates of perceived versus actual sexual orientation in men's speech. *Folia Phoniatrica et Logopaedica* 50(1). 35–48.

Mendoza, Dozandri C. in prep. That /s/ tiene tumbao: Chonga-fied sibilants and the disidentificatory sociophonetics of Latinx drag.

miles-hercules, deandre. 2022. The Real Tea: Language at the Intersections. In *The Oxford Handbook of Language and Sexuality*, Rusty Barrett & Kira Hall (eds.). United Kingdom: Oxford University Press. URL `https://doi.org/10.1093/oxfordhb/9780190212926.013.70`.

Moonwoman-Baird, Birch. 1997. Towards the study of lesbian speech. In *Queerly Phrased*, Anna Livia & Kira Hall (eds.), pp. 202–213.

Muchlinski, David, David Siroky, Jingrui He, & Matthew Kocher. 2016. Comparing Random Forest with Logistic Regression for Predicting Class-Imbalanced Civil War Onset Data. *Political Analysis* 24(1). 87–103. URL `https://doi.org/10.1093/pan/mpv024`.

Munson, Benjamin. 2007. The acoustic correlates of perceived masculinity, perceived femininity, and perceived sexual orientation. *Language and Speech* 50(1). 125–142.

Munson, Benjamin & Molly Babel. 2007. Loose Lips and Silver Tongues, or, Projecting Sexual Orientation Through Speech. *Language and Linguistics Compass* 1(5). 416–449. URL `http://doi.wiley.com/10.1111/j.1749-818X.2007.00028.x`.

Munson, Benjamin, Sarah V. Jefferson, & Elizabeth C. McDonald. 2006b. The influence of perceived sexual orientation on fricative identification. *The Journal of the Acoustical Society of America* 119(4). 2427–2437. URL `http://asa.scitation.org/doi/10.1121/1.2173521`.

Munson, Benjamin, Elizabeth C. McDonald, Nancy L. DeBoe, & Aubrey R. White. 2006a. The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech. *Journal of Phonetics* 34(2). 202–240. URL `http://linkinghub.elsevier.com/retrieve/pii/S0095447005000379`.

Murphy, M. Lynne. 1997. The elusive bisexual: Social categorization and lexico-semantic change. In *Queerly Phrased: Language, Gender, and Sexuality*, Anna Livia & Kira Hall (eds.), pp. 35–57. New York & Oxford: Oxford University Press.

Pharao, Nicolai, Marie Maegaard, Janus Spindler Møller, & Tore Kristiansen. 2014. Indexical meanings of [s+] among Copenhagen youth: Social perception of a phonetic variant in different prosodic contexts. *Language in Society* 43(1). 1–31. URL `https://www.cambridge.org/core/article/indexical-meanings-of-s-among-copenhagen-youth-social-perception-of-a-phonetic-variant-in-different-prosodic-contexts/76B4354FEDB55A819FADDF321620BC33`.

Pierrehumbert, Janet B., Tessa Bent, Benjamin Munson, Ann R. Bradlow, & J. Michael Bailey. 2004. The influence of sexual orientation on vowel production. *The Journal of the Acoustical Society of America* 116(4). 1905–1908. URL `http://asa.scitation.org/doi/10.1121/1.1788729`.

Podesva, Robert J & Janneke Van Hofwegen. 2014. How Conservatism and Normative Gender Constrain Variation in Inland California: The Case of /s/. *University of Pennsylvania Working Papers in Linguistics* 20(4).

Project, Movement Advancement, BiNetUSA, & Bisexual Resource Center. 2014. *Understanding Issues Facing Bisexual Americans*. URL https://www.lgbtmap.org/file/understanding-issues-facing-bisexual-americans.pdf.

R Core Team, 2022. *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Vienna, Austria. URL https://www.R-project.org/.

Rayson, Paul & Amanda Potts. 2020. Analysing Keyword Lists. In *A Practical Handbook of Corpus Linguistics*, Magali Paquot & Stefan Th. Gries (eds.), pp. 119–139. Cham: Springer International Publishing. ISBN 978-3-030-46216-1. URL https://doi.org/10.1007/978-3-030-46216-1_6.

Rendall, Drew, Paul L. Vasey, & Jared McKenzie. 2008. The Queen's English: An Alternative, Biosocial Hypothesis for the Distinctive Features of "Gay Speech". *Archives of Sexual Behavior* 37(1). 188–204. URL https://doi.org/10.1007/s10508-007-9269-x.

Rogers, Henry & Ron Smyth. 2003. Phonetic Differences Between Gay- and Straight-Sounding Male Speakers of North American English. *15th International Congress of Phonetic Sciences*. 4.

Rogers, Henry, Ron Smyth, & Greg Jacobs. 2000. Vowel and Sibilant Duration in Gay- and Straight-Sounding Male Speech. *International Gender and Language Association Conference* 1.

Rothauser, E.H. 1969. IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics* 17. 225–249.

Shadle, Christine H. 1990. Articulatory-Acoustic Relationships in Fricative Consonants. In *Speech Production and Speech Modelling*, William J. Hardcastle & Alain Marchal (eds.), pp. 187–209. Dordrecht: Springer Netherlands. URL http://link.springer.com/10.1007/978-94-009-2037-8_8.

— 1991. The Effect of Geometry on Source Mechanisms of Fricative Consonants. *Journal of Phonetics* 19(3-4). 409–424.

Shapiro, Michael. 1995. A case of distant assimilation: /str/ → / tr/. *American Speech* 70(1). 101–107.

Smyth, Ron, Greg Jacobs, & Henry Rogers. 2003. Male voices and perceived sexual orientation: An experimental and theoretical approach. *Language in Society* 32(3). URL http://www.journals.cambridge.org/abstract_S0047404503323024.

Sommers, Stephen. 1999. *The Mummy*.

Strobl, Carolin, James Malley, & Gerhard Tutz. 2009. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods* 14(4). 323–348. URL http://doi.apa.org/getdoi.cfm?doi=10.1037/a0016973.

Tagliamonte, Sali A. & R. Harald Baayen. 2012. Models, forests, and trees of York English: *Was/were* variation as a case study for statistical practice. *Language Variation and Change* 24(2). 135–178. URL https://www.cambridge.org/core/product/identifier/S0954394512000129/type/journal_article.

Thomas, Erik. 2017. *Sociophonetics: An Introduction*. Bloomsbury Publishing.

Thorne, Lisa. 2013. "But I'm attracted to women": Sexuality and sexual identity performance in interactional discourse among bisexual students. *Journal of Language and Sexuality* 2(1). 70–100. URL http://www.jbe-platform.com/content/journals/10.1075/jls.2.1.03tho.

Todd, Simon & Chloe Willis. 2023. *Tools for Analyzing Linguistic Keyness (TALK)*. URL https://github.com/chloemwillis/Tools-for-Analyzing-Linguistic-Keyness.

Trousdale, Gary & Kirk Wise. 2001. *Atlantis: The Lost Empire*.

Turai, Ráhel K. 2021. Bisexuality: Narrating Sexual Fluidity. In *The Oxford Handbook of Language and Sexuality*. Oxford University Press. ISBN 9780190212926. URL https://doi.org/10.1093/oxfordhb/9780190212926.013.68.

Urban Dictionary (2021). *Urban Dictionary: bisexuality*. URL https://www.urbandictionary.com/define.php?term=bisexuality.

Van Borsel, John, Jana Vandaele, & Paul Corthals. 2013. Pitch and Pitch Variation in Lesbian Women. *Journal of Voice* 27(5). 656.e13–656.e16. URL https://linkinghub.elsevier.com/retrieve/pii/S0892199713000842.

Waksler, Rachelle. 2001. Pitch range and women's sexual orientation. *WORD* 52(1). 69–77. URL https://doi.org/10.1080/00437956.2001.11432508.

Wilkinson, Mark. 2019. 'Bisexual oysters': A diachronic corpus-based critical discourse analysis of bisexual representation in The Times between 1957 and 2017. *Discourse & Communication* 13(2). 249–267. URL https://doi.org/10.1177/1750481318817624.

Willis, Chloe. 2021. Bisexuality and /s/ production. *Proceedings of the Linguistic Society of America* 6(1). 69–81. URL https://journals.linguisticsociety.org/proceedings/index.php/PLSA/article/view/4942.

— forthcoming. Bisexuality in Experimental Sociophonetics: Ideologies and Implications. *Journal of Language and Sexuality* 13 (1).

Willis, Chloe & Chadi Ben Youssef. 2023. Random bisexual forests: Intersections between gender, sexuality, and race in /s/ production. *Proceedings of the Linguistic Society of America* 8(1). 5504. URL https://journals.linguisticsociety.org/proceedings/index.php/PLSA/article/view/5504.

Zimman, Lal. 2010. Female-to-male transsexuals and gay-sounding voices: A pilot study. 22.

— 2013. Hegemonic masculinity and the variability of gay-sounding speech. *Journal of Language and Sexuality* 2 (1). 1–39.

— 2017. Variability in /s/ among transgender speakers: Evidence for a socially grounded account of gender and sibilants. *Linguistics* 55(5). URL http://www.degruyter.com/view/j/ling.2017.55.issue-5/ling-2017-0018/ling-2017-0018.xml.

— 2018. *Duration & moments for /s/ [Praat script]*.

— 2020. Transgender Language, Transgender Moment: Toward a Trans Linguistics. In *The Oxford Handbook of Language and Sexuality*, Kira Hall & Rusty Barrett (eds.). Oxford University Press. ISBN 978-0-19-021292-6. URL https://oxfordhandbooks.com/view/10.1093/oxfordhb/9780190212926.001.0001/oxfordhb-9780190212926-e-45.