NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

LANGUAGE ACCOMMODATION AND GENDER IN SINGAPORE ENGLISH: A CORPUS-BASED APPROACH LOOKING AT SAME-GENDER AND MIXED-GENDER INTERACTIONS

KOH JIA JUN

SCHOOL OF HUMANITIES

2023

Language Accommodation and Gender in Singapore English: A Corpus-based Approach Looking at Same-gender and Mixed-gender Interactions

KOH JIA JUN

School of Humanities

A thesis submitted to the Nanyang Technological University in partial fulfilment of the requirement for the degree of Master of Arts

2023

Statement of Originality

I certify that all work submitted for this thesis is my original work. I declare that no other person's work has been used without due acknowledgement. Except where it is clearly stated that I have used some of this material elsewhere, this work has not been presented by me for assessment in any other institution or University. I certify that the data collected for this project are authentic and the investigations were conducted in accordance with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

Date:	13th August 2023
Name of student:	Koh Jia Jun
Signature of student:	

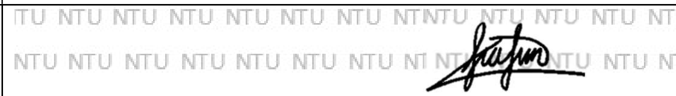
Supervisor Declaration Statement

I have reviewed the content of this thesis and to the best of my knowledge, it does not contain plagiarised materials. The presentation style is also consistent with what is expected of the degree awarded. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accordance with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

Date:	14th August 2023
Name of Supervisor:	Ng Bee Chin
Signature of Supervisor:	

Authorship Attribution Statement

This thesis **does not** contain any materials from papers published in peer-reviewed journals or from papers accepted at conferences in which I am listed as an author.

Date:	13th August 2023
Name of student:	Koh Jia Jun
Signature of student:	

ACKNOWLEDGEMENTS

Usually, people will call their theses a *labour of love*. I don't disagree, it really was a *labour*. While writing this thesis over the past two years, I often tell myself I'm "nurturing my intellectual baby", so that I will treat it with care, responsibility, and love — and now I finally gave birth to it. Many hands have helped me during the gestation period, for this piece of work is not something I could have achieved solely by myself.

Firstly, I'd like to thank my supervisor, Prof Bee, for the guidance since the start. You have been a supportive mentor, in school and in life. You were the fire under me to propel me forward because of your belief in me and what I can do when I did not believe in myself. Your push has allowed me to transcend my own limiting thoughts and to power through what I thought was undoable. I will always remember your teachings and this life experience.

Next, I want to thank the stellar research team led by, Prof Ying Ying, that adopted me into their family and treated me as one of them. Thank you Junwen for being the best stats consultant that I basically owe half my thesis to, Yun Sxin for being my graduating sister so that I never had to go through any hurdle alone, Xin Hwee for being the most supportive and chaotic force for good, Rae for being the stable senior I can pester, and Sophia for being my tutor in undergrad and now a friend in postgrad. We might have different "mothers", but all of you are the village that raised this fledgling academic. Lastly, I want to thank Prof Ying Ying herself, for offering me this opportunity to do a Master's degree, for being a fun and dedicated professor to tutor for, and for the laughs we shared together with the research team.

Thank you to all the year one undergraduate students I have taught for the short two semesters. Your enthusiasm to cocreate your own learning experiences with me made me realise that teaching is something I love to do, and I'm honoured to have been your tutor. Teaching all of you has been the highlight of this two-year voyage and I wish I could have done more of it.

I want to thank my family for the patience they have for me, and especially my partner, Dorothy, for being by my side through all this. Balancing work, self, and relationships is the hardest act in this journey, and I'm grateful to have you by my side through all the trials and tribulations and, of course, all the wins that came by. Thank you for imbuing me with love and meaning through it all.

Lastly, I’d like to thank myself. This endeavour into academia taught me many things about myself. Thank you, Jerry, for completing this, even when you thought at some points that it was not possible and felt like you were an imposter, that you should give it up because all this was a huge mistake. But you did it. You finished it in the end. You are not a fraud. You are capable of so much. This thesis is dedicated — to you.

This research is supported by the Ministry of Education, Singapore, under its Academic Research Fund Tier 2 (MOE2019-T2-1-084).

TABLE OF CONTENTS

SUMMARY

Do men and women speak differently? Language differences between men and women have always been a popular topic across various cultures and languages. Singapore English is no exception but research on gendered interactions have received little attention in Singapore. This thesis utilises conversations extracted from Singapore's National Speech Corpus (NSC), a three-part corpus of speech data produced by the Infocommunications and Media Development Authority of Singapore (IMDA), consisting of read scripts, local words, and dyadic interactions. These dyadic interactions will be viewed through the lens of Communication Accommodation Theory (CAT) and the Difference and Dynamic approaches in language and gender to examine how certain linguistic features that mark social distance are influenced by speaker gender and the dyad condition — i.e., same-gender dyads or mixed-gender dyads. The three linguistic features explored in this thesis are the use of backchannels, filled pauses, and Singlish particles, with an additional supplementary measure that employs the use of Language Style Matching (LSM) scores. Using linear mixed effects modelling, results show that speaker gender is a significant predictor for all three variables, with women using more backchannels than men, and men using more filled pauses and Singlish particles than women. For accommodation patterns from same-gender dyads to mixed-gender dyads, the use of backchannels diverged while the use of filled pauses and Singlish particles converged. Filled pauses and especially Singlish particles are shown to be good indicators of low social distance between interlocutors and reflect strategies used by both men and women to accommodate to each other. Results for LSM scores revealed that same-gender female dyads match more in linguistic style than same-gender male dyads, indicating the difference in communication styles between women and men — with women's communication style being more cooperative than men's. This study is the first quantitative analysis of linguistics features drawn from a large corpus of natural speech focusing on gender differences within the framework of CAT in Singapore. The findings give us a snapshot of how feminine and masculine speech styles are indexed in the Singapore context, and the usefulness of comparing same-gender and mixed-gender dyads to provide an interactional dimension to the analysis.

Chapter 1 Introduction

The purpose of this thesis is to bring together the topics of communication accommodation, language and gender, and Singapore English. It seeks to find out how language differs between men and women in the context of Singapore, and how it changes when participants speak to conversational partners of the same or different gender as them.

Currently, little had been done to investigate the role of gender in conversational interactions in Singapore. In the research of language variation within communities, gender constraints consistently show up as a crucial sociolinguistic variable (Xu, 2015), hence, this thesis will look into how these gender differences manifest in language and compare them with theoretical models that have attempted to explain these differences (and/or similarities). This study can therefore serve as a snapshot of Singapore's sociolinguistic landscape on the topic of language and gender in interactions both between and within men and women.

Even though this thesis makes use of several theoretical perspectives such as Communication Accommodation Theory (CAT) and Language Style Matching (LSM) to look at the differences of men and women in language and gender, it also serves as an exploratory piece of work to look at language and gender in Singapore in a much more discrete and data-heavy approach, analysing word and phrase tokens in a large corpus of around 1,000 hours of transcribed speech from more than 600 individuals. This way of leveraging large sample and token sizes can allow us to see broader trends more quantitatively, and hopefully reveal more statistically accurate patterns in language use. A linguistic study like this thesis can allow us to look at some of the claims from previous language and gender studies in a more localised context, with the data from a national corpus that is considerably voluminous and diverse.

Chapter 2 Literature Review

This literature review will firstly discuss Communication Accommodation Theory (CAT), followed by the history and major movements of language and gender literature to set some context. The review will then combine the two fields of research, delving into the effects of language accommodation in gendered contexts, with a focus on the four linguistics features this thesis is interested in.

2.1 Communication Accommodation Theory

2.1.1 An Overview

This section covers what Communication Accommodation Theory (henceforth, CAT) is, how its exhibited, and most importantly why we do it and how it benefits us as speakers. CAT started with the concept of Speech Accommodation Theory (SAT), which is a keystone idea in explaining the changes people exhibit when speaking to each other (Giles, 1973). Subsequently, after the inclusion of paralinguistic and other nonverbal form of communication in interactions, SAT was reconceptualised as Communication Accommodation Theory and, for decades, served as a robust theory for understanding how people interact with one another — whether it be differences in socioeconomic, linguistic, ethnic, gender, age, occupational status, culture, or any other factors in which different social groups possess (Giles, Coupland & Coupland, 1991).

The term *accommodation* here entails that the interlocutors either move away from or towards each other in terms of social distance by shifting and changing their communicative behaviours (Giles & Ogay, 2007). These communicative behaviours include linguistic features such as lexical choice and sentence structure, paralinguistic features such as pauses and speech rate, non-verbal features such as proximity and facial expressions (Giles & Ogay, 2007; Elhami, 2020), and even language varieties (Giles, Taylor & Bourhis, 1973). By altering these communicative behaviours, one can achieve *convergence* (reducing social distance) or *divergence* (increasing social distance) within the interaction. *Maintenance* (unchanged difference in social distance) is a specific form of divergence (Giles, 2016) wherein the speaker remains on their linguistic and interactional style without shifting or changing, which is, in essence, a form of non-converging behaviour with their conversation partner (Gallois, Ogay & Giles, 2005).

2.1.2 Why We Accommodate

There are several reasons why we accommodate linguistically (or not) with our interlocutors, which largely have to do with the social identities we choose to exhibit (Giles 1978, Tajfel & Turner, 1979). Divergence and maintenance happen when speakers want to emphasise the differences between themselves and their interlocutors, accentuating the dissimilarities. This is done usually as an act of displaying certain social and ingroup identities which require them to dissociate themselves personally from whoever their speaking to (Giles et al., 1991; Gallois et al., 2005), displaying a sense of ingroup pride and bringing about increased feelings of self-worth (Giles & Ogay, 2007).

Convergence, on the other hand, happens primarily due to similarity attraction (Byrne, 1971), wherein the speaker seeks to increase similarities with their interlocutor in order to gain approval and obtain a more positive impression from their speech partner (Giles & Smith, 1979). Being more similar to our interlocutors generally fosters an increased sense of liking and respect from them (Giles & Ogay, 2007; Wang & Fussell, 2010), and even improves communication effectiveness and mutual understanding by making the interaction more predictable, and lowering uncertainty, interpersonal anxiety (Gudykunst, 1995).

The process of accommodation is largely an unconscious process (Dragojevic et al., 2016) and attempts to explain these processes birthed various other models. In the interactive-alignment model outlined by Pickering and Garrod (2004), they explained that interlocutors align various aspects of an interaction, such as situation models of the conversation, as well as the linguistic aspects of lexicon, semantics, syntax, et cetera through priming cues. Similarly, in studies rooted in mimicry, humans have a natural tendency to imitate our conversation partners behaviourally (*chameleon effect,* Chartrand & Bargh, 1999) and verbally (*echo effect*, Kulesza et al., 2014). However, convergence can also be employed consciously to gain practical and social rewards for the speaker. In several studies, convergence has shown to function as a sound strategy to gain more positive evaluations and advantages in many areas, such as tipping (van Baaren et al., 2003), sales (Jacob et al., 2011), negotiation (Richardson et al., 2019) and courtship (Guéguen, 2009).

Now, the question is, how does language accommodation manifest itself in gendered interactions? However, before moving on to that topic, it would be necessary to go over the language and gender literature first to set some context in that field of research.

2.2 Language and Gender

The discourse surrounding Language and Gender can be broadly categorized into 4 major movements and approaches throughout time — namely the (1) Deficit approach, (2) Dominance approach, (3) Difference approach, and (4) Dynamic approach.

2.2.1 The Deficit Approach

The Deficit approach can be effectively attributed to Robin Lakoff's 1973 article (and subsequent book in 1975) *Language and Woman's Place*. Lakoff's, which claims the existence of a "women's language" that is distinctive in its characteristic of being uncertain, weak, excessively polite, and features such as the use of specific colour words, weaker expletives, women-only adjectives (e.g. lovely, divine, sweet), tag questions, the intensive "so", uncertain speech, polite forms (e.g. leaving a decision open, not imposing one's views), and requestive (as opposed to directive) speech.

However, this instrumental work by Lakoff is not without its shortcomings. Firstly, one of the biggest criticisms of Lakoff's work is its heavy reliance on introspection and its non-scientific approach (Siddiqui, 2014), as Lakoff even disclaimed in her original article, "The data on which I am basing my claims have been gathered mainly by introspection: I have examined my own speech and that of my acquaintances, and have used my own intuitions in analyzing it." (Lakoff, 1973, p.46). Secondly, Lakoff assumes a set of linguistic features forms a "women's language" when they are not necessarily gendered (Eckert & McConnell-Ginet, 2003), but instead, these "powerless" features (e.g., hedges, tag questions, polite terms) had been found to be more affected by a speaker's social status and familiarity with the social settings than with the speaker's gender (O'Barr & Atkins, 1987). Lastly, Lakoff's work was said to put women in a position where inferiority and deficiencies are seen and enhanced, as Dale Spender (1998, p. 8) puts it,

> "She takes male language as the norm and measures women against it, and one outcome of this procedure is to classify any difference on the part of women as 'deviation'. Given these practices, it is unlikely that Lakoff could have arrived at positive findings for women, for any differences revealed, where a product of language or of sex, would be predisposed to interpretation as yet more evidence of female deficiency." (Spender, 1998, p.8).

2.2.2 The Dominance Approach

Continuing from Spender, the Dominance approach is attributed to her 1985 book, *Man Made Language*, focusing less on women's inferiority, and more on men's dominance in society. Essentially, she raised the idea of how language, in and of itself, "has been literally man made and that it is still primarily under male control" (Spender, 1985, p. 12).

Since men have been holding dominant positions in society, this dominance is also reflected in how language is used and how meaning is constructed. The most straightforward example is the semantic pejoration of words involving women over time (Endendijk, van Baar & Deković, 2020). However, under scrutiny, the claim of language being made by men and fundamentally disadvantageous to women can be easily refuted with counterexamples of women using language against men in any way (Talbot, 2019). This indicates that the dominance approach's blanket claim that English, or any other language, is created by men to dominate women can be rather hyperbolic and a disingenuous characterisation of language and gender inequality.

Nevertheless, as pointed out by Baxter (2006), this approach remains an essential contribution to the feminist movement, illuminating a fundamental aspect of how language has been used (and is still used) to disparage women. It reflects to us the deeply entrenched sexism present in language, and a wider social phenomenon of gender inequalities, such as status and power between men and women.

2.2.3 The Difference Approach

The Difference approach is usually credited to Deborah Tannen's (1990) book *You Just Don't Understand*. Tannen's main thesis refutes the deficient status of women and the dominant position of men in society that the previous two approaches claim. Instead, her approach adopts a neutral stance, avoiding value judgements, and describing the differences in the language between men and women to be equally valid, with neither one being "better" or "inferior" to the other, which aligns with the neutrality and relativism in the field of linguistics.

This approach is built upon the work by Maltz and Borker (1982), who reasoned that the differences in girls' and boys' method of socialisation give rise to their respective ways of speech. Tannen proposes that men and women belong to two different and distinct subcultures and this subcultural difference is akin to a form of social separation, such as class or ethnic difference, in which boys and girls undergo very separate and distinct ways of socialisation growing up, resulting in a difference between men's and women's language use (Tannen, 1990).

However, the Difference approach saw insufficient quantitative evidence as much of the influential works by Maltz and Borker (1982) and Tannen (1990) relied on qualitative and anecdotal data, and thus lack empirical support for their claims (MacGeorge et al., 2004).

Moreover, maintaining Tannen's neutral approach here can overlook the socially ingrained power differences that do exist between men and women in various social, political, and socioeconomic contexts (Cameron, 1995; Talbot, 2019). It is still true that women still face ingrained, systemic sexism in institutional settings like business, government, politics, education, law, religion, and media (Baxter, 2006) which cannot be disregarded and explained away by mere "difference in culture" between men and women. As Penelope Eckert and Sally McConnell-Ginet (2003) wrote in the beginning of their book *Language and Gender*, "…the study of difference became an enterprise in itself and was often detached from the wider political context." (p.2). There is a bigger picture of the sociopolitical landscape where various gender-based inequalities stem from that cannot be ignored and overlooked, leading us to the final and most recent approach in the language and gender movement.

2.2.4 The Dynamic Approach

We can already find inklings of the Dynamic approach in earlier works, for example, Judith Butler (1988; 1993) on *performing gender*, Elinor Ochs on *indexing gender* (1992), and Erving Goffman (1976) and West and Zimmerman (1989) on *doing gender*. This performative view of gender harkens back to the work by Simone de Beauvoir (1949), who puts forth a succinct keystone idea, "One is not born, but rather becomes, a woman." (p.283).

The Dynamic approach is regarded as a social constructionist approach, positing that instead of being a biological correlate of sex, gender is a socially constructed identity which can dynamically change in different situations and with different individuals (Coates, 2015). Butler's book *Gender Trouble: Feminism and the Subversion of Identity* (1999) played an influential role in propagating this view of gender and its performative nature. Contrary to the previous approaches which view men and women as binary opposites that inhabit exclusive social and symbolic spheres the performative view sees gender not as something that is accomplished at birth or early on in life (Bucholtz, 2014). Rather, gender needs to be repeatedly affirmed by being performed and displayed publicly, or as Butler (1999) puts it, "gender is an identity tenuously constituted in time, instituted in an exterior space through a stylized repetition of acts" (p.179), which we can see in clothing (Entwistle, 2000), behaviour (Goffman, 1976; 1977), and of course, language.

This is an important step for the study of language and gender, especially when there are deviances in traditional gender norms, such as "tomboys" and "sissies", which the previous approaches cannot adequately describe and explain (Hall, 2003); biologically ambiguous peoples who do not fall neatly into sexual categories (Sax, 2002); and even possible contentions that biological sex itself could be a social construct (Davis & Preves, 2017). All these add on to the multiplicity, fluidity, and at times conflicting femininities/masculinities that one uses to construct and present themselves to the world.

Why the Dynamic approach is crucial in this thesis and our understanding of language and gender is because it adds much more nuance and flexibility to what "men's language" and "women's language" mean and other social factors come into play to affect the way people construct themselves through language. In the Dynamic approach, we can see the reversed directionality of effect between language and gender — it is not so much about how one's gender affects the language used, but rather how speakers use language as a tool to express and construct their gender, as can be seen in a lot of Kiesling's (1998, 2007) work on language, gender, and masculinities. Rather than looking at speech and interactional traits as inherent to men and women *per se*, we need to see these traits as "social performances which are semiotically linked (indexed) to [men/women], through cultural discourse and cultural models." (Kiesling, 2007, p.659).

It is difficult to talk about language and gender, and by extension, the terms "men" and "women" without talking about "masculinities" and "femininities". When talking about gender, the terms "masculine" and "feminine" — and their noun forms "masculinity" and "femininity" — frequently appear in the language and gender literature. Most works do not explicitly offer a working definition of masculinity or femininity, but that does not mean other social scientists have not attempted to do so (e.g, Spence and Buckner, 1995; Blechner, 1998; Hoffman et al. 2005). For the purposes of our discussion, we will adopt the definitions used in language and gender literature and masculinity and femininity will refer to the normative traits and behaviours that associate with the male and the female gender (c.f. Cameron, 1997; Eckert & McConnell-Ginet, 2003; Kiesling, 2007; Coates, 2013).

Unlike "men" and "women", which are social identities, "masculinity" and "femininity" are continuums with intensities that lie on a spectrum. That is, there exists gradations of masculine, more masculine, most masculine, but not male, more male, most male (Talbot, 2019) and likewise for feminine and female. Some researchers such as Paechter (2003) even use "masculinities" and "femininities" in place of "gender" to avoid using the term completely because of the constantly shifting definitions and connotations of "gender".

The reason why this distinction and differentiation between men/masculinity, and women/femininity is important is because when the topic on language and gender is approached, it is reductive and essentialist to have the generalisations of "men's language" or "women's language". Rather, what is more meaningful is what can data and observations tell us about how different variations of language index masculinity and femininity. To put it in another way, taking women and femininity as an example, instead of concluding "women speak with this trait because they are women", it may be more productive to say, "this trait indexes femininity in which females are more likely to employ to perform their gender".

2.2.5 Intersectionality

When discussing the construction of social identities in the Dynamic approach, the concept of intersectionality is bound to be evoked, as each individual is mélange of various social identities. The term *intersectionality* often comes up in language and gender studies, gaining a form of "citational ubiquity" (Wiegman, 2012) and became a buzzword in the spheres of gender, racial, and other forms of social activism (Cooper, 2015). In a nutshell, intersectionality is a framework first coined by Kimberlé Crenshaw (1989) when she spotlighted and expounded on inequalities faced by not just women, but black women. At the heart of it, intersectionality is about the different social identities an individual can embody and possess, including and not limited to gender, race, sexuality, and class, that "mutually constitute, reinforce, and naturalize one another" (Shields, 2008, p.302), meaning that the different social identities we claim for ourselves interact together to produce a unique experience that is formed and practiced.

As Eckert and McConnell-Ginet warns, "If we were to talk about gender as if it were independent of other categorization schemes, we would effectively erase the vast range of gendered experience." (2003, p.50). Therefore, for us to meaningfully talk about gender differences and similarities in terms of language use, we need to consider how gender overlaps with these various other social categories (see Hall, Rodrigo & Hiramoto, 2020). Gender does not exist independently in a vacuum separated from other social factors that influence human behaviour, as seen in Hare-Mustin and Maracek's (1988) work that questioned *The Meaning of Difference* between genders, they provided an example that aptly illustrates the intersectional nature of gender and power differences,

> "Typically, those in power advocate rules, discipline, control, and rationality, whereas those without power espouse relatedness and compassion. Thus, in husband-wife conflicts, husbands call on rules and logic, whereas wives call on caring. When women are in the dominant position, however, as in parent-child conflicts, they emphasize rules, whereas children appeal for sympathy and understanding. Such a reversal suggests that these differences can be accounted for by an individual's position in the social hierarchy rather than by gender." (p. 459).

In this scenario, a pattern of interaction that typifies a "men's style" of speech and "women's style" of speech is superseded by the family relational dynamic in the form of parental power over the child, that is, the mother-child effect eclipsed the male-female effect regarding power.

Each of us has a unique blend of social identities and accumulated life experiences, and we are community-driven beings that perform these traits through our speech and behaviour. As Eckert and McConnell-Ginet wrote, "Too much abstraction spoils the broth." (p.89), that is, humans are not just a "man" or a "woman". Some of us are husbands, wives, parents, business owners, students — all these social roles intersect to have influence over how we speak.

Intersectionality theory is markedly relevant in contexts where multiculturalism and multilingualism pervade, as will be seen later in this thesis. This also informs the approach and methodology of doing linguistic studies regarding gender.

2.3 Communication Accommodation and Gender

This section seeks to amalgamate the theories in the previous two sections. Now that the scene of language and gender has been set in the previous section, this section can move on to look at how language accommodation manifests itself in gendered interactions.

2.3.1 Application of CAT in Language and Gender Studies

Since CAT is a theory that is applied to speakers belonging to different social groups, it is sensible for CAT to be applied in the domain of gender given the presence of social disparities of power and inequality between men and women that has been presented in the previous sections. This is especially relevant to the two latest approaches in language and gender research — the Difference and Dynamic approach. The Difference approach centres around the idea that men and women belong to different cultures, and hence, the interaction between men and women would constitute an event of intercultural communication (Maltz & Borker, 1982). The Dynamic approach focuses on the performance of social identities the speaker wants to exhibit, which is also exactly what CAT is primarily focused on (Giles 1978, Tajfel & Turner, 1979), and it is up to the speaker to employ strategies (consciously or unconsciously) to converge, diverge, or maintain based on their conversational goals.

When it comes to interaction between men and women, it seems to be the case that speakers accommodate more to their interlocutor's speech style than to their actual gender (Hannah & Murachver, 1999), meaning that the accommodation process of convergence is a default process in mixed-gender interactions. Similarly, comparing mixed-gender to same-gender interactions, Mulac and colleagues (1988) concluded, "in mixed-sex dyads, it appears that both genders adopted a linguistic style more like that of their out-group partner than they would have maintained with an in-group partner." (p.331). This means that, if speakers perceive the interlocutor to be different from themselves, they tend to use the language of their interlocutors. This makes sense given what we have discussed in Section 2.1.2 *Why We Accommodate* — in most regular contexts, communication is largely a cooperative, productive, and prosocial phenomenon that benefits from accommodative behaviour.

2.3.2 Studies on Accommodation in Language and Gender

One of the earliest papers on the topic of communication accommodation between men and women is by Bilous and Krauss (1988), who concluded several key findings in same-gender and mixed-gender dyads in their study involving 30 males and 30 females. They found that women speak more and interrupt more in same-gender dyads but less in mixed gender dyads (convergence), while males did not change their use of interruptions in both contexts (maintenance). For mean utterance length, both men and women converged when speaking in mixed-gender dyads, with females reducing their mean utterance length and men increasing their mean utterance length. Both men and women used more laughter and backchannels in mixed-gender dyads than in same-gender dyads, with backchannels converging and laughter diverging (Bilous & Krauss, 1988).

We can see from above that accommodation is a natural process in mixed-gender interaction — when men and women interact with each other, they shift their language according to whom they are talking to. However, this shifting seems to be an asymmetric process as it has been shown that women tend to accommodate more readily than men are (Fitzpatrick, Mulac & Dindia, 1995; Namy, Nygaard & Saurteig, 2002; Stupka, 2011). Fitzpatrick et al. (1995), who used same-gender strangers, mixed-gender strangers, and married couples, found convergence of gender-preferential styles of speaking to be present in all interactions. However, women seem to be better able to shift their language style in all conversation contexts — i.e., with their husbands, with stranger men, and with fellow women. Whereas for men, they are less able to accommodate to a female style of speech, especially if they are a more traditional and sex-typed male, they are less capable to shift to a more feminine linguistic style. On the flipside, however, the strongest style shift observed in the study came from men who were talking with their spouses, wherein husbands accommodated to their wives' style twice as much as the wives accommodated to them. This study shows us two things — firstly, in general, women are more able to accommodate to their interlocutor than men, and secondly, relationship between individuals play a huge role in how much men are able to accommodate to women. The researchers attribute the latter point to a stronger and more salient couple identity than individual gender identity, concluding that both the gender and the relationship between the dyads are important factors to consider, such as whether they are strangers, friends, or romantic partners, and is very relevant point to this thesis' methodology (see Section 3.3 *Participants*).

A more recent work in language accommodation and gender also demonstrated that women accommodated to men more than men to women (Namy et al., 2002). However, this is done with speech shadowing task that involves only repeating word lists, which is not representative of actual interactions involving men and women. Similarly, in Stupka's (2011) undergraduate paper, she also showed women exhibited more convergence while men exhibited minimal convergence. While the methods used in her study was unclear and vague, she did raise good points on the intersection of gender with age and relationship status.

2.3.3 Overaccommodation

With the previous section broadly showing how women have a higher propensity to accommodate to their interlocutors, the phenomenon of overaccommodation becomes a particularly noteworthy aspect to characterise instances when accommodation is overdone. A study by Bayard (1995) in New Zealand showed that women and men swore at similar frequency in same-gender conversations. However, in mixed-gender conversations, women actually swore more than men due to their expectation and preconception that men tend to swear more, and hence overaccommodated in their use of swearing. This aspect of CAT is something to keep a lookout for as it can reveal certain social perceptions a speaker has of their interlocutors (Giles & Ogay, 2007).

2.3.4 Speaker Gender vs. Conversation Partner Gender

This phenomenon of overaccommodation links to a particularly illuminating study by Hancock and Rubin (2015), who reported that instead of the *speaker's gender* being a factor for linguistic variation, the *conversation partner's gender* is the more crucial and relevant factor affecting language use. The study saw both males and females using significantly more dependent clauses and interruptions when speaking to females, even when there was no evidence of gendered language — i.e., dependent clauses and interruptions are not found to be feminine ways of speaking, but both males and females tend to use more of it with female conversation partners. The researchers suggested that rather than accommodating to the linguistic features or the style of the conversation partner *per se*, speakers instead accommodate to a *mental schema* regarding how a particular gender speaks, even when this schema is not reflected in reality during the interaction. For example, when speaking to a man, a speaker would have an internal "men's language schema" that they are converging to, even if said man is not exhibiting such linguistic features (Hancock & Rubin, 2015). In this case, the theory behind CAT needs to be adjusted slightly, as accommodation is not just about the speaker's reaction towards what and how their interlocutor is speaking (Giles & Gasiorek, 2013), but more specifically about who the interlocutor is, or perceived by the speaker to be.

2.3.5 Speech Complementarity

From the above studies we have discussed, we can see that linguistic convergence is to be expected in gendered interaction as we can assume communication itself is largely a prosocial and cooperative phenomenon (e.g. Giles & Smith, 1979; Fitzpatrick et al., 1995; Kulesza et al., 2014). However, we also have to consider the nuances in *linguistic divergence* as they could be a sign of *psychological convergence* (Thakerar et al., 1982) arising from the gender of the speakers. In mixed-gender interactions, mutual divergence has been observed where men lowered their pitch make their masculinity more pronounced (Hogg, 1985), while women accentuate their femininity with increased pitch (Montepare & Vega, 1988). This has been labelled "speech complementarity" (Giles et al., 1991, p.33), where the goal of the conversation is the display of gender identity and distinction from the interlocutor for heterosexual appeal and social attractiveness (Goffman, 1977) and can happen concurrently with linguistic convergence in other variables to form a holistic, coherent narrative of this psychological convergence (Giles et al., 1991).

2.4 Linguistic Features in Language and Gender

With communication accommodation in gendered interactions having been covered above, this section will now focus on specific linguistic features that are commonly researched in the language and gender literature. The features usually reflect the tendency of speaking style of a given group — such as whether women talk more than men, use more backchannels, behave more tentatively in speech, use less swearing, and are more stylistically similar in speech (through the use of Language Style Matching scores). These features are presented with two factors in mind — prevalence in the literature and relevance to the current study that is conducted.

2.4.1 Verbosity

There exists a cultural stereotype that "women speak more than men", and various sexist idioms across cultures have exemplified this saying (Coates, 2015; Siddiqui, 2014). A generic statement like that is sure to draw many investigations from linguists and social scientists to prove the veracity of this statement. Some studies have found it to be true that women speak more (e.g. Reid et al., 2003), while other studies have found the opposite that men speak more (e.g. Bortfeld et al., 2001; Waara & Shaw, 2006), and there are studies that find that there is no significant difference between how talkative men and women are (e.g. Leaper & Ayres, 2007; Hannah & Murachver, 2007). However, one would actually be hard-pressed to find substantial evidence supporting the claim that women speak more, as most research have, in fact, found the opposite (see James & Drakich [1993] for a list of studies examining verbosity). However, these conclusions are meaningless without the context of speaking the studies are focusing on — from the setting and goal of conversation (Onnela et al., 2014), to the underlying social relationships in the interaction such as physician-patient dynamics (Hall & Roter, 2002). In Tannen's (2017) book regarding female friendships, women friend groups tend to talk more as compared to men — in terms of frequency, utterance length, and during more personal topics. However, this is only true in private settings as men talk far more in public spheres such as business and meetings. Therefore, when examining verbosity between the genders, the various extraneous factors as seen above have to be considered that can contribute to the overall amount a person would talk.

Another challenge to measure verbosity is also the method being used. There are several ways to do so, and different studies employ different methods in measuring it — some examples include the total word count output by a speaker, floor time, number of turns, number of words per turn. The issue is that these different measures for verbosity might not even be consistent with one another, such as Frances' (1979) study which showed that women take more conversational turns but men talk for a longer duration. Therefore, the investigation and measure for verbosity between gender differences is more complex than it first sounds.

2.4.2 Supportive communication

Backchannelling is a verbal or non-verbal behaviour of a listener that signals to the speaker their attention and support without taking over the conversational floor (Yngve, 1970), such as "mm", "uh-huh", "yeah" (verbal) and head nods (non-verbal). These backchannels are also often referred to as minimal responses and are regularly used as a strategy for speakers to establish a collaborative floor for greater friendship and closeness through the continued display of interest and presence in the conversation (Coates, 2015).

In the language and gender literature, it is common to find conclusions that women use backchanneling more than men do (e.g. Thorne & Henley, 1975; Duncan & Fiske, 1977; Fishman, 1978; Roger & Nesshoever, 1987; Bilous and Krauss, 1988, Mulac & Bradac; 1995), and the most recent meta-analysis conducted by Plug et al. (2021) also came the conclusion that women use more supportive language than men do.

The reasons for women's predominant use of supportive backchanneling are subjected to debate depending on which approach one takes. On one hand, in line with the dominance approach, Pamela Fishman (1980) suggests that the lower power of women in interactions manifest in a readiness to support men's conversation topics; whereas on the other hand, in the Difference approach, Maltz and Borker (1982) suggests that women learned to make interested listener sounds as they have socialized to do so while men do not and tend to stay silent instead when listening. Mulac and colleagues (1998) investigated backchanelling based on the Dominance and Difference approach in their study involving 96 conversations with mixed-gender and same-gender dyads and found support for the Difference approach and not the Dominance approach.

Even so, whether it is dominance-based or difference-based, the fact of whether women backchannel more than men depends on the methodology used, the languages under study, and the region/culture in question. Dixon and Foster (1998) found no significant effects on speaker gender and both verbal/non-verbal backchannels using mixed-sex and same-sex dyadic interactions in white English speakers in South Africa; while a computational approach using a Japanese corpus showed that women backchanneled more than men (Ptaszynski, Hasegawa

& Masui, 2014). Again, harking back to the workings of intersectionality theory and complexity of social categories, we can see that every single piece of research and its conclusions can only be interpreted within the bounds of the data collected.

2.4.3 Tentativeness and Filled Pauses

When Lakoff (1973) noted that "women's language" has a tendency to convey uncertainty, she referred to lexical and phonological strategies that mitigate forthrightness, such as the use of tag questions (e.g. *isn't it?*), filler words (e.g. *well, you know, like, uh, um*), empty adjectives (e.g. *divine, sublime*), tentative language (e.g. *probably, kind of, sort of, quite*), and rising intonation at the end of declarative sentences (or uptalk/upseak, see Warren, 2016). These strategies, Lakoff claimed, aim to soften the strength of assertions and allow the speaker to avoid being too direct and possibly come into conflict with their interlocutor. Much of these claims of early feminist work see that the greater use of hedging/tentative language by women is a testament to their uncertainty and position of powerlessness (Lakoff, 1973), and the domination of men (Spender, 1985).

Dubois and Crouch (1975) were the first to investigate Lakoff's claim and found that men tend to use more tag questions than women, whereas McMillan et al. (1977) found that women use more tag questions in the task-oriented communication, and that tentative speech can function to express interpersonal sensitivity rather than the lack of power. Like the above two examples, early studies often have conflicting results and are subjected to methodological issues (see Holmes, 1984). In a similar vein, Holmes (1982) and Cameron et al. (1989) noted the complex nature of tags as not just a function that seeks confirmation, but also have facilitative, softening, or challenging effects (outlined in Holmes, 1995). Both Holmes and Cameron et al. found women's use of tags to be facilitative or mitigating and men's to be confirmation-seeking which further complicates matters because Cameron et al.'s (1989) study also saw that individuals in more powerful social and interactional positions, such as teachers, doctors, and parents, tend to use tag questions in more facilitative and mitigating ways while the party with the lower status (students, patients, children) used more for confirmation-seeking. This subverts what we expect from gender and power relations as this would logically (and erroneously) conclude that women are in a position of *higher* power than men since women use more facilitative and mitigating tags and men use more confirmation-seeking tags. This speaks to the complexity and difficulties with using tags as a measure for gender differences.

More recently, Reid et al.'s (2003) study on tentative language, that is, hedges, disclaimers, and tag questions, found that women used more tentative language in cooperative situations, when gender awareness was primed in a pre-study. Laserna, Seih and Pennebaker (2014) also looked at filler words and pauses and found women tend to use more filler words (e.g., *I mean*, *you know*, *like*), whereas disfluencies like filled pauses (*uh*, *um*) had no correlation with gender, but correlated with age instead.

However, researchers such as Shriberg (2001), Tottie (2011) and Liberman (2014) have found sizeable differences between men's and women's use of filled pauses (*uh*, *um*). Liberman reported that men used filled pauses 38% more than women in the Fisher corpus (Cieri et al. 2004; 2005) with the overwhelmingly greater usage of *uh*, which is consistent with the findings by Shriberg and Tottie who reported similar outcomes in the Switchboard corpus (Godfrey, Holliman & McDaniel, 1992) and the British National Corpus (BNC) respectively.

On the topic of filled pauses, there is a particularly interesting pattern in terms of gender differences and the type of filled pauses used. Acton (2011), using the Speed Dating Corpus and Switchboard Corpus, saw a difference in the frequency of *um* and *uh* between men and women, with the *um*/*uh* ratio of women to be more than men by 3.5 times and 2.5 times respectively for each corpus. Similarly, Liberman's (2014) and Tottie's (2011) research corroborated Acton's findings, meaning that across research, *um* is more indexical of feminine speech while *uh* is much more common in masculine speech comparatively.

Apart from English, a similar pattern is observed in Mandarin Chinese in the sounds 嗯 *en* and 呃 *e* (Yuan et al., 2016), a parallel to um/uh with a filled pause — one with a nasal consonant at the end, and one with an no syllabic coda. Their study found that gender have effects on the frequency of filled pauses used, with men using significantly more filled pauses than women, which is in line with the previous studies in English.

It is noteworthy here that filled pauses function more than mere "hesitation markers" (Tottie, 2011, p.193). The importance and function of filled pauses has been debated over the years (see Clark & Fox Tree, 2002) — such as being errors and thus irrelevant parts of speech, to being useful meaningful units of natural language use. However, one particular aspect of relevance to highlight is that the frequency of filled pauses is not automatic and have consistently been shown less frequent in formal contexts compared to informal contexts (Schachter et al., 1991; Broen & Siegel, 1972). This can be useful in helping us see how a speaker positions themself in relation to the interlocutor, with the number of filled pauses being an indicator of the level of formality and social distance between them.

2.4.4 Swearing

Lakoff (1973) stated in her book, "In any event, it is a truism to state that the 'stronger' expletives are reserved for men, and the 'weaker' ones for women." (p.50). In line with this, Precht (2002) did a corpus study of recorded conversations and found that men were found to say *shit* significantly more than the women, and women use the euphemistic *gosh* significantly more than the men. Uses of *damn* or *god*, on the other hand, was not found to have significant sex differences. Additionally, in Jay's (2009) extensive research in swearing, he found that women were indeed five times more likely to say the weaker swear phrase *oh my god*, accounting for a quarter of women's swearing, and corroborates the claim that men tend to say more offensive words (e.g. *fuck*, *shit*) more frequently than women.

While researchers like Güvendir (2015) had explained gender differences in swearing on a *biological* basis, we also have to look at the *sociological* aspect of swearing when it comes to interactions between people and gender. Sara Mills, in her book, *Gender and Politeness* (2003) provided a critical review of previous politeness models by Brown and Levinson (1978, 1987), and argued for a more flexible way of seeing swearing not just as a speech act of crudeness and impoliteness, but also one of social closeness and solidarity. Incidentally, Mills (2003) also champions the performativity of gender through language use, where women can construct their femininities even through highly masculinised language, such as swearing, and this phenomenon can also be found in non-western regions (Okamoto & Sato, 1992)

This can be seen in Jay's (2009) research that concluded both men and women tend to swear more often in same-gender contexts than in mixed-gender contexts, meaning that there is a gender effect on how likely a person is to use swear words. Such intergender phenomenon alludes to the Goffman's (1977) idea of how there is a heterosexual incentive for men to be more courteous in the presence of women, given that the performance of heterosexuality is an important display of masculinity (Coates, 2013; Kiesling, 2007). As for women, swearing is often viewed as an impolite and comparatively inappropriate behaviour to exhibit, and women's performance of femininity tends to be more deferential and "nice" (Mills, 2003, p.204). However, as mentioned in Section 2.3.3 on overaccommodation, it is also possible for women to swear more than men when speaking to men, perhaps as a display of perceived solidarity and closeness (Bayard, 1995).

Therefore, this means that an analysis of the amount and type of swear words used between same-gender and mixed-gender dyads can tell us several things about the speakers themselves, such as camaraderie, comfort level with the conversation partner, formality, self-consciousness and self-presentation to the same/opposite-gender interlocutor.

2.4.5 Discourse Particles

The field of research surrounding discourse markers is voluminous with numerous perspectives that cannot be fully covered in this subsection (see Maschler and Schiffrin, 2015), but a subset of it, discourse particles, will be looked at in more detail as it is related to the study conducted in this thesis.

In Deborah Schiffrin's (1987) pioneering book *Discourse Markers*, discourse particles are defined as "sequentially dependent elements which bracket units of talk" (p.31), which include conjunctions, interjections, adverbs (e.g. *now*, *then)*, and lexicalised phrases (e.g. *y'know*, *I mean*). However, the particular aspect of discourse markers this thesis is interested in concerns discourse particles, which has a starker presence in Asian languages such as Mandarin, Japanese, Korean, Cantonese, et cetera.

Several studies on Mandarin Chinese showed the usage of different types of sentence-final discourse particles and the different functions they serve (Chan & Lin, 2019). For example, Farris (1995) identified the particle 嘛 *mà*, as a feature of feminine speech during 撒娇 *sājiāo*, a coy-like way of speaking specifically done by women; Shih (1998) also reported that women use a great deal more sentence-final particles and interjections than men, including particles like 呀 *ya*, 呢 *nē*, 啦 *là*, and 耶 *yè*; Hu (1981) suggested that Beijing women's greater usage of 吧 *bà* is a hedge to sound less definite and more tactful, softening imperatives and interrogatives as a politeness strategy (Chan, 1998), aligning to Chinese norms that women are expected to be polite in speech (Hong, 1997). This means that Lakoff's (1973) and Holmes' (1995) claim that women using more standard forms than men is also supported in Chinese communities (Van den Berg & Xu, 2010), as well as women's less frequent and less intense use of profanities (Chan, 1998) as it would lower a woman's social status (Shih, 1984) — constraints that Chinese men are not bound by.

Similarly for Japanese, in Okamoto's (1997, 2013) in-depth work on language and gender and Inoue's (2002) history of women's language in Japanese, the main indicators of gendered speech, termed "men's language" (*danseigo* 男性語*)* and "women's language" (*joseigo* 女性

語) lies in the use of polite forms of honorifics and, more particularly, sentence-final particles, where there exists male-exclusive and female-exclusive particles, referencing classic works from Silverstein (1979) and Ochs (1990), who showcased how gender expression in speech is accomplished. Moreover, these studies on Japanese are also congruent with the western studies in previous sections on the tendency for women's language to be more polite.

From the above examples of Mandarin Chinese and Japanese, it can be seen that discourse particles serve as an effective window to look into gendered speech, and examining the amount and type of discourse particles can shed light on how men and women interact in dyadic contexts.

2.4.6 Function Words: Language Style Matching

One other way to make use of speech data to measure linguistic style is to look at the amount of the types of function words used by each participant (as opposed to content words like verbs, nouns, and adjectives). Function words are useful because of their high frequency of occurrence in speech (Gonzales, Hancock & Pennebaker, 2010), which makes it easy to measure in interactions, and they also have the added advantage making up a substantial part of speech — more than 50% of what we say in speech despite being only 0.004% of a speaker's lexical repertoire (Chung & Pennebaker, 2007; Rochon et al., 2000). Hence, this is where Language Style Matching (LSM) comes into play.

Language Style Matching is an index ranging from zero to one ($0.0 < LSM < 1.0$) that indicates how similar two speeches/texts are in terms of linguistic style. The closer the index is to 1, the closer the two texts match in linguistic style. The idea of LSM is first introduced in the paper by Niederhoffer & Pennebaker (2002) and is based on function words usage of the two pieces of text. The nine function word categories used for the calculation of LSM are as follows:

1) auxiliary verbs (e.g., to be, to have)
2) articles (e.g., an, the)
3) common adverbs (e.g., hardly, often)
4) personal pronouns (e.g., I, they, we)
5) indefinite pronouns (e.g., it, those)
6) prepositions (e.g., for, after, with)
7) negations (e.g., not, never)
8) conjunctions (e.g., and, but)
9) quantifiers (e.g., many, few)

It is only until Gonzales et al.'s (2010) paper that a formula for calculating LSM score is formalised and used to calculate the LSM score (see Section 3.6.5 *Measuring Language Style Matching*).

LSM has been used in various relationship contexts, such as romantic interest and relationship stability between dating pairs (Ireland et al., 2011), friendship closeness between famous authors/poets (Ireland & Pennebaker, 2010), as well as relationship commitment and satisfaction (Bierstetel et al., 2020). Although LSM has not been used to analyse differences in linguistic style within and across gender in the literature, higher LSM scores have been shown to promote feelings of liking (Gonzales et al., 2010) and relationship stability (Ireland et al., 2011). Meaning that, to a considerable extent, LSM can be used as a tool to analyse same-gender and mixed-gender dyads to see how these groups differ, and what it can tell us about linguistic style in gendered interactions.

2.5 Language and Gender in Singapore

This section now zooms in to the target demographic this thesis is interested in — the study of language and gender in the context of Singaporeans and Singapore English. While there are studies on language differences in Singapore with gender as a factor, studies that involve language use in *interaction* within and across genders are deficient. Nevertheless, this section shall attempt to synthesise the literature reviewed above with studies in Singapore to come up with hypotheses that will serve as the backbone of this thesis.

2.5.1 Research Questions and Hypotheses

A common thread joining most of the linguistic features discussed in the previous section is the theme of cross-cultural communication and a change in formality in cross-gender interactions due to an increase in social distance.

The crux of this thesis is derived from combining the theories of CAT in Section 2.1 (*Communication Accommodation Theory*) together with the Difference and Dynamic approaches presented in Section 2.2 (*Language and Gender*). The impetus is based on the premise that men and women inhabit different gender spheres, and hence a speaker would have a greater social distance with an interlocutor of a different gender than an interlocutor of the same gender as them (e.g. Mulac et al., 2001). As such, in interactions between mixed-gender dyads, it would be expected to contain more formal features, less solidarity markers, and less similar in linguistic style due to the larger social distance; and vice versa for same-gender dyads due to the closer social distance.

Hence, the main research question guiding this thesis is:

RQ: How do men and women in Singapore negotiate social distance through language use as it relates to (1) their own gender, (2) their conversation partner's gender, and consequently, (3) whether they are in same-gender or mixed-gender pairings?

This will be investigated by looking at linguistic variations between men and women in same-gender dyads (male-male, female-female) and mixed-gender dyads (male-female). This means examining the effect of the speaker's gender and the conversation partner's gender on certain linguistic features, and seeing how the results align with the more recent models in the language and gender discourse, and what it can tell us about intergender and intragender communication in Singapore.

There will be four hypotheses made in this thesis, presented in the next four subsections. For the first three hypotheses: backchannelling, filled pauses, and Singlish particles, they will be seeded in two parts:

(a) In terms of the effect of gender on the dependent variable
(b) In terms of mixed-gender pairings in relation to same-gender pairings (F vs. FF; M vs. MM)[1], i.e. accommodation.

Part (a) gives insight to differences in linguistic behaviours between men and women, while part (b) goes a step further to see how these linguistic behaviours change between same-gender and mixed-gender contexts.

Lastly, the fourth hypothesis on Language Style Matching will be slightly different, as it will be based solely on the three gender pairing configurations — male-male (MM), female-female (FF), and male-female (MF) pairings — due to the nature of how LSM scores are calculated (see Section 3.6.5 *Measuring Language Style Matching*).

2.5.2 Hypothesis 1: Backchannelling

There have been no studies done in Singapore regarding supportive communication and gender. It is therefore in the interest of this thesis to investigate and draw conclusions on backchannels/minimal responses used in dyadic interactions of different genders.

In alignment with most research on this subject (see Section 2.4.2 *Supportive communication*), I would expect women to use higher rates of backchannelling than men as backchannels are typically associated to feminine ways of speaking (e.g. Thorne & Henley, 1975; Duncan & Fiske, 1977; Fishman, 1978; Roger & Nesshoever, 1987, Plug et al., 2021). However, in mixed-gender interactions, I would expect a convergence for both men and women in backchannel usage with increased usage for both genders (Bilous & Krauss, 1988) due to the increase in social distance. Hence, the first hypotheses are as follows:

H1a: Women will use more backchannelling than men.

H1b: Both men and women will converge in mixed-gender dyads — with both women and men increasing in their use of backchannels.

[1] The abbreviation convention for this thesis is such that FF represents women from same-gender dyads, MM represents men from same-gender dyads, F represents females from mixed-gender dyads, and M represents males from mixed-gender dyads.

2.5.3 Hypothesis 2: Filled Pauses in Singapore English

As we have seen in Section 2.4.3 *Tentativeness*, this area involving tentative language covers a wide range of linguistic features such as tag questions, filler words, filled pauses, etc., but not all are immediately relevant.

Using data from the International Corpus of English (ICE), Hoffman, Blass and Mukherjee (2014) investigated the tag question in various Asian Englishes and found that Singapore English uses tag questions almost four times less than British English. Both Hoffman et al. (2014) and Takahashi (2014) reported Singapore English uses tag questions in distinctly different ways from the canonical usage of tags (e.g., "She's more controlled now *is it*" [p.704] instead of *isn't it*). Additionally, Hoffman and colleagues also found that other invariant tags (e.g., *right, okay*) and Singlish particles like *lah*, *hah*, *hor*, *leh*, *lor*, *mah*, and *meh* can be used to replace a wide-ranging use of tags, which could account for the low frequency of tag questions in Singapore English. Therefore, with Singapore English's deviation from canonical tag question usage, coupled with the complexity of tag questions tackled earlier (e.g., facilitative vs. mitigating vs. confirmation seeking), a study on tags in Singapore English is not warranted.

Similarly, expressions such as *you know*, *I mean*, *well* can be complex to capture and categorise. They have been classified differently by various researchers — which includes "filler words" (Laserna et al., 2014), "discourse particles" (Eckert & McConnell-Ginet, 2003), "pragmatic particles" (Holmes, 1984), or even just under the broader term "tentative language" (Leaper & Robnett, 2011), often with these terms overlapping with one another and used interchangeably. Also, these terms carry with them their original semantic meanings alongside their function as filler words and can be hard to distinguish, especially in large untagged corpora, which makes them problematic to analyse.

Instead, more practically, we can look towards the use of filled pauses (*uh* and *um*) to give us various indicators, such as verbal delays in speech (Clark & Fox Tree, 2002), as well as the level of formality in the interaction as filled pauses are consciously used less in formal settings than in informal settings (Schachter et al., 1991). Following this logic, I can expect both men and women will reduce their use of filled pauses in the mixed-gender context. In terms of intergender difference, I can also expect men to use more filled pauses than women (Shriberg, 2001; Tottie, 2011; Liberman, 2014; Yuan et al., 2016).

Hence, Hypothesis 2 are formulated as follows:

H2a: Men will use more filled pauses (um/uh) than women.

H2b: In mixed-gender dyads, both males and females (M, F) will have lower usage of filled pauses than in same-gender dyads (MM, FF).

2.5.4 Hypothesis 3: Discourse Particles is Singapore English

One of the most recent and relevant study on this topic is the paper by Leimgruber, Lim, Gonzales and Hiramoto (2020), who looked at gender variation in the use of Singlish discourse particles in text messaging. They concluded that ethnicity and gender are predictors of particle usage, and specifically, they found that men use discourse particles significantly much more than women, with the exception of *leh* amongst the 10 discourse particles under their investigation. Leimgruber and colleagues also investigated the use of particles in relation to ethnicity and found that there were no significant differences between ethnicities, except Indians who use significantly fewer particles than the Chinese and Malays. Interaction effects that are reported between gender and ethnicity are only found for the particles *lah* and *sia*.

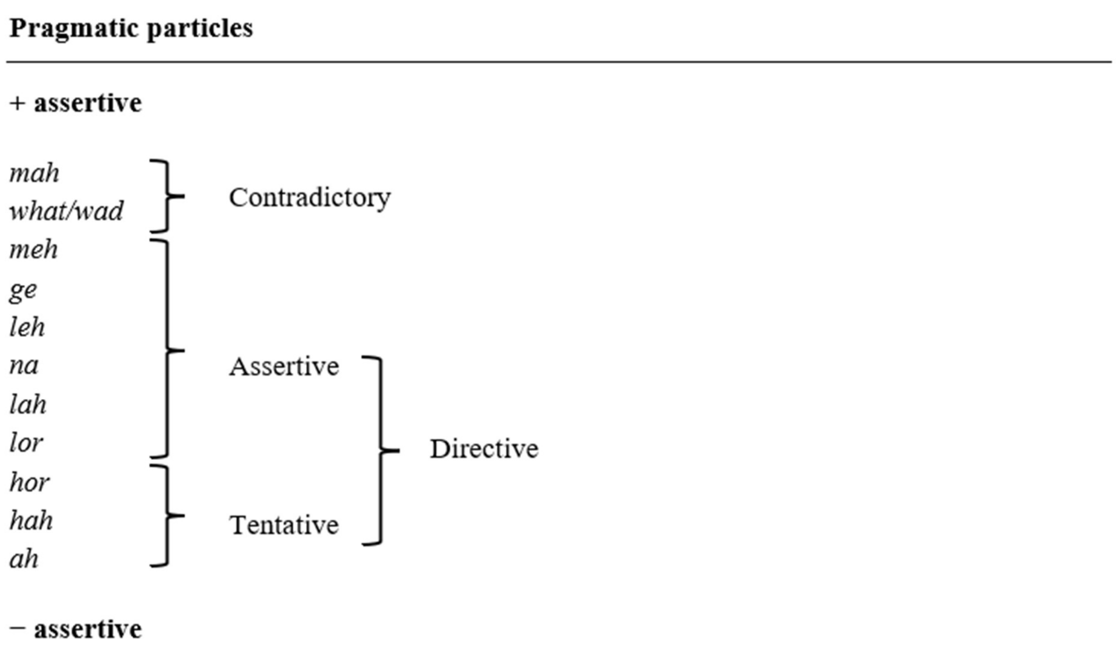
The insight gained from this study could be useful when we look at particle usage in dyadic interactions as indicators of formality, status, and solidarity, which is extremely valuable and useful, but this does not form the full picture of gendered interaction. Crucially, there are missing information about the interlocutor whom participants are communicating with — such as their relationship to the speaker, their gender, which informs how particle usage differs because of these factors. These are important, connected pieces when looking at interactions between men and women, which is why in this thesis, this dimension will be explored.

The analysis of the frequency and type of Singlish particles used by a speaker can provide us with a unique window to look at gendered interactions. Singlish, or sometimes termed Singapore Colloquial English (SCE), is the colloquial low(L)-variety of Singapore English is often used among friends and everyday language use (e.g. Alsagoff, 2010; Gupta, 1994; Rubdy, 2001). Hence, the use of Singlish particles in a dyadic interaction can indicate a kind of in-group solidarity between speakers. Even though Cavallaro and Ng (2009) have shown that covert prestige and solidarity traits of SCE were not evident in the results of their matched-guise test, this is because of the various confounds such as status-consciousness and government language policies that shaped language attitudes towards Singlish in public spheres and in private spheres (Cavallaro, Ng & Seilhammer, 2014). In practice, Singlish still very much remains as an essential "'glue' that binds Singaporeans together" (Cavallaro & Ng, 2009, p. 154) and Singlish particles, such as *lah*, has been treated as a marker of solidarity in various classic works on Singapore English (e.g., Platt & Weber, 1980; Platt, 1987).

Additionally, particles are discourse markers that are peripheral to the utterance, meaning that their presence or absence will not change the meaning of the utterance, and only speaker's attitudes on that utterance changes (Gupta, 2006). Hence, looking at the different types of Singlish particles can show the stance that speakers are taking in their interactions based on the many different functions Singlish particles have. Table 1 below shows how Gupta (1992) classified 11 Singlish discourse/pragmatic particles into three main categories: *contradictory*, *assertive*, and *tentative*.

Table 1

The table showing the Singlish discourse particles on an assertive continuum (Gupta, 1992, p.37).



These categories on the assertive continuum represent the level of commitment the speaker has to what they are saying. From the high end of the assertive continuum are the contradictory particles which are used to express disagreement and contradiction with something the conversation partner previously has said. Next, the assertive particles are used to express the speaker's positive commitment to what they are saying while lastly, the tentative particles express lesser positive commitment. Directives as labelled in Table 1 here are a group of assertive and tentative particles that express varying degrees of command/requests. This arrangement of particles by Gupta is useful because an additional particle-level analysis can be used to compare relative usage of these particles in gendered interactions. Thus, using corpus data drawn from actual speech and interactions, I hope to be able to capture Singlish particles in interactional use, which will better allow us to examine how solidarity and assertiveness can play out in same-gender and mixed-gender interactions.

Therefore, for this thesis, seeing how particles are an essential part of solidarity, I can expect there to be more usage of Singlish particles in same-gender dyads where speakers have less social distance with their interlocutors, compared to mixed-gender dyads. I would also expect men to use more Singlish particles than women as reported by Leimgruber and colleagues (2020), and for both men and women to use fewer solidarity markers in mixed-gender dyads due to the larger social distance between men and women:

H3a: Men will use more Singlish sentence-final particles than women.

H3b: In mixed-gender dyads, both men and women (M, F) will use fewer Singlish sentence-final particles than same-gender dyads (MM, FF).

2.5.5 Hypothesis 4: Language Style Matching in Singapore English

Ho and Chionh (2009) reported the use of personal pronouns, articles, determiners, among other miscellaneous lexical tokens in Singapore teenage blogs. Incidentally, LSM score is calculated based on some of these features among others (see Section 2.4.6 previously), wherein an index ranging from zero to one is derived, and the closer the value is to one, the more the two interlocutors match in linguistic style. Unfortunately, their work is purely a report of raw percentages, with no statistical analyses to show any significance nor theoretical underpinnings to form meaningful conclusions.

However, motivated predictions can still be made from the literature surrounding gendered interactions, especially with regards to the Difference approach in which the premise is that cross-gender interaction would be subjected to a cross-cultural effect, with increased formality and social distance between interlocutors of different genders. Therefore, this thesis expects same-gender dyads to match more in terms of style (LSM closer to one) than mixed sex dyads:

H4: Same-gender dyads (FF, MM) will have a higher LSM score than mixed-gender dyads (MF).

2.6 Positionality of This Paper

Firstly, this thesis does not contest or dispute the patriarchal view of men's domination of women in the society and certain areas of language, like the semantic derogation of women (Spender, 1985), that is, the language used to talk *about* women (Siddiqui, 2014). This characteristic of patriarchal systemic oppression women in history, as well as in modern society, is firmly recognised.

However, as Tannen (1993) clarifies, "Likewise, recognizing that men dominate women in our culture does not preclude the existence of patterns of communication that tend to typify women and men." (p.5). We have the belief that men and women speak differently, and an obsession with the gender difference (Talbot, 2019), so it is in our interest to investigate the presence (or absence) of this difference. While it is not in the purpose of this thesis to comment specifically on *why* men and women use language differently, there has been a longstanding importance for the documentation study of *how* language differs in male and female talk. This thesis takes a step further to look at how language differs in male and female *interactions* in same-gender and mixed-gender pairings, not just the speaker's gender *per se*.

Also, it is not the intention of this thesis to exclude certain communities of people, especially individuals that inhabit the LGBTQ+ spectrum, nor to perpetuate hegemonic heteronormative perspectives of gender, especially with regards to heterosexuality and cisgenderism. This is purely a result of the corpus data which limits gender to a binary choice for the participants. Therefore, unlike more qualitative studies in which linguistic expressions of masculinity/femininity can be examined on the individual level, the quantitative approach in this thesis utilises the reported genders of the participant to form models for the linguistic features that correlate with gender. Hence, this thesis serves an initial and primary look at gender and language in Singapore, which, in turn, can be used to inform future potential research on the same topic.

2.7 Significance of this Thesis

The aim of this thesis is to consolidate the areas of communication accommodation, language and gender, and Singapore English. It seeks to investigate the linguistic disparities between men and women in Singapore, consider how their language usage changes when conversing with individuals of the same or different gender, and at the same time taking into account possible intersecting social identities together with gender.

Currently, there is a paucity of research regarding the gendered interactions in Singapore. Given that gender consistently emerges as significant sociolinguistic variables when examining language variation within communities (Xu, 2015), it is crucial to explore how these gender-related differences manifest in language and compare them with existing theoretical frameworks. Additionally, it is archivally important to obtain a comprehensive overview of Singapore's sociolinguistic landscape on the topic of language and gender.

This thesis aims to examine and apply conclusions put forth by prior studies, utilizing a corpus-driven approach that rely on a local database with appropriate statistical modelling. By adopting this method, I attempt to evaluate the validity of these claims and their applicability to the Singapore context with a quantitative lens. This research aims to serve as one of the first study that systematically investigates the differences in language use between men and women in Singapore English using spoken data, and with varying social distances of interlocutors in mind.

In the current landscape of gender research, it may seem that studies analysing gender difference in speech runs the risk of being outdated and unnuanced, but ultimately, a large corpus study and analysis of a specific culture and language is extremely beneficial to the continuously growing language and gender literature, such as what Newman et al. did (2008) in their study on language and gender using 14,000 text samples. Large quantitative corpus studies can serve as a clear springboard for future work to be done on the topic and refine it. Emerging patterns illuminated in this thesis can be used to look into more specific language features or phenomenon through more targetted qualitative methods if a more detailed question is to be asked in the field of language and gender, and of Singaporean speech and interaction.

Works such as this thesis can allow us to open a window to shed light on the linguistic practices of local communities and the linguistic decisions they make while living in a very heavily gendered world. Borrowing the words of Eckert and McConnell-Ginet (2003),

> "After all, we don't just flop through the world, but we have plans — however much those plans may change from moment to moment. And these plans and the means by which we carry them out are strongly affected by gender." (p.6).

Chapter 3 Methodology

3.1 Approach

As Coates (2012) noted in her chapter on *Gender and discourse analysis*, the field of language and gender consists of two main strands. One is a part of discourse, and the other is on quantitative sociolinguistics. While Coates' chapter focused on the former, this thesis will focus on the latter, analysing the correlation and variation of language and gender differences, akin to Trudgill's (1974) early study on the variation of English in Norwich.

Thus, this thesis approaches gender and language variation using the combination of quantitative variationist methodology (Romaine, 2003) and corpus linguistics, in which verbal data is systematically collected, in the form of a corpus in this thesis, and examined for the relationship between social factors and linguistic variation. These linguistic variables are quantitative in nature and rest on the Principle of Accountability (Anderson et al., 2022), wherein every speech feature will be calculated with the same baseline — in this case, as a fraction of a million words uttered by the participant (see Section 3.6.1 *Extractability and Calculation of Dependent Variables*). It is in the interest of this thesis and the benefit of the research field to leverage on a large transcribed local database of conversations between and within men and women, in order to reveal patterns of language use in these contexts.

Statistical modelling, more specifically, linear mixed-effects models (LMMs) and linear regression models, will then be used to examine the relationship between the independent variables, mainly the gender of the participant, the conversation partner, and the dyad type they are in, together with the four dependent variables outlined in the previous chapter with the hypotheses.

3.2 National Speech Corpus (NSC) of Singapore

3.2.1 Overview of NSC

The data used in this thesis is obtained from the National Speech Corpus (NSC) by the Infocommunications and Media Development Authority (IMDA) of Singapore (see Koh et al., 2019). It is the first large-scale repository of spoken English in Singapore functioning as an important resource of open speech data that can be used for automatic speech recognition (ASR) research and speech-related applications. A total of 1,379 participants were recruited for the creation of this corpus, comprising Singaporeans of various gender, age, occupation, education, ethnic and linguistic backgrounds.

The NSC data is split into three parts: Part 1 is a phonetically balanced (PB) corpus of read sentences sourced from news sources that include phones, diphones, and triphones to document the sounds of Singapore English. Part 2 is a local words (LW) corpus that include readings of Singapore-related names, places, and other linguistic terms that are locally unique and not found in other parts of the Anglosphere. Part 3 is a conversation corpus in which participants are organised to into pairs to converse with each other — either face-to-face in the same room, or through a telephone in separate rooms, for a timespan of approximately 2 hours and 15 minutes (Koh et al., 2019). Since the aim of this thesis is to investigate spoken interactions between people, the data used henceforth will be only be drawn from Part 3 of the NSC.

3.2.2 Data Used in This Thesis: NSC Part 3

In Part 3 of the corpus, spoken interactions between every dyad have been transcribed by IMDA and associated vendors in the format of Praat's .TextGrid files (Boersma & Weenink, 2023) at the participant level. Every file contains the start and end times for words, silences, and non-speech sounds for the entire duration of the conversation.

In general, the content of the participants' conversations is divided into two segments — a diapix task segment and a free conversation segment. For the diapix segment, participants were each given slightly different pictures drawn from DiapixUK picture materials (Baker & Hazan, 2011) and instructed to find 12 differences without looking at each other's picture. This task stimulates speech and collaboration as dyads can only relay visual information verbally by communicating what they see in their own pictures with each other. Subsequently, in the free conversation segment, participants were provided conversation cards (*smol tok* for face-to-face dyads and *Hypothetically Fun* for separate room dyads) to get the conversation started. Dyads will converse with each other until the recording session ends. In the event that the deck of conversation cards has been exhausted, a list of suggested topics and questions are provided to further conversation until the time of the recording session is up.

3.3 Participants

A total of 912 individuals participated in Part 3 of the NSC and no subjects were repeated or participated in the recordings more than once. For the analyses conducted in this thesis, only 694 participants' data are used (N=694; 347 pairs) out of all the participants. In a nutshell,

Table 2 below shows the distribution of these participants in terms of the dyad type (same-gender or mixed gender dyad) and gender of individual (male or female).

Table 2

The number of individuals in same-gender pairing and mixed-gender pairing among friends.

Individuals in same-gender condition		**Individuals in mixed-gender condition**	
512		182	
Female (FF)	*Male (MM)*	*Female (F)*	*Male (M)*
288	224	91	91

The reason why 694 participants' data are used out of 912 recorded data is due to two reasons: procedural issues in the recording process, and the relationship between the participants. For the procedural issues, participants who deviated from the instructed order of procedure during recording of the NSC are removed from data processing to ensure consistency and accuracy of results (see Section 3.5.1 *Separating Conversations from Diapix Task*).

For the relationship of the participants, I decided to only take friend pairings into account for analyses, rejecting pairs who are either romantic couples, family, or strangers. The reasons for choosing friend pairs are two-fold: Firstly, we can see in Table 3 below that friends comprise most of the data collected in this portion of the NSC, which allow for the sufficiently large statistical power that is ideal for analyses.

Table 3

The number of individuals organised in terms of gender, the room condition the participants are in, and the relationship between participants.

	Same Room (482)				**Separate Room (416)**			
	Friends	*Couples*	*Family*	*Strangers*	*Friends*	*Couples*	*Family*	*Strangers*
Female (493)	182	49	26	0	197	6	12	21
Male (405)	164	49	12	0	151	6	2	21

Secondly, and more importantly, it is to set the scope for this thesis and exclude potential confounds in the analyses. Recalling Section 2.2.5 *Intersectionality* and 2.3.2 *Studies on Accommodation in Language and Gender*, it is wise to not include more intimate pairings in the analyses as there would be an additional relational complexity that needs to be considered for married couples (Fitzpatrick et al., 1995) and families (Hare-Mustin & Maracek, 1988). Friends are social groups in which are people know each other well enough, but not so intimately like partners and families, so by looking only at interactions between friends, the relationship type between participants can be controlled for, and gender differences can be observed without complications from factors such as couple identity or parent-child power dynamics as mentioned in the literature.

3.4 Tools and Software

The main computer programmes used for this thesis is the programming language R (R Core Team, 2022) in the integrated development environment R studio (Rstudio Team, 2020). These programmes are used for the large scale data manipulation and calculation in the data processing phase (described in Section 3.5 *Data Processing*) as well as creating the statistical models in the analyses seen in the Results chapter.

Two other computer programmes are used for obtaining some of the measures presented in this thesis. AntConc, the corpus analysis software developed by Laurence Anthony (2020), is used for the extraction and consolidation of words/tokens from the corpus that are relevant to the specific measure in question, and LIWC2015 (Pennebaker et al., 2015) is used to obtain the percentages of function words used in each transcript to be used to calculate LSM scores.

3.5 Data Processing

3.5.1 Separating Conversations from Diapix Task

With the method of collection in the Part 3 of the NSC data, there exists a potential issue due to the nature of talks in these dyadic interactions. In its attempt to elicit natural speech from Singaporean participants, the NSC utilise the help of different stimuli (diapix pictures and conversation cards/prompts) to help participants get started on interacting. However, the use of different stimuli to generate talk also introduced different types of talk in the interaction between the dyads, which should be considered before analysis. This is an issue because the diapix portion of the interaction would be considered a more task-oriented type of talk than the conversation portion and previous studies have shown that task-oriented communication has the potential to influence speakers to exhibit different linguistic behaviours compared to non-task-oriented speech (e.g., tag questions in McMillan et al., 1977; backchannels in Bilous & Krauss, 1988).

Therefore, to overcome this issue on different nature of talk and ensure more accuracy of the analyses and conclusions, the transcripts is split into two separate files — one containing speech from the diapix task and one from the free conversation. Three undergraduate students were recruited for this process, working with text (.txt) files of the transcripts that were previously extracted from the .TextGrid files using the rPraat package in R (Bořil & Skarnitzl, 2016). The student assistants were required to read through the transcripts and mark out the location in the transcript with three asterisk signs (***) where the diapix task ends and the conversation segment starts. Any oddities or deviations from the dipaix-conversation sequence in a transcript is flagged and passed on to me and, depending on the severity of the oddity, I will either accept the transcript or reject it for analysis. For example, oddities include instances where participants did the diapix task, switched to conversation, and then did the diapix task again; or ambiguous transitions between the diapix segment and the conversation segment. Following this marking and flagging procedure, 14 participants' transcripts were rejected out of a total of 912 collected.

After validating that all transcripts are eligible for analysis, I used R to split each participant's transcripts into two separate text (.txt) files and saved them in different folders. Due to the aforementioned issue with task-oriented speech versus conversational speech, and to keep the scope of this thesis more targeted, only the *conversation* transcripts will be used since the focus of this thesis is more interested in analysing conversations between dyads, instead of task-oriented speech. Additionally, since the conversations usually take more than an hour compared to the diapix which takes about about 15 minutes, the conversation transcripts will be more voluminous in duration and word count, which is advantageous in statistical analysis.

3.5.2 Extracting Word Count from Transcripts

In the NSC data, apart from the utterances, silences and non-speech sounds are also transcribed, meaning that determining the word count from the transcript is not as straightforward as counting the number of strings separated by spaces.

Hence, the data is cleaned up by removing silences (indicated by "<S>" and "<Z>" in the transcripts) and non-speech sounds such as coughs "(ppc)", laughter "(ppl)", breathing noises "(ppb)", and other unidentified sounds "(ppo)". This process is done in R by using regular expressions to capture instances of the appearances of these silence and non-speech markers. Subsequently, since non-word transcriptions are removed, the total number of words uttered by each individual during the course of the free conversation segment can be obtained by counting the number of items contained within word boundaries (e.g., white spaces, punctuation).

3.6 Dependent Variables

This section is focused on the dependent variables (DVs) outlined in the hypotheses above in Section 2.5 — what contributed to these measures, how they are obtained, and how they are calculated.

3.6.1 Extractability and Calculation of Dependent Variables

The reasons that the linguistic features were selected is based on of two points of consideration — relevance and extractability. Firstly, the dependent variables in question should be relevant to what this thesis seeks to investigate, which is that these variables should be able to effectively reflect the effects of gender pairings have on formality and solidarity in interactions as outlined in Section 2.5.1 (*Research Questions and Hypotheses*). Secondly, the features selected need to be easily segmentable into tokens to be counted and calculated. Since the NSC has *not* undergone part-of-speech tagging, there would be a greater reliance on tagging conventions idiosyncratic to the NSC. For example, round parentheses `()` are used for hesitation/filled pauses, squared parentheses `[]` for Singlish particles, and exclamation points `!!` for interjections, to name a few. Hence, backchannels, filled pauses, Singlish particles are features that fit these criteria of relevance and extractability, and thus chosen as variables to be analysed in this thesis. Unfortunately, the analysis for swearing was aborted due to censorship limitations imposed on the data (see 6.4 *Swearing and Corpus Limitations*).

For calculation of dependent variables, quantitative studies rests on the Principle of Accountability (Anderson et al., 2022) wherein each participant and conversation is measured along the same base rate. For the first three measures explored in this thesis (backchannelling, filled pauses, and Singlish particles), each participant's usage of features contributing to that measure is summed up and normalised to the denominator of a million words. Taking Singlish particles as an example, the usage of all particles is first counted, then divided by the total word count of the participant, and multiplied by one million words to get the count per million words for total particles used. The reason why count per million words is used over other base numbers such as per 1,000 words (Leimgruber et al., 2020) is to eliminate instances of false zeros that can cause zero-inflation in the count data. To illustrate this, for example, if a participant were to use 0.04 particles per 1,000 words, count data would round the value off to zero per 1,000 words. However, if count per million is used instead, we can see that the particle usage of that same participant is now 40 per million words. Thus, using a smaller denominator would result in zero-inflation that can distort the data and analysis results and a larger denominator of a million is used instead.

The count per million data of each of these three measures is then fitted to a linear mixed-effects model with gender and dyad type as the fixed effects, and participant pairs included as a random effect to capture variations that are idiosyncratic to each pair of participants (more is explained in Chapter 4 *Results*).

3.6.2 Measuring Backchannelling

The category of backchannels can be a very broad and nebulous category (Duncan & Niederehe, 1974; Drummond & Hopper, 1993). For example, one important backchannel, *okay*, was excluded in the analysis. Although it was initially defined in Yngve's (1970) pioneering work where the term *back channels* was originally coined, Drummond and Hopper (1993) noted the issue of over-capturing if all tokens of brief acknowledgment are lump into a single category of "backchannels" or "minimal responses". Hence, they only took into account three of such tokens — *yeah*, *uh huh* and *mm hm* in their work. Additionally, while Yngve describes *okay* as a backchannel, Fellegy (1995) treated it as a discourse marker instead. These examples show the complexity of classifying these certain sounds, words, and phrases as backchannels.

Therefore, for the analysis of supportive communication in this thesis will focus on these terms — *mm*, *mmhmm*, *uh huh*, *yeah*, and *right* for analysis, drawing from several important works on the subject (e.g. Yngve, 1970; Drummond & Hopper, 1993; Heinz, 2003; Mulac, 1998, Mott & Petrie, 1995; Fellegy, 1995; etc.). Additionally, I also identified *ya* (a variation of *yeah*) as a common backchannel used by Singaporeans, and sometimes accompanied by Singlish particles for emphatic agreement, e.g. *ya lor* (Goh, 2016).

In the extraction process, it is important to identify instances of these words that solely appear within a turn because a backchannel is a signal to indicate that one is listening and the conversation partner who is speaking still has the floor. Therefore, appearances of these backchannels should not encompass more content than the backchannel itself. For example, “mmhmm” (single backchannels) and “ya ya” (stacked backchannels) were accepted as an instance of backchannelling, but “ya but he’s not wrong though” was not accepted as the conversational floor is considered to have been taken away.

Identifying these backchannels was done using the computer programme Antconc (Anthony, 2020) to control for this criterion using targeted regular expressions and extract the eligible tokens from the transcripts.

3.6.3 Measuring Filled Pauses

Continuing in this approach, Antconc is also used to extract instances of tentative language use. However, as explained in Section 2.5.3 in the previous chapter, tag questions will not be examined because of its atypical usage in Singapore English where an invariant tag is favoured in this variety of English. Additionally, the aforementioned complexity of terms like *you know*, *I mean*, *I think*, and *sort of* were excluded in my analysis. Instead, a meaningful proxy for tentative language that is more measurable can present itself through verbalisation of filled pauses (*uh* and *um*) which have been shown to vary with gender (Anton, 2011) and formality contexts (Schachter et al., 1991; Clark & Foxtree, 2002).

In the NSC, filled pauses are marked in round parentheses `()`. Hence, using Antconc, the presence of *(uh)* and *(um)* can be easily captured, along with their alternate spellings *(er)*, *(err)*, and *(erm)*.

3.6.4 Measuring Singlish Particles

Wee (2004) has noted that the actual number of Singlish discourse particles is unclear, which is evident in the studies across time. For example, Gupta (1992), Wee (2004), Ler (2006), Botha

(2018), Lee (2018), Leimgruber et al. (2020), and most recently Boo, Lee & Tan (2023) have respectively identified 11, 8, 10, 16, 18, 10, and 16 particles in Singapore English. Therefore, for this thesis, after considering the various literature surrounding Singlish particle research, I have decided to look at 18 of these particles, drawing from an amalgamation of the four most recent studies stated above — Botha (2018), Lee (2018), Leimgruber et al. (2020), and Boo et al. (2023). The 18 particles taken into consideration are *ah*, *lah*, *eh*, *lor*, *leh*, *what*, *sia*, *meh*, *mah*, *one*, *hor*, *ha*, *she*, *bah*, *liao*, *wor*, *siol*, and *nia*.

Since Singlish particles were transcribed in the NSC transcripts in square brackets (e.g. [lah], [meh], [sia], etc.), an R script was written to search through all transcripts and tally up the frequency of all 18 particles used for each participant and calculated to the base of per million words.

3.6.5 Measuring Language Style Matching

Lastly, for Language Style Matching (LSM) scores, transcripts of each participant are fed into LIWC2015 (Pennebaker et al., 2015). The programme processes each transcript individually and outputs 93 measures, of which only nine of them are relevant:

1) "ppron" – personal pronouns
2) "ipron" – indefinite pronouns
3) "article" – articles
4) "prep" – prepositions
5) "auxverb" – auxiliary verbs
6) "adverb" – common adverbs
7) "conj" – conjunctions
8) "negate" – negations
9) "quant" – quantifiers

As described in Section 2.4.6 *Function Words: Language Style Matching,* a formula is used to calculate the individual LSM score of each particular parameter before taking the average across all nine parameters to obtain the overall LSM score for each pair of participants. The formulae for calculating LSM score is carried out as follows (Gonzalez et al., 2010; Ireland et al., 2011):

$$LSM_{prep} = 1 - \frac{|prep_1 - prep_2|}{prep_1 + prep_2 + 0.0001}$$

$$LSM = avg\begin{pmatrix} LSM_{prep} + LSM_{article} + LSM_{auxverb} + LSM_{adverb} \\ LSM_{conj} + LSM_{ppron} + LSM_{ipron} + LSM_{negate} \end{pmatrix}$$

Since only a single LSM score is generated from the calculation of two transcripts obtained from the pair of participants, instead of having 694 points of data for each participant like the previous variables, the number of data points is halved for LSM scores (N=347).

3.7 Independent Variables

3.7.1 Interdependency of Independent Variables

Before moving on to the Results chapter, there is something important to note here regarding the independent variables (IVs) that are used as predictive factors for the dependent variables (DVs) outlined above.

Since this thesis is interested in language variation in gendered interactions, as mentioned in Section 2.5.1 (*Research Questions and Hypotheses*) it will look at:

(1) **Speaker gender**,
(2) **Conversation partner's gender**, and
(3) **Dyad type** the speakers are in (same-gender or mixed-gender).

However, these three variables are interdependent, meaning that only two of these variables are required to conclude findings for the third variable. Table 4 below shows the interdependency of the three variables — for example, if the speaker gender is female, and the partner gender is also female, they would be a same gender dyad; Or if the speaker gender is male and the dyad type is a mixed-gender dyad, then the corresponding conversation partner would necessarily be female; etc. Essentially, these are the four distinct groups of individuals this thesis is looking at, represented by each row shown in Table 4.

Table 4

A table showing the interdependency of speaker gender, conversation partner gender, and dyad type.

(1) Speaker Gender	(2) Partner Gender	(3) Dyad Type
F Speaker	M Partner	Mixed
M Speaker	F Partner	Mixed
F Speaker	F Partner	Same
M Speaker	M Partner	Same

Table 4 is important to note because it informs the statistical model used in the analyses to avoid including an IV that is codependent with the other two IVs. The model only needs two variables to draw conclusions for the third. In other words, for the three IVs of (1) speaker gender, (2) partner gender, and (3) dyad type, the *interaction effect* between any two of the IVs is the same as the *main effect* of the third IV — as an example, an interaction effect between speaker gender and dyad type would mean there is a main effect of partner gender on the dependent variable.

3.7.2 Accommodation Model vs. Partner Model for Result Interpretation

This brings us to the two ways the results will be visualised — the *Accommodation Model* and the *Partner Model*. Firstly, Accommodation Model is based on CAT where the pattern of linguistic variation is visualised as an accommodation effect due to in-group/out-group differences between same-gender and mixed-gender dyads. Secondly, the *Partner Model* is based on patterns of linguistic variation influenced by the gender of the interlocutor (Hancock & Rubin, 2015), in which a certain linguistic feature's usage is affected due to the mental schemas participants have regarding the gender of their interlocutor (explained in Section 2.3.4 *Speaker Gender vs. Conversation Partner Gender*).

For the Accommodation Model, the DVs are modelled against (1) speaker gender and (3) dyad type, since the Accommodation Model is interested in whether a speaker of a certain gender will use language differently in the same-gender context or mixed-gender context. Whereas for the Partner Model, the DVs are modelled against (1) speaker gender and (2) partner gender since it is interested in whether a speaker will use language differently when interacting with a female conversation partner or male conversation partner. Table 5 below shows how the independent variables fit in each of the two statistical models.

Table 5

A table showing the three independent variables and their roles in the linear mixed models when interpreting results.

	Accommodation Model	**Partner Model**
Main effects	Speaker Gender & Dyad Type	Speaker Gender & Partner Gender
Interaction effect	Partner Gender	Dyad Type

The important thing to note here is that, essentially, these two models are theoretically and statistically the exact same models, only that they present and visualise the results differently, and will be particularly helpful in contextualising the graphs that are being plotted. In the results presented in the next chapter, the values of the statistical modelling of the two models are presented as the same table — such as F-tests, estimated values, standard errors, and p-values. The differences between the two models only lie in the visualization tools such as graphs and trend lines, which will be grouped differently to show a different perspective.

Chapter 4 Results

In this section, I will be presenting the results for the four measures delineated in my hypotheses, namely, backchanneling, filled pauses, swearing, Singlish particle usage and Language Style Matching. Each measure will be tackled in their own subsection, presented as the dependent variable against the two models — Accommodation Model and Partner Model.

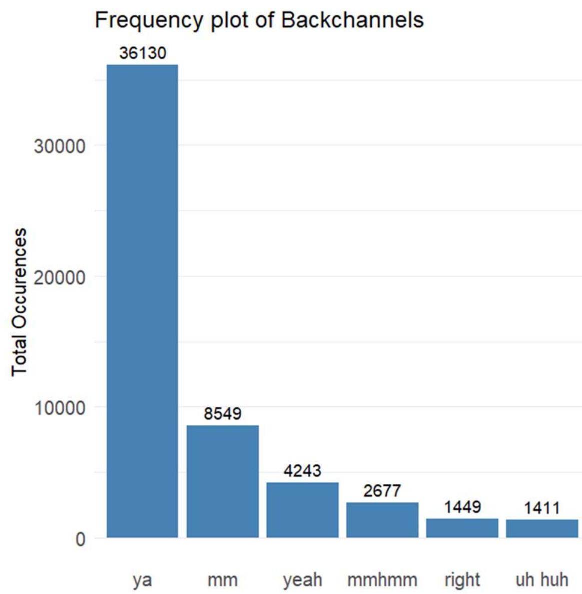
4.1 Backchannelling

4.1.1 Descriptive Statistics

Taking into account the words *mm*, *mmhmm*, *uh huh*, *yeah*, *right* and *ya*, a total of 54,459 backchannels are used by the 694 participants. Figure 1 on the right shows the total number of occurrences for all the minimal responses in order of decreasing frequency — with *ya* taking up most of the occurrence of minimal responses (66.3%).

Figure 1

A frequency graph of all backchannels uttered by every participant, in decreasing order.



4.1.2 Model Fitting Results

Linear mixed-effects modelling (LMM) using the *lme4* package (Bates et al. 2015) is used to assess the effects of speaker gender and the dyad type on rates of backchannelling based on the Accommodation Model. The model includes the participant pairs as a random effect to account for each unique pair to be a possible source of random variation:

```
lmer(log(backchannelling) ~ Gender*dyad + (1|Pair)
```

An ANOVA test confirmed that a mixed-effects model with the pairs as a random effect is a significantly better model than a regular linear regression without this random effect, and hence, all subsequent models will have participant pairs as a random effect to account for pair-level variations.

Similarly, for the Partner Model, an LMM assessing the effects of speaker gender and the conversation partner gender on rates of minimal responses are also carried out:

```
lmer(log(backchannelling) ~ Gender * Partner.Gender + (1|Pair)
```

A logarithmic transformation has been done on the rates minimal responses to ensure that normality assumptions for the residuals hold for the model. Appendix A contains the normal quantile-quantile (Q-Q) plots for all measures discussed in this chapter.

Figure 2

Results of ANOVA of the linear mixed effects model for backchannelling

Predictor	df_{Num}	df_{Den}	***Sum Square***	***Mean Square***	***F-value***	***p-value***	
Speaker Gender	1	686	3.94	3.94	6.65	0.010	*
Partner Gender	1	686	0.04	0.04	0.068	0.378	
Dyad Type	1	344	0.46	0.46	0.779	0.794	

ANOVA F-test results of the mixed models showed a significant main effect of speaker gender [$F(1, 686.4) = 6.65$, $p = 0.010$*], but not dyad type [$F(1, 344) = 0.779$, $p = 0.40$] nor partner gender [$F(1, 344) = 0.068$, $p = 0.79$)].

Subsequent pairwise comparisons of the independent variables were conducted with the use of estimated marginal means (also called least-squares means) using the *emmeans* package (Lenth, 2023). Estimated marginal means (EMMs) was used because they represent the average response of the dependent variable after adjusting for the effects of other variables in the model, and thus can be effectively used to carry out post-hoc pairwise comparisons between participants.

Table 6 below shows the estimated marginal means for each of the four groups of participants this thesis is interested in: female speakers with male partners (mixed-gender), male speakers with female partners (mixed-gender), female speakers with female partners (same-gender), and male speakers with male partners (same-gender). In order to meaningfully interpret these values, Figure 3 and Figure 4 on the next page plots these EMMs for the four groups of individuals for visualisation.

Table 6

A table showing the estimated marginal means of each of the four group of participants.

Speaker Gender	Partner Gender	Dyad Type	*Estimated Marginal Mean*	*Standard Error*	*df*	*Lower Confidence Level*	*Upper Confidence Level*
F Speaker	M Partner	Mixed	4773	501	590	3883	5865
M Speaker	F Partner	Mixed	3826	402	590	3113	4702
F Speaker	F Partner	Same	5121	359	344	4462	5877
M Speaker	M Partner	Same	4275	340	344	3657	4998

Figure 3

Graph of EMMs of backchannels used per million words, grouped by dyad type.

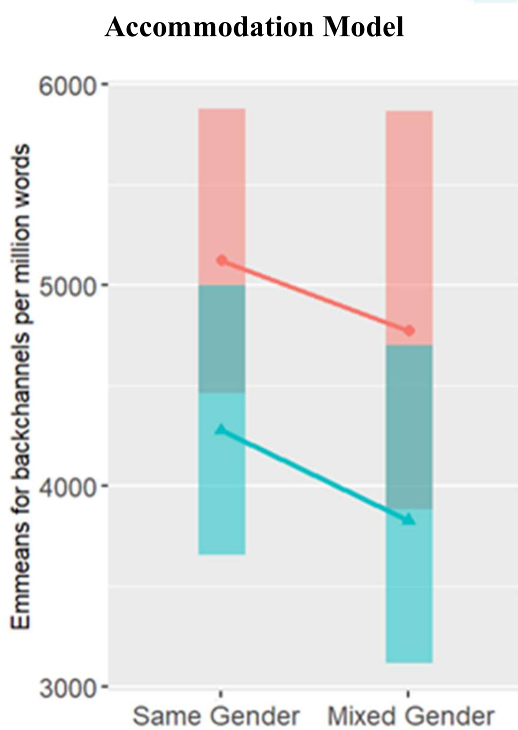


Figure 4

Graph of EMMs of backchannels used per million words, grouped by partner gender.

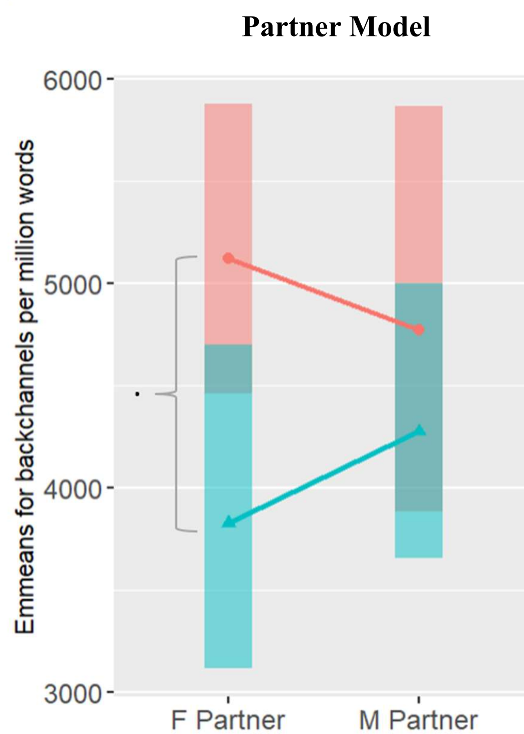


Table 7 below shows the pairwise comparisons between the four groups. The contrast between same-gender female group and the mixed-gender male group is noteworthy (marked by a period "." in Table 7 and Figure 4). Both these two groups feature female conversation partners as interlocutors, which is relevant to the Partner Model seen in Figure 4 and explained further below.

Table 7

A table showing the six pairwise comparisons of each of the four groups of participants.

Pairwise Contrasts	*Estimate*	*SE*	*df*	*t-ratio*	*p-value*	
F Speaker F Partner - M Speaker F Partner	1295	538	344	2.405	0.0781	.
F Speaker F Partner - F Speaker M Partner	348	616	344	0.565	0.9424	
F Speaker F Partner - M Speaker M Partner	845	494	344	1.712	0.3192	
M Speaker F Partner - F Speaker M Partner	-947	498	590	-1.902	0.2283	
M Speaker F Partner - M Speaker M Partner	-449	526	344	-0.855	0.8280	
F Speaker M Partner - M Speaker M Partner	497	605	344	0.822	0.8442	

Interpreting the results from Figure 3 and Figure 4 which plots the EMMs for the output of this linear mixed-effects model, it can immediately be seen in the graphs that there are the generally higher rates of backchannels used by female speakers (pink) than male speakers (blue), as evidenced by the main effect of speaker gender reported earlier [$F(1, 686.4) = 6.65$, $p = 0.010$*]. This supports Hypothesis 1a which predicted the use of backchannels as a tendency of feminine speech style, used more frequently by women than men.

However, looking at the trendlines in Figure 3 (Accommodation Model plot), it shows a *decrease* in usage of backchannels in mixed-gender dyads for both male and female speakers. Even though these decreases in backchannel usage in mixed-gender dyads for both genders are not significant (Table 7), their downward trend is not expected, going against the convergence and increase predicted in Hypothesis 1b.

In Figure 4 (Partner Model plot), this deviation from expectation can be seen more obviously — when conversing with a female partner, women use *more* backchannels while men use *less* backchannels with a female partner, with a difference that is close to significance [$t(344) = 2.405$, $p = 0.078$.]. Therefore, looking at Accommodation Model and Partner Model, they both show the unexpected linguistic behaviour where men did not accommodate to women's style of speech by using more backchannels.

4.1.3 Interim Conclusions for Backchannelling

Linear mixed-effects modelling conducted in this section showed that there is a significant main effect of speaker gender on backchannel usage, in which women use more backchannels than men, supporting Hypothesis 1a and in agreement with most of the literature's findings that women engage in more supportive communication. On the other hand, Hypothesis 1b is contradicted. While H1b predicted the convergence of backchannel usage for female speakers and male speakers in mixed-gender dyads with *increased* usage from genders, the model refutes this hypothesis by showing that speakers of both genders *decreased* backchannel usage in mixed-gender contexts (Figure 3), effectively going against CAT's prediction that instances of supportive communication like backchannelling *should increase* for more socially distant interactions like mixed-gender interactions.

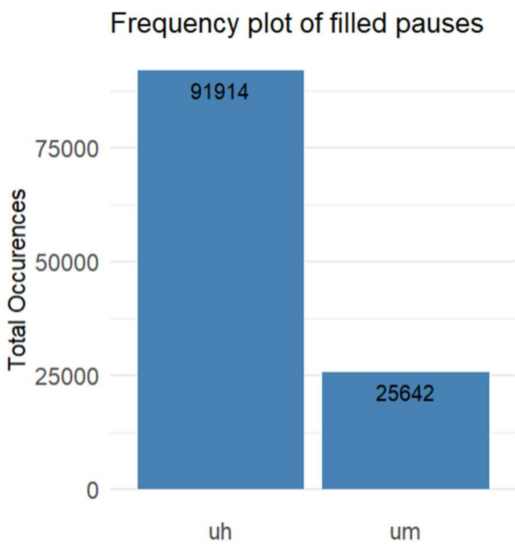
Curiously, the Partner Model in Figure 4 showed that men using less backchannels than women while talking to female conversation partner with marginal significance. This also runs contrary to expectations because CAT predicts that men use more supportive communication with women and accommodate to women's higher tendency to use backchannels.

4.2 Filled Pauses

4.2.1 Descriptive Statistics

Figure 5

Frequency plot of total filled pauses, (uh) and (um), used in the corpus data.



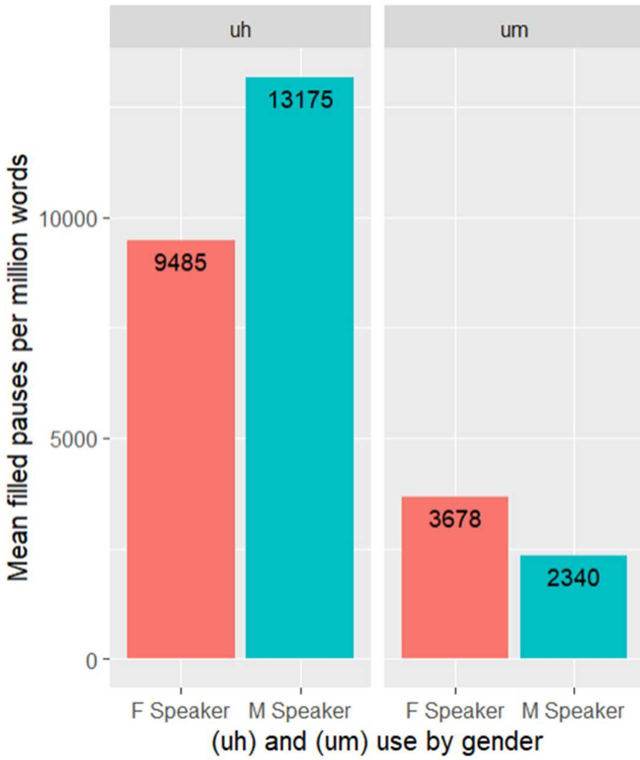
A total of 117,556 filled pauses of *uh* and *um* are captured in the NSC corpus of friends' interactions. Figure 5 on the left shows the distribution of filled pauses used, with 91,914 tokens of *uh* and 25,542 tokens of *um*. It can be seen that the use of *uh is* overwhelmingly more than the use of *um* with the former being about 350% more frequent than the latter.

This pattern persists when looking at gender differences for the use of *uh* and *um* between men and women, where both men and women use *uh* overwhelmingly more than *um*, which can be seen on the next page in Figure 6 showing the mean use of filled pauses per million words for speaker gender and the filled pause type.

Even though both men and women were shown to use more *uh* than *um*, there were differences within each filled pause type with respect to speaker gender. Men (M = 13175, SD = 12197) have a higher mean usage than women (M = 9485, SD = 9653) for the filled pause *uh*, while women (M = 3678, SD = 4732) have a higher mean usage than men (M = 2340, SD = 3408) for the filled pause *uh*. This reflects the literature on filled pauses that also observed gender differences in usage of the *uh* and *um* (see section 2.4.3) with men using more *uh* than women, and women using more *um* than men. In line with these findings, the *um*/*uh* ratio in this NSC data set shows that women have a 2.45 times higher *um*/*uh* ratio than men, validating Acton's (2011) study and Liberman's (2014) observation that *um* is a filled pause more likely to be utilised by women.

Figure 6

A bar graph showing the mean filled pauses used per million words for uh and um across female and male speakers.



These trends, while interesting and corroborates with previous findings, fall outside the scope of this thesis, and more importantly, these observations have to be taken with a grain of salt as the variation within each group is extremely large as evidenced by the large standard deviations.

More crucially the next subsection will explore how filled pauses as a whole varies in interaction with respect to the speaker gender, partner gender, and dyad type.

4.2.2 Model Fitting Results

Similar to the previous section on backchannelling, a linear mixed-effects model is used on the log-transformed values of filled pauses per million words, with speaker gender and the dyad type as fixed effects and participant pairs as a random effect,

```
lmer(log(filledpauses) ~ Gender * dyad + (1|Pair))
```

and an LMM assessing the effects of speaker gender and the conversation partner gender on rates of filled pauses are also carried out:

```
lmer(log(filledpauses) ~ Gender * Partner.Gender + (1|Pair))
```

Table 8

Results of ANOVA of linear mixed effect model for filled pauses.

Predictor	df_{Num}	df_{Den}	*Sum Square*	*Mean Square*	*F-value*	*p-value*	
Speaker Gender	1	684	1.49	1.49	3.72	0.054	.
Partner Gender	1	684	1.42	1.42	3.54	0.060	.
Dyad Type	1	344	0.005	0.005	0.01	0.914	

ANOVA F-test results of the mixed models as seen in Table 8 showed that there are no significant effects in any of the predictors. However, the effect of speaker gender [$F(1, 684.5) = 3.72$, $p = 0.054$] and partner gender [$F(1, 684.5) = 3.54$, $p = 0.060$] have on the rates of filled pauses used trend towards significance and cannot be disregarded yet.

This can be better explained by plotting the EMMs estimated by the model. Table 9 below shows the estimated marginal means for the four groups of participants and can be seen graphed in Figure 7 and Figure 8 on the next page, while Table 10 further down tabulates the pairwise comparisons for each of the groups.

Table 9

A table showing the estimated marginal means of filled pauses used per million words for each of the four group of participants.

Speaker Gender	**Partner Gender**	**Dyad Type**	***Estimated Marginal Mean***	***Standard Error***	***df***	***Lower Confidence Level***	***Upper Confidence Level***
F Speaker	M Partner	Mixed	10609	905	596	8972	12545
M Speaker	F Partner	Mixed	10641	908	596	8999	12582
F Speaker	F Partner	Same	9330	528	344	8346	10429
M Speaker	M Partner	Same	11886	763	344	10476	13486

Figure 7

Graph of EMMs of filled pauses used per million words, grouped by dyad type.

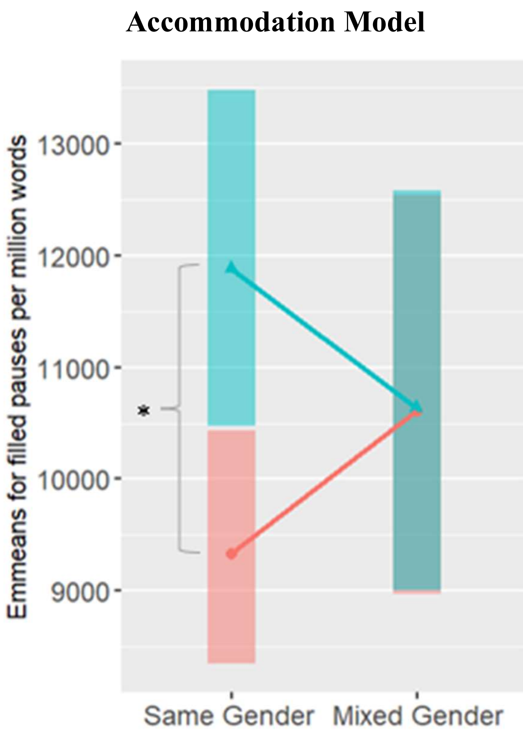


Figure 8

Graph of EMMs of filled pauses used per million words, grouped by partner gender.

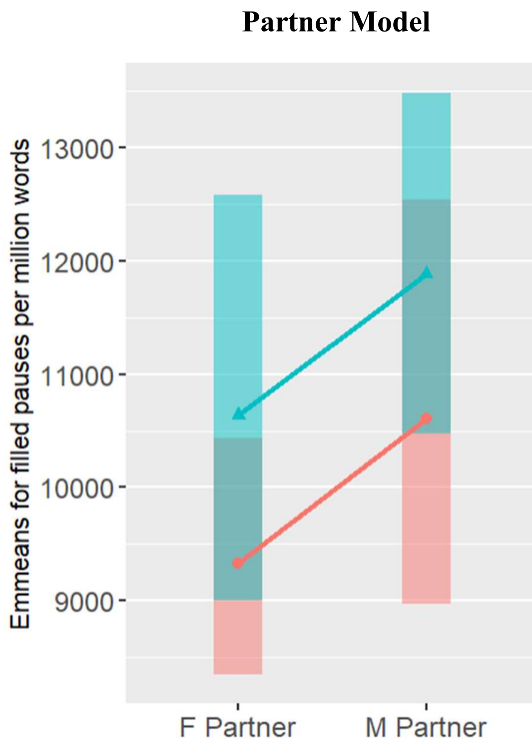


Table 10

A table showing the six pairwise comparisons of filled pauses used per million words for each of the four groups of participants.

Pairwise Differences	*Estimate*	*SE*	*df*	*t-ratio*	*p-value*	
F Speaker F Partner - M Speaker F Partner	-1331	1050	344	-1.248	0.5966	
F Speaker F Partner - F Speaker M Partner	-1279	1048	344	-1.221	0.6142	
F Speaker F Partner - M Speaker M Partner	-2556	928	344	-2.754	0.0314	*
M Speaker F Partner - F Speaker M Partner	31.9	998	596	0.032	1.0000	
M Speaker F Partner - M Speaker M Partner	-1245	1186	344	-1.049	0.7204	
F Speaker M Partner - M Speaker M Partner	-1277	1184	344	-1.078	0.7031	

While the effect of speaker gender on filled pauses usage presented earlier (Table 9) did not exactly fall below the 5% confidence level ($p = 0.054$), pairwise comparisons of groups shown in Table 10 above show that there are significant differences specifically between same-gender female and same-gender male speakers [$t(344) = -2.754$, $p = 0.0314$*]. This means that the analysis still gives some support to Hypothesis 2a that there is an effect of speaker gender on filled pauses usage where men use more filled pauses than women, and this effect is especially apparent in same-gender contexts.

Figure 7 (Accommodation Model plot) above illustrates this significant difference between women in same-gender dyads and men in same-gender dyads. Additionally, following the trend lines to compare with mixed-gender dyads, there is a mutual convergence to practically the same mean for men and women in mixed-gender dyads [$t(597) = 0.032$, $p = 1.00$]. This convergence counters Hypothesis 2b which predicted an overall *decrease* in filled pauses usage for both gender in mixed-gendered interactions due to speakers adopting more formality with increased social distance. While the results here agree that men do *decrease* their usage of filled pauses in mixed-gender dyads, they do not show the same for women. Instead, female speakers *increased* their rates of filled pauses to the same level as men in mixed-gender dyads.

However, interestingly, Figure 8 (Partner Model plot) shows another interpretation of the results. It was also shown earlier in Table 9 above that conversation partner gender has a main effect very close to significance [$F(1, 684.5) = 3.54$, $p = 0.060$], and this is shown in the graph as an overall increasing trend in filled pauses usage when the conversation partner is male compared to the condition where the conversation partner is female. This means that in the context where the conversation partner is male, it is likely that speakers will generally use more filled pauses when interacting with them.

4.2.3 Interim Conclusions for Filled Pauses

For this analysis on filled pauses usage, linear mixed-effects modelling showed that there are marginal main effects of speaker gender and partner gender on filled pause usage. Results show that male speakers likely use more filled pauses than female speakers, and also that conversations with male interlocutors brought about more filled pause usage from speakers.

This means that Hypothesis 2a is partially supported, especially in the case of same-gender dyads where the differences between male speakers and female speakers become highly significant and more pronounced. However, convergence took place from same-gender to mixed-gender dyads in which men reduce their use of filled pauses, and women increased their use of filled pauses. This refutes Hypothesis 2b's prediction that both speakers will reduce their use of filled pauses in mixed-gender contexts.

The Partner Model also reflects the marginal main effect of partner gender, wherein conversations involving male partners are characterised by higher use of filled pauses by speakers, and conversations involving female partners show a lower use of filled pauses. This gives us an indicator that, instead of strictly being a marker of formality and social distance, filled pauses is also shown to be a feature of speech associated with men's style of speaking (see Section 5.2.1 *Socialisation Strategies in Closer Social Distances* for discussion).

4.3 Use of Singlish Particles

4.3.1 Descriptive Statistics

Table 11 on the right shows the total number of occurrences for each of the 18 particles of interest, together with the percentages of how much each particle was used in the corpus as a proportion of all particles used.

In order to better visualise the frequency of use of Singlish particles, Figure 9 on the next page is a bar chart showing the number of occurrences of each particle, a parallel to the bar chart presented in Botha's (2018) paper, featured below in Figure 10.

Looking at both figures, we can see similarities of particle distribution between the NSC and Botha's (2018) data — particles *ah*, *lah*, *eh*, *lor*, and *leh* are the five most commonly used particles, accounting for the majority of particle usage. However, the slight difference is that these five most frequent particles accounts for a much higher percentage of particles used in this corpus (91.78%) compared to Botha's study (62.9%). The difference between these two percentages could be a due to the differences in data type and data volume of the two corpora — the NSC features *spoken* data of 6,748,062 words over 694 participants, while Botha uses *text messaging* data with 300,000 words over 57 participants.

Table 11

A table showing the frequency of each particle used in the NSC data.

Particles	***Total occurrences***	***Percentage of particles***
ah	60898	39.9%
lah	50449	33%
eh	17645	11.6%
lor	6716	4.4%
leh	4399	2.88%
what	2964	1.94%
sia	2243	1.47%
meh	1658	1.09%
mah	1772	1.16%
one	1461	0.96%
hor	1357	0.889%
ha	848	0.556%
seh	73	0.048%
bah	74	0.048%
liao	53	0.035%
wor	18	0.012%
siol	14	0.009%
nia	6	0.004%

Figure 9

Frequency plot of particle use for each particle in the NSC with total of 152,648 particle tokens (N=694).

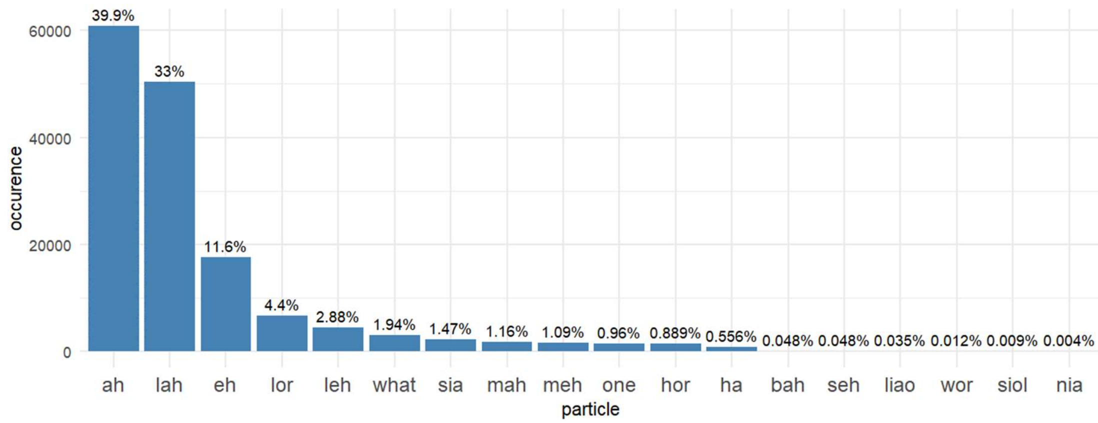
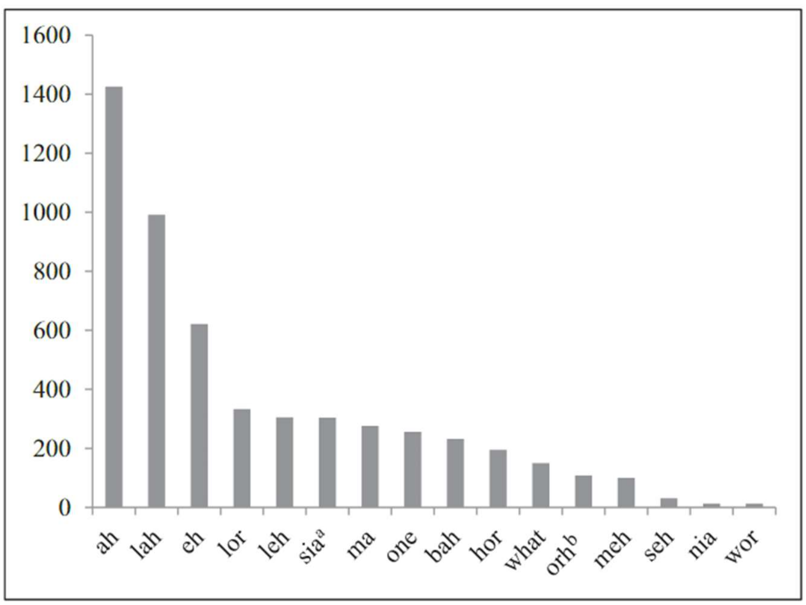


Figure 10

Frequency plot of particle obtained from Botha (2018) with total of 5,355 particle tokens.



Particle	In %
ah	26.6%
lah	18.5%
eh	11.6%
lor	6.2%
leh	5.7%
sia[a]	5.7%
ma	5.2%
one	4.8%
bah	4.3%
hor	3.6%
what	2.8%
orh[b]	2.0%
meh	1.9%
seh	0.6%
nia	0.2%
wor	0.2%

N=5355

a *siak* and *sial* also used.
b status as particle not confirmed, and the orthography for this particle was checked against the text data.

4.3.2 Model Fitting Results

Similar to the previous sections, an LMM is used, but in this instance on the square-root-transformed values of filled pauses per million words, with speaker gender and the dyad type as fixed effects and participant pairs as a random effect,

```
lmer(sqrt(particles) ~ Gender * dyad + (1|Pair))
```

and an LMM assessing the effects of speaker gender and the conversation partner gender on rates of filled pauses are also carried out:

```
lmer(sqrt(particles) ~ Gender * Partner.Gender + (1|Pair))
```

The reason that a square root transformation was used here instead of log transformation because square root transformations are able to handle zero values while at the same time still keep the residuals in the data relatively normally distributed (Appendix A), because unlike the previous DVs, the data on Singlish particles contain instances where some participants used no particles at all throughout the conversation.

Table 12

Results from ANOVA of the linear mixed effects model for Singlish particles

Predictor	df_{Num}	df_{Den}	***Sum Square***	***Mean Square***	***F-value***	***p-value***	
Speaker Gender	1	616	8935	8935	8.47	0.0037	**
Partner Gender	1	686	1534	1534	1.45	0.228	
Dyad Type	1	616	3419	3419	3.24	0.072	.

ANOVA F-test results of the mixed models showed that there is a strong main effect of speaker gender [$F(1, 616) = 8.47$, $p = 0.0037$**], and on the other hand, no main effects for partner gender [$F(1, 616) = 1.45$, $p = 0.23$] and dyad type [$F(1, 344) = 3.24$, $p = 0.073$]. However, it is worthy to note that the main effect of dyad type is very close to significance.

Table 13 on the next page tabulates the EMMs of particles used per million words for the four groups of participants that were estimated from the model. These values are plotted in Figure 11 and Figure 12 below it.

Table 13

A table showing the estimated marginal means of Singlish particles used per million words for each of the four group of participants.

Speaker Gender	Partner Gender	Dyad Type	*Estimated Marginal Mean*	*Standard Error*	*df*	*Lower Confidence Level*	*Upper Confidence Level*
F Speaker	M Partner	Mixed	15889	1566	462	12961	19115
M Speaker	F Partner	Mixed	17745	1655	462	14642	21145
F Speaker	F Partner	Same	17693	1211	344	15391	20154
M Speaker	M Partner	Same	22589	1551	344	19641	25743

Figure 11

Graph of EMMs of Singlish particles used per million words, grouped by dyad type.

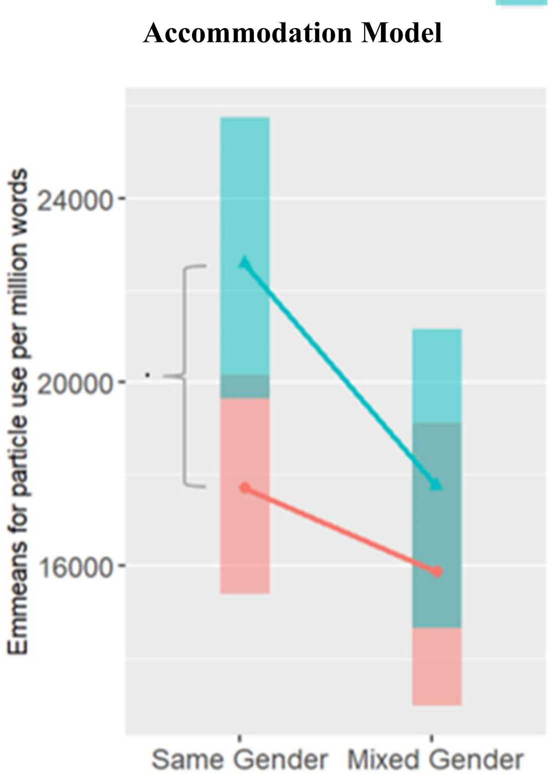
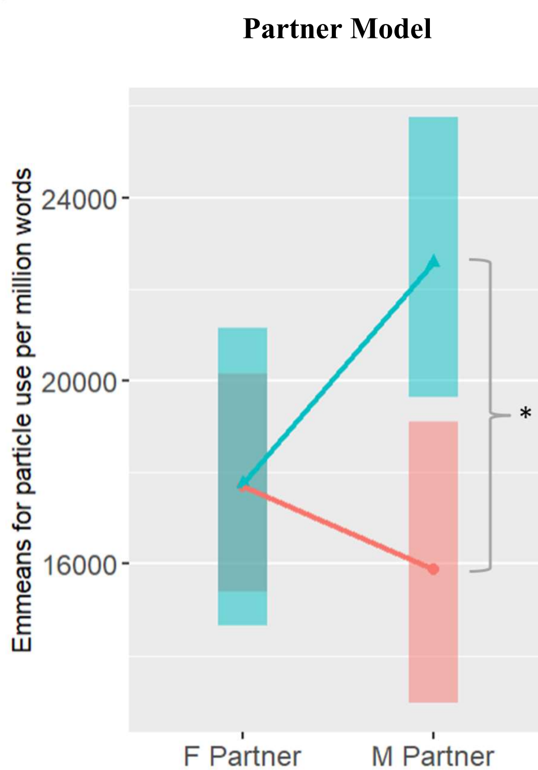


Figure 12

Graph of EMMs of Singlish particles used per million words, grouped by Partner Gender.



Pairwise comparisons are also done for these four groups of participants and are presented in Table 14 below. Two pairwise differences are of note — the differences between same-gender female dyads and same-gender male dyads (p = 0.064), as well as participants whose conversation partner is male (p = 0.0135*).

Table 14

A table showing the six pairwise comparisons of Singlish particles used per million words for each of the four groups of participants.

Pairwise Differences	*Estimate*	*SE*	*df*	*t-ratio*	*p-value*	
F Speaker F Partner - M Speaker F Partner	-52.2	2050	344	-0.025	1.0000	
F Speaker F Partner - F Speaker M Partner	1803.4	1979	344	0.911	0.7898	
F Speaker F Partner - M Speaker M Partner	-4896.2	1968	344	-2.488	0.0635	.
M Speaker F Partner - F Speaker M Partner	1855.6	1251	462	1.483	0.4486	
M Speaker F Partner - M Speaker M Partner	-4844.0	2268	344	-2.136	0.1439	
F Speaker M Partner - M Speaker M Partner	-6699.6	2204	344	-3.040	0.0135	*

Firstly, in the Figure 11 and Figure 12 above, plotting the estimated marginal means (Table 13) reveals that male speakers generally used Singlish particles more frequently than female speakers. With the highly significant main effect of speaker gender reported earlier in Table 12 [$F(1, 616) = 8.47$, $p = 0.0037$**], we can thus conclude that Hypothesis 3a is fully supported.

While the pairwise comparison between same-gender male dyads (MM) and same-gender female dyads (FF) is not technically significant [$t(344) = -2.49$, $p = 0.064$], it is extremely close. With the main effect of speaker gender being so highly significant in the model, it lends probable justification to distinguish the MM and FF groups in their use of Singlish particles, with the same-gender male pairs using more than the female pairs. Subsequently, looking at Figure 11 (Accommodation Model plot), the trendlines from the same-gender dyads to mixed-gender dyads show a decreasing trend in particle usage with the convergence of male and female speakers. This shows support for Hypothesis 3b, showing that there is an accommodation effect wherein both men and women reduce their use of Singlish particles in mixed-gender contexts, especially with the main effect of dyad type also being very close to significance [$F(1, 344) = 3.24$, $p = 0.073$] as reported previously.

In the other graph, Figure 12 (Partner Model plot) shows an interesting pattern. When the conversation partner is male, men use significantly more Singlish particles compared to women

[$t(344) = -3.04$, $p = 0.0135$*] (Table 14). However, when speaking to a female conversation partner, there is very clear a convergence by male speakers (decreasing) and female speakers (increasing) to the point of equivalence [$t(344) = -0.025$, $p = 1.00$]. This tells us two things: Firstly, as reported earlier, the strong main effect of speaker gender would mean that men tend to use Singlish particles more frequently than women, however, this is largely attributed to the high usage in same-gender male interactions and low usage by women in mixed-gender interactions. Secondly, this also serves as evidence that Singlish particles are particularly salient solidarity markers for same-gender male speech, and indicative of masculine language in Singapore English, corroborating the findings reported in Leimgruber et al.'s (2020) study.

4.3.3 Modelling by Particle Type

Investigating further into particle use, I extracted the top five most used particles — *ah*, *lah*, *eh*, *lor*, *leh,* which accounted for 91.78% of all particles used — and modelled their frequency per million words with speaker gender, partner gender and they particle type, with random effects on the pair level and the individual level:

```
lmer(sqrt(particles) ~ Gender * Partner Gender * Particle Type
+ (1|Pair) + (1|ID))
```

Results show that the type of particle use is a highly significant factor and interacts strongly with speaker gender and partner gender. Pairwise comparisons in Table 15 below show that for all particles (except for eh), same-gender male pairs (MM) consistently use more of these particles than same-gender female pairs (FF), further supporting that the use of Singlish particles is highly indexical of masculine ways of speaking and were utilised significantly more in in-group socialisation between men.

Table 15

Table of pairwise comparisons of ah, lah, eh, lor *and* leh *for Speaker Gender and Partner Gender*

contrast	Estimate	SE	df	z-ratio	p-value	
Particle: ah						
F Speaker F Partner - M Speaker F Partner	-810.71	606.19	Inf	-1.34	0.539	
F Speaker F Partner - F Speaker M Partner	481.01	562.16	Inf	0.86	0.828	
F Speaker F Partner - M Speaker M Partner	-4311.61	573.14	Inf	-7.52	0.000	***
M Speaker F Partner - F Speaker M Partner	1291.72	547.53	Inf	2.36	0.085	*
M Speaker F Partner - M Speaker M Partner	-3500.90	701.88	Inf	-4.99	0.000	***
F Speaker M Partner - M Speaker M Partner	-4792.62	664.22	Inf	-7.22	0.000	***
Particle: lah						
F Speaker F Partner - M Speaker F Partner	-382.53	558.90	Inf	-0.68	0.903	
F Speaker F Partner - F Speaker M Partner	140.57	539.94	Inf	0.26	0.994	
F Speaker F Partner - M Speaker M Partner	-2441.60	510.32	Inf	-4.78	0.000	***
M Speaker F Partner - F Speaker M Partner	523.09	513.60	Inf	1.02	0.739	
M Speaker F Partner - M Speaker M Partner	-2059.07	626.54	Inf	-3.29	0.006	
F Speaker M Partner - M Speaker M Partner	-2582.17	609.68	Inf	-4.24	0.000	
Particle: eh						
F Speaker F Partner - M Speaker F Partner	405.82	336.98	Inf	1.20	0.624	
F Speaker F Partner - F Speaker M Partner	346.28	340.48	Inf	1.02	0.739	
F Speaker F Partner - M Speaker M Partner	585.81	277.25	Inf	2.11	0.149	
M Speaker F Partner - F Speaker M Partner	-59.54	306.31	Inf	-0.19	0.997	
M Speaker F Partner - M Speaker M Partner	179.99	334.27	Inf	0.54	0.950	
F Speaker M Partner - M Speaker M Partner	239.53	337.80	Inf	0.71	0.894	
Particle: lor						
F Speaker F Partner - M Speaker F Partner	282.40	180.31	Inf	1.57	0.398	
F Speaker F Partner - F Speaker M Partner	281.55	180.40	Inf	1.56	0.401	
F Speaker F Partner - M Speaker M Partner	492.18	139.16	Inf	3.54	0.002	**
M Speaker F Partner - F Speaker M Partner	-0.86	154.80	Inf	-0.0055	1.000	
M Speaker F Partner - M Speaker M Partner	209.78	159.16	Inf	1.32	0.551	
F Speaker M Partner - M Speaker M Partner	210.64	159.26	Inf	1.32	0.548	
Particle: leh						
F Speaker F Partner - M Speaker F Partner	321.40	136.42	Inf	2.36	0.086	
F Speaker F Partner - F Speaker M Partner	245.25	147.12	Inf	1.67	0.341	
F Speaker F Partner - M Speaker M Partner	366.68	116.63	Inf	3.14	0.009	**
M Speaker F Partner - F Speaker M Partner	-76.15	115.21	Inf	-0.66	0.912	
M Speaker F Partner - M Speaker M Partner	45.28	114.71	Inf	0.39	0.979	
F Speaker M Partner - M Speaker M Partner	121.43	127.25	Inf	0.95	0.775	

The results in this subsection show the impact of the difference in particle types and how differences were reflected in their use by men and women in different gender contexts. In Section 2.5.4, I mentioned the different pragmatic functions of Singlish particles, particularly in relation to Gupta's (1992) model on the assertive continuum and how analysis on particle type can be helpful. However, in practice, it seems that only a few particles contribute to majority of the particle usage, leaving the remaining particles too infrequent to be statistically useful in this analysis. In this corpus, out of 18 particles, five particles account for more than 90% of all particle usage, meaning that the 13 other Singlish particles take up less than 10% of particle usage. This large disparity in particle frequency makes analysis on the particle-level tricky and potentially misleading if analyses were to treat all particles equally.

Nonetheless, this subsection explores the potential of particle type differences in gendered interaction that can inform future studies if they are interested to look more into this topic.

4.3.4 Interim Conclusions for Singlish Particles

In this section's analysis on Singlish particle usage, linear mixed-effects models show that there is a strong main effect of speaker gender with male speakers using more Singlish particles than female speakers, strongly supporting Hypothesis 3a.

There is also a main effect of dyad type that is close to significance, with mixed-gender dyads using fewer Singlish particles than same-gender dyads. This supports Hypothesis 3b and can be seen in the Accommodation Model plot with a converging trend and a decrease in Singlish particle usage for both male and female speakers from same-gender to mixed-gender dyads. Since Singlish particles are seen as informality/solidarity markers, the drop in Singlish particle usage is evidence that mixed-gender interactions indicate larger social distance between participants.

The Partner Model shows a clear difference between male and female speakers when conversing with a male conversation partner. In interactions with men, male speakers used significantly more Singlish particles than female speakers. However, when speaking with women, men reduced their use of particles while women increased their use of particles.

These results in this section suggest that Singlish particles is a strong marker for solidarity and low social distance between Singaporean speakers as both genders use more Singlish particles with their own in-group members. Especially for men, Singlish particles possibly play a big role in socialisation among men as MM pairs use more particles than FF pairs with great significance. In the case of Singlih particles, it would seem that similar to filled pauses, a high usage of Singlish particles is a feature of speech associated with typically masculine ways of speaking.

4.4 Language Style Matching Scores

4.4.1 Descriptive Statistics

Figure 13

Histogram showing a frequency plot of all 397 pairs and their LSM scores, coloured by gender pairing type.

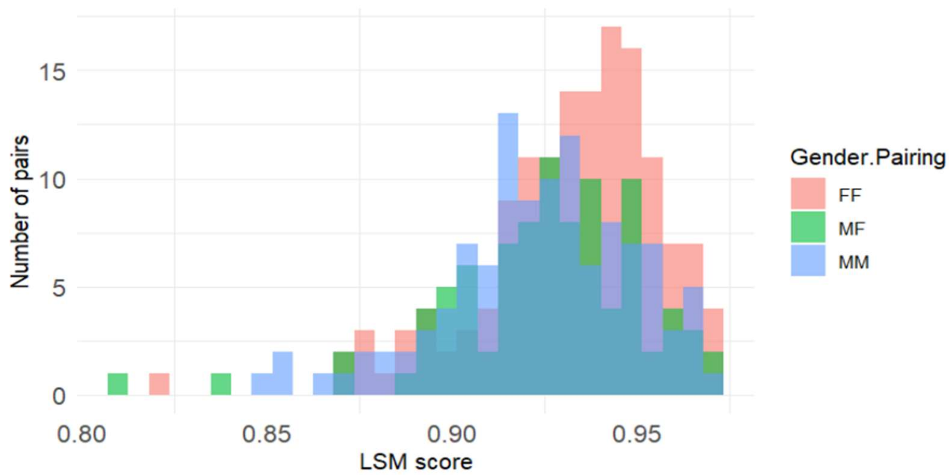


Figure 13 above shows the distribution of all 397 pairs in relation to their LSM scores, with female-female pairs represented in pink, male-male pairs in blue, and male-female pairs in green. The x-axis is a scale of LSM score and the y-axis represents the number of pairs that attain a certain LSM score. From the figure, it can be seen that all of the participants have LSM scores above 0.80, with most concentrating in the 0.90 to 0.95 range. Also clearly shown in the figure, FF pairs occupy the higher LSM scores compared to MF and MM pairs, which are generally more spread out in comparison. However, it is worth noting that there is an unequal number of pairs for each gender pairing, with 144 FF pairs, 112 MM pairs, and 91 MF pairs. Therefore, statistically modelling is required to fully ascertain the differences in LSM score between each gender pairing.

4.4.2 Model Fitting Results

For this analysis on Language Style Matching, a slightly different approach is taken compared to the previous dependent variables. A linear regression model is used instead as there are no random effects of pairs. This is because the calculation of LSM scores already took into account pair-level variations by generating a single LSM score from the two participants (see Section 3.6.5 *Measuring Language Style Matching*). As such, this also means that there are three levels for gender pairing as the independent variable — female-female pairs (FF), male-male pairs

(MM), and male-female pairs (MF). The linear regression used below seeks to model the LSM score as predicted by these three gender pairings:

```
lm(LSM ~ Gender.Pairing)
```

Table 16

Results of ANOVA of linear regression model on LSM scores

	df	*Sum Square*	*Mean Square*	*F-value*	*p-value*
Gender Pairing	2	0.0054	0.0027	4.18	0.016*
Residuals	344	0.222	0.00065		

F-test ANOVA results show that there is an effect of gender pairing on LSM scores [$F(2, 344) = 4.18$, $p = 0.016$*], meaning that the combination of gender pairs in dyadic interactions has a significant impact on how similarly interlocutors are to each other in terms of linguistic style. Table 17 below shows the EMMs of LSM scores for each gender pairing, which are plotted in Figure 14 on the next page.

Table 17

A table showing the estimated marginal means of Language Style Matching (LSM) scores for each of the three pair types — female-female (FF), male-male (MM), and male-female (MF).

Gender Pairing		***Estimated Marginal Mean***	***Standard Error***	***df***	***Lower Confidence Level***	***Upper Confidence Level***
Female-Female	(FF)	0.9325	0.002118	344	0.9283	0.9366
Male-Male	(MM)	0.9238	0.002402	344	0.9191	0.9285
Male-Female	(MF)	0.9256	0.002665	344	0.9203	0.9308

Figure 14

Graph plotting the Estimated Marginal Means (EMMs) of each of the three pair types.

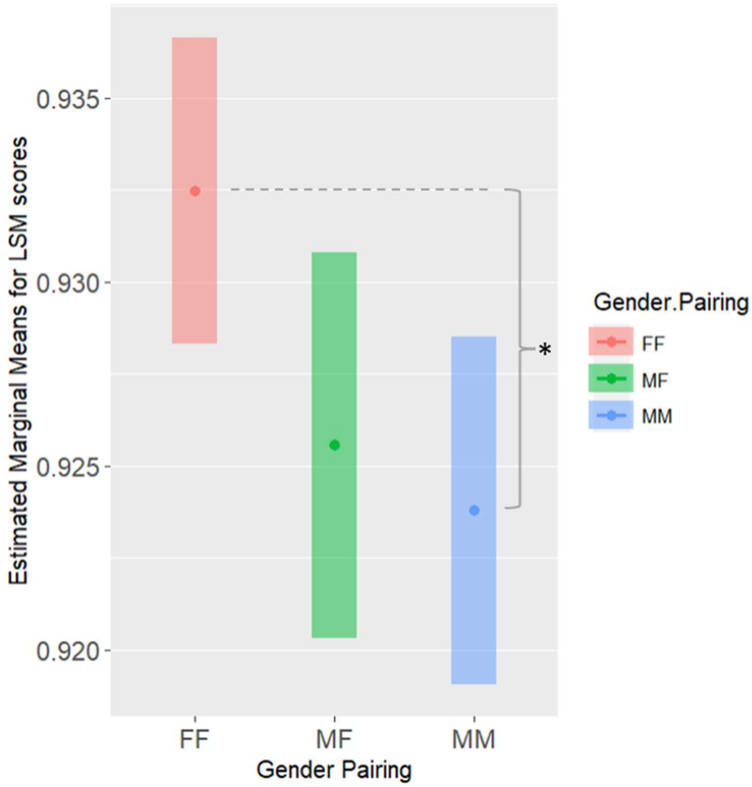


Table 18 below shows the pairwise differences between each of the groups. There is only one comparison with significant differences, which is between same-gender female (FF) and same-gender male (MM) pairs [$t(344) = 2.71$, $p = 0.019$*] and it is indicated in Figure 14 above.

Table 18

A table showing the three pairwise comparisons of LSM scores between each of the three pair types.

Pairwise Differences	*Estimate*	*SE*	*df*	*t.ratio*	*p-value*	
FF – MF	0.00691	0.00340	344	2.030	0.1066	
FF – MM	0.00868	0.00320	344	2.712	0.0192	*
MF – MM	0.00177	0.00359	344	0.494	0.8740	

Figure 14 above plots the estimated marginal means of each gender pairing. It is worth noting that just within the small range of 0.92 to 0.94 in LSM score, there exists observable differences between the pairings. The estimated mean LSM score for FF pairs is the highest, followed by mixed-gender MF pairs, and lastly MM pairs, which is significantly lower than the FF pairs.

This means that same-gender female pairs match linguistically with each other significantly more than same-gender male pairs, with mixed-gender pairs somewhere in between.

This result goes against Hypothesis 4 which predicted that same-gender pairs (FF and MM) would both have a higher LSM score than mixed-gender pairs (MF) due to the increased social distance in mixed-gender interactions. Particularly for male participants, there is a likelihood that they might have style-matched *more* with their female friends than with fellow men.

However, this result can still be meaningfully interpreted with the use of other theoretical models explored in the literature of language and gendered interactions. Researchers such as Coates (1996, 1997) have concluded that talk among women friends highly cooperative, hence the higher LSM score; and works by Cameron (1997) and Kiesling (2007) show that men's talk contains more competitive elements and distancing forms as a display of solidarity and masculinity, which could suggest why there is significantly lower LSM score for MM pairs compared to FF pairs. However, as these researchers also warn, it is important not to solely attribute cooperative as a feminine style of speech and competitive as a masculine style of speech as these traits are not neatly dichotomous in gender categories (Coates, 2012), as men's talk also feature many highly cooperative devices in women's speech (Cameron, 1997).

Nevertheless, the insight gain from analysing LSM score gives us a good picture and valuable information regarding the interactive styles between friends of different gender pairings in Singapore.

4.4.3 Interim Conclusions for LSM Scores

A linear regression model across the three gender pairing types is conducted and results show that there is a significant difference in LSM scores between same-gender female pairs and same-gender male pairs — FF pairs match each other in linguistic style significantly more than MM pairs, with MF pairs somewhere between the two groups. Hence, Hypothesis 4 is contradicted as it expected that both MM and FF would style-match significantly *more* than MF pairs. These results point to the possibility that same-gender female dyads were more cooperative, reflected in their higher LSM scores; same-gender male dyads were more competitive that resulted in the lower LSM scores; and lastly, mixed-gender dyads lie between the two, possibly indicating a process of accommodation by both the FF and MM pairs.

4.5 Considering other social factors

While this thesis is interested in gender effects on linguistic features that indicate formality and social distance, there is an additional motivation to look at other social factors mediating these features. In Section 2.2.5 *Intersectionality*, the importance of these different social factors, such as race, age, class play in language use. In this section, I will present some findings on these dimensions using the same approach that has been done above for gender. There will be a particular attention paid on participant ethnicity as ethnic variation in language use is a common research area in Singapore English (e.g., Lim, 2000; Leimgruber et al., 2020; Soh, Lee & Tan, 2022). There will also be an exploratory look at age groups as some previous research on language and gender (e.g., Laserna et al., 2014; Stupka, 2011) suggested age to be a more important factor instead. However, it is important to stress that these analyses are supplementary to the discussion on language and gender in Singapore and lie outside the main scope of this thesis.

4.5.1 Considering Ethnicity

Singapore is a multicultural society with three major ethnicities — Chinese (74.3%), Malay (13.5%), and Indian (8.96%). The paragraphs below will describe the results of backchanneling, filled pauses, Singlish particles, and LSM scores in relation to ethnicity. Two sets of results will be shown for each variable. The first is a comparison between pairs consisting of the three ethnicities: Chinese-Chinese (342 participants) vs. Malay-Malay (82 participants) vs. Indian-Indian (90 participants). The second set of results is a comparison between participant pairings in same-ethnicity condition (514 participants) versus mixed-ethnicity condition (180 participants).

Table 19

Table showing the participant numbers for the three major ethnic groups and participants in the same-ethnicity condition and mixed-ethnicity condition

Comparison 1: Between Ethnicities	***Chinese***	***Indian***	***Malay***
	342	90	82

Comparison 2: Same vs. Mixed Ethnicities	***Same Ethnicity***	***Mixed Ethnicity***
	514	180

The following pages will give a brief description of these results, including additional models with speaker gender to account for possible intersectional effects. Additional details, such as Q-Q plots, graphs, and especially tables of figures for each variable being examined can be found in Appendix A.

In the analyses below, some conclusions will be drawn relating to formality and social distance for ethnicity and age, similar to what was done above with gender. However, to emphasise again, since ethnicity and age effects are not the focus of this thesis, results presented would serve only as an exploratory analysis to address concerns on intersectionality with these other prominent social factors.

Backchannels

Figure 15

Graph of EMMs of backchannels used per million words, grouped by ethnicity.

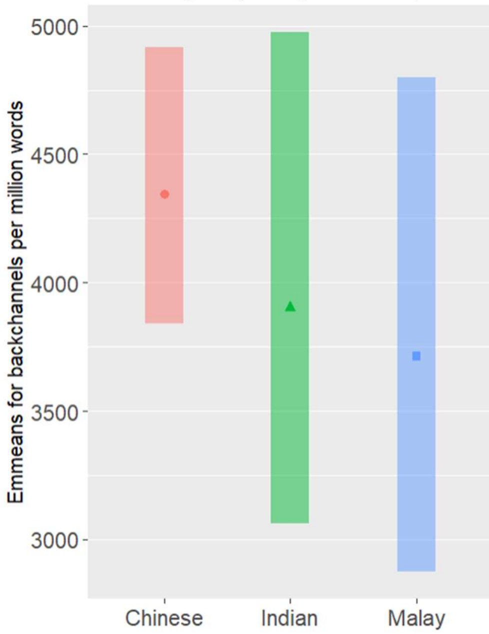
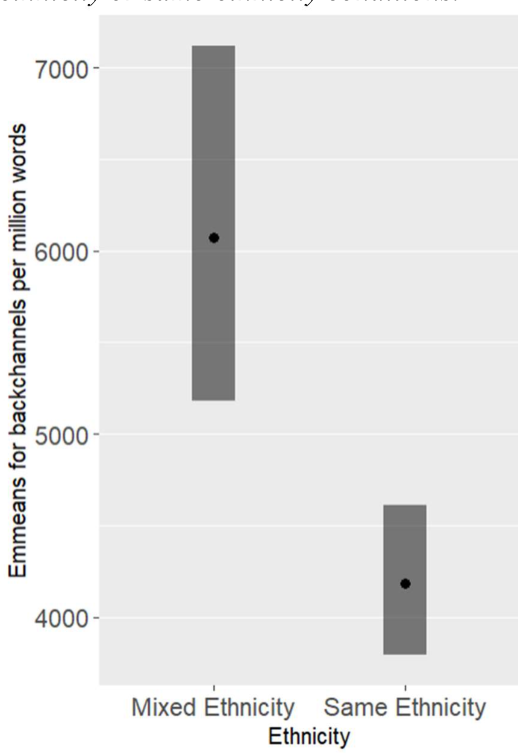


Figure 16

Graph of EMMs of backchannels used per million words between pairs in mixed-ethnicity or same-ethnicity conditions.



ANOVA F-test results show no significant effects of ethnicity on backchannel usage [$F(2, 260) = 0.756$, $p = 0.47$]. Figure 15 on the left plots the EMMs of backchannels per million words for the three ethnicities wherein there are no significant differences between them. Additionally, a linear mixed effects model with both ethnicity and gender as fixed effects also did not yield significant interaction effects [$F(2, 503) = 0.100$, $p = 0.90$], suggesting that dyads of different ethnicities did not use backchannels differently from one another.

However, Figure 16 shows that there is significant a difference between dyads of mixed-ethnicity and same-ethnicity. The EMM plots for these two groups on the left shows that dyads of mixed ethnicity used significantly more backchannels than dyads in same ethnicity [$F(1, 583) = 16.5$, $p < 0.001$], which falls in line with predictions of increased supportive communication with increased social distance. However, similar to ethnicity types, there were no significant interaction effects of ethnicity pairing condition with gender [$F(2, 260) = 0.212$, $p = 0.65$].

Filled Pauses

Figure 17

Graph of EMMs of filled pauses used per million words, grouped by ethnicity.

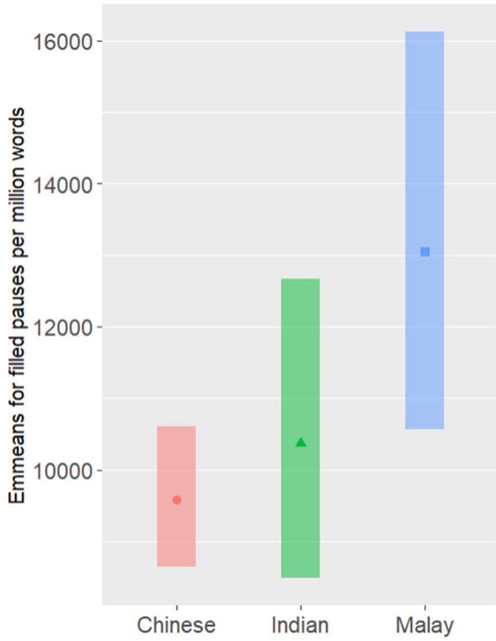
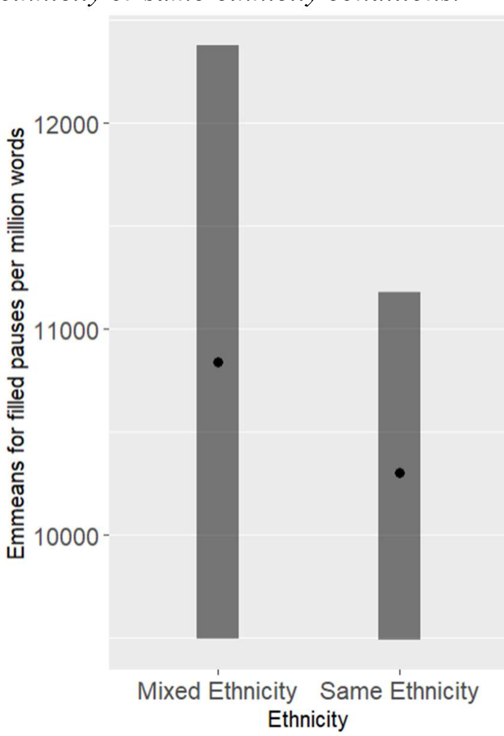


Figure 18

Graph of EMMs of filled pauses used per million words between pairs in mixed-ethnicity or same-ethnicity conditions.



ANOVA F-test shows that there is a significant effect of ethnicity on filled pauses usage [$F(2, 268) = 3.38$, $p = 0.036$*]. Figure 17 on the left plots the EMMs of the three groups of ethnicities and shows that the Malay pairs used significantly more filled pauses than the Chinese pairs [$t(259) = -2.33$, $p = 0.05$]. However, when modelled with participant gender, no interaction effects were found between ethnicity and gender [$F(2, 507) = 0.879$, $p = 0.42$].

Comparing mixed-ethnicity and same-ethnicity in Figure 18, there is no significant difference between these two groups [$F(1, 596) = 0.452$, $p = 0.51$]. However, modelling ethnicity pair type and gender yields significant interaction effects $F(1, 688) = 6.29$, $p = 0.012$*], with the only significant difference between male speakers and female speakers in same-ethnicity condition [$t(563) = -2.98$, $p = 0.016$]. These results reinforce the effect of speaker gender on filled pause usage, as this gender differences become more pronounced in interactions between participants of the same ethnicity.

Particles

Figure 19

Graph of EMMs of Singlish particles used per million words, grouped by ethnicity.

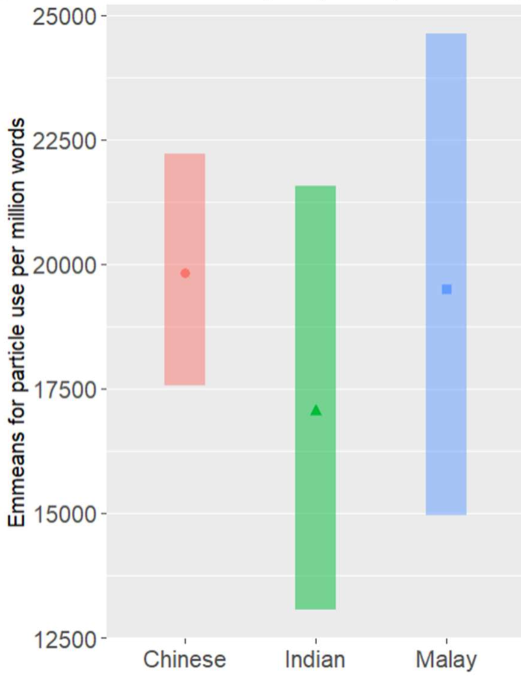
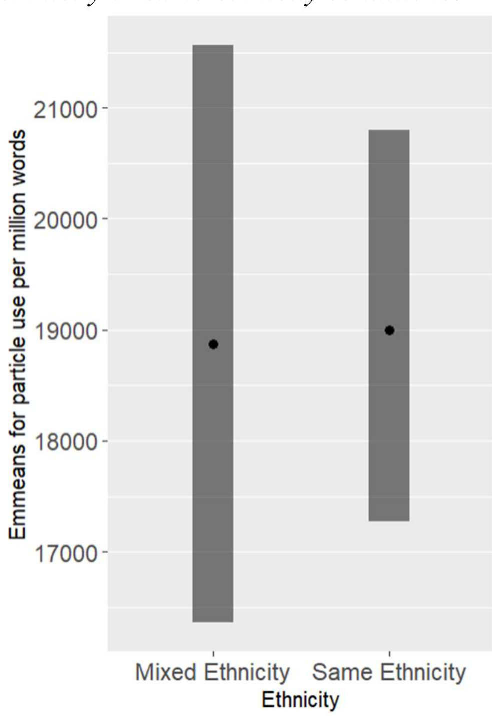


Figure 20

Graph of EMMs of Singlish particles used per million words between pairs in mixed-ethnicity or same-ethnicity conditions.



ANOVA F-tests show no significant effects of ethnicity on particle usage, [$F(2, 269) = 0.613$, $p = 0.54$], which is contrary to Leimgruber et al.'s (2020) conclusion which showed that Indians use significantly fewer particles than Chinese and Malay speakers. One crucial difference between their study and this thesis is in the methodology. While Leimgruber et al. looked at participant ethnicity as a factor, I compared same-ethnicity pairs across all three ethnicities (Chinese-Chinese vs. Malay-Malay vs. Indian-Indian) in which ethnic variations can be analysed at an interactional level instead of just the individual level. Figure 19 on the left plots the EMMs for particle usage per million words, showing no significant differences across the ethnic pairs. Interaction effects with gender are also not observed [$F(2, 418) = 1.83$, $p = 0.16$].

In Figure 20, no significant differences were found between mixed-ethnicity and same-ethnicity pairs [$F(1, 691) = 0.691$, $p = 0.93$]. Neither were there interaction effects with gender [$F(1, 601) = 0.199$, $p = 0.66$].

These results highly suggest that Singlish particles were neither affected by different ethnicity groups, nor whether speakers were interacting with someone of the same ethnicity or not.

LSM Scores

Figure 21

Graph of EMMs of LSM scores, grouped by ethnicity.

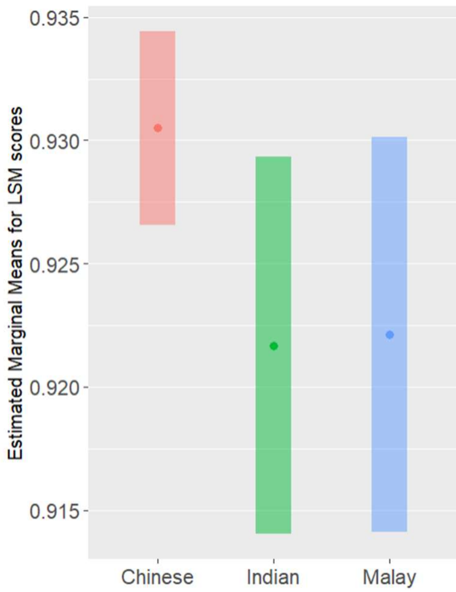
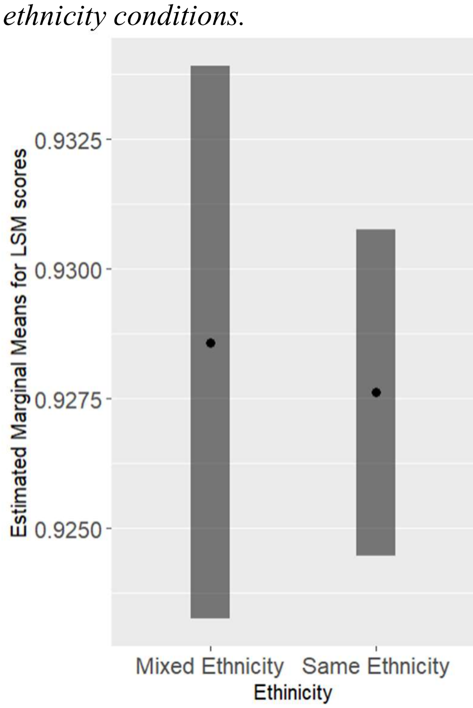


Figure 22

Graph of EMMs of LSM scores between pairs in mixed-ethnicity or same-ethnicity conditions.



ANOVA F-test on the model showed that there is a significant effect of the ethnicity of the pairs on LSM scores [$F(2, 254) = 3.13$, $p = 0.045$*], however post-hoc pairwise comparisons did not show any significant differences between groups. Figure 21 plots the EMMs for LSM scores for the three ethnicity groups. It shows that Chinese-Chinese pairs seem to have a higher mean LSM score than Malay-Malay and Indian-Indian pairs, but the confidence intervals suggest that this might just be at the threshold of significance. Additionally, a multiple regression model with ethnicity groups with gender pairing did not show a significant interaction between these two factors [$F(2, 248) = 1.56$, $p = 0.19$].

Examining participants in mixed-ethnicity and same-ethnicity condition, Figure 22 clearly shows that there is no differences between these groups [$F(1, 345) = 0.0948$, $p = 0.76$], and neither were there interaction effects when modelled with gender pairings [$F(2, 341) = 0.33$, $p = 0.72$].

4.5.2 Considering Age

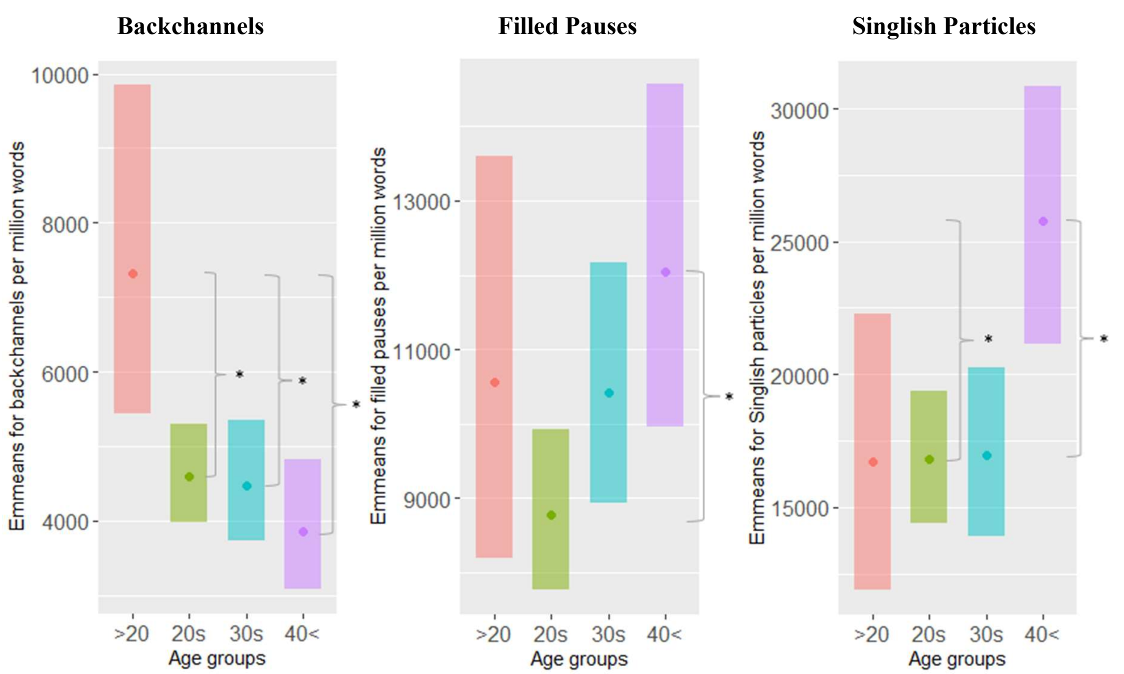
This section will focus on age group differences for the variables being examined. Most of the friend pairs in the data used are in the same age range, hence, in the analysis for age in this subsection, participants will be grouped in age groups of under 20 (N=54), 20 to 29 years-old (N=234), 30 to 39 years-old (N=146), and above 40 years-old (N=96). There was originally more granularity with 40, 50 and 60 year-olds, but due to their sparser participation, they are combined into one group.

The same linear mixed effects models were conducted with age groups as the predictor, like the one for ethnicity groups above. ANOVA F-tests on the models reveal main effects of age groups on all the examined variables of backchannelling [$F(3, 261) = 4.011$, $p = 0.0082$**], filled pauses [$F(3, 261) = 2.84$, $p = 0.038$*], and Singlish particles [$F(3, 261) = 4.42$, $p = 0.0047$**].

In Figure 23 below, the graphs for backchannelling, filled pauses, and Singlish particles are plotted to show their distribution across age groups.

Figure 23

Graph of EMMs of backchannelling, filled pauses, and Singlish particles, grouped by age groups.



Pairwise comparisons for each variable are done and the tables of results can be found in Appendix A. For the use of backchannels, participants under 20 years-old contrasts with all other age groups; for filled pauses, only participants aged 40 and above contrast with participants in their 20s; and for Singlish particles, participants aged 40 and above use significantly more particles than participants in their 20s and 30s.

While the significant effects of age group on the dependent variables are indicative of a possible relationship between age and these linguistic features that mark formality/social distance. There is a caveat to these conclusions, which is the unbalanced sample sizes across age groups, especially for the two extreme groups — participants below 20 years-old are relatively small (N=54), and the group above 40 years-old is a very broad age range (comprising 40, 50, and 60 year-olds, an age range that spans 30 years) which could blanket some meaningful variations within the group of older participants.

To investigate intersectional effects, linear mixed models are conducted with both gender and age as predictors for the three variables, wherein no interaction effects are observed between speaker gender and the age groups of the participants for backchannels [$F(3, 506) = 1.82$, $p = 0.142$], filled pauses [$F(3, 500) = 1.62$, $p = 0.183$], and Singlish particles [$F(3, 478) = 1.81$, $p = 0.144$]. This means that, similar to most of the analyses with gender and ethnicity, intersectional effects between gender and age are not statistically present.

Lastly, for age group effects on LSM scores, linear modelling showed no significant main effect [$F(3, 261) = 1.50$, $p = 0.22$], as well as no interaction effects with gender [$F(6, 253) = 1.40$, $p = 0.21$].

4.5.3 Interim Conclusions for Analysis with Other Social Factors

In this section, ethnicity and age were considered as potential predictors for the variables looked at in this thesis. For ethnicity comparisons, two methods were carried out — comparison between pairs of the three ethnic groups were compared (ie. Chinese-Chinese vs. Malay-Malay vs. Indian-Indian pairs), and also comparisons were made between participants who were in either the same-ethnicity or mixed-ethnicity context. For age comparisons, dependent variables are measured across four age groups (i.e., <20 years-old, 20s, 30s, 40 years-old<).

By modelling the data with LMMs, analyses on pairs of different ethnicities only showed filled pauses to have significant main effects. No interaction effects were found between ethnicity of pairs and speaker gender on any of the dependent variables analysed. Analyses on same-ethnicity and mixed-ethnicity pairs revealed a main effect of ethnicity for backchannelling only. As for interaction effects with gender of participant, only filled pauses usage showed this interaction effect.

For age groups of participants, there is a main effect of age groupings for all variables except LSM scores. This means that age as a factor mediating the features that indicate social distance between interlocutors is a topic worth exploring, at least more so than ethnicity effects. However, no interaction effects were observed for any of the variables, indicating that intersectional effects of gender and age might not be of great concern.

4.6 Summary of Results

This chapter covered the analyses on how speaker gender, partner gender and dyad type influenced the three dependent variables (backchannels, filled pauses, Singlish particles) using linear mixed effects models to obtain estimated marginal means for gender comparisons and pairwise contrasts. On top of that, an analysis of the LSM scores modelling against gender pairing variables of male-male, female-female, and male-female was also conducted for a supplementary view on gendered interactions with regards to matching of linguistic styles between participants.

4.6.1 Summary of Main Effects

Table 20 shows a summary table of the results, with each row representing the dependent variables (backchannels, filled pauses, Singlish particles, LSM scores) and the columns representing the independent variables (speaker gender, partner gender, and dyad type; or gender pairing type for LSM score). The reason why there is no "interaction effects" presented in Table 20 is because the interaction effect of any two variables is equivalent to a main effect of the third. For more detail, this was addressed in Section 3.7.1 (*Interdependency of Independent Variables*) previously. Within the boxes in the table, *Yes* indicates the presence of main effect of that independent variable, while *No* indicates the absence of main effect.

Table 20

Table of the summary of results for the four dependent variables against speaker gender, partner gender and dyad type.

	Speaker Gender	**Partner Gender**	**Dyad Type**
(1) **Backchannels**	Yes, F higher. ($p = 0.010$)	No.	No.
(2) **Filled pauses**	Yes, M higher. ($p = 0.054$)	Yes, M higher. ($p = 0.060$)	No.
(3) **Singlish Particles**	Yes, M higher. ($p = 0.0037$)	No.	Yes, mixed-gender lower. ($p = 0.072$)
		Gender Pairing Type	
(4) **LSM score**		Yes, FF > MM. ($p = 0.016$)	

4.6.2 Summary of Hypotheses and Outcomes

Table 21 below presents an overview of the results presented in the previous chapter with respect to the hypotheses formed in this thesis. To iterate again, part (a) of the hypotheses focuses on the differences in linguistic behaviours between men and women in their in-group contexts of same-gender, while part (b) goes a step further to postulate the changes in linguistic behaviours in mixed-gender contexts (see Section 2.5.1 *Research Questions and Hypotheses*).

Table 21

A table showing the summary of results as it relates to the hypotheses outlined in this thesis.

	Hypothesis (a)	**Results (a)**	**Hypothesis (b)**	**Results (b)**
(1) Backchannels	Women > Men Backchannels are features of feminine styles of speech (cooperative)	Fully supported.	Accommodation. Men and women to *increase* in usage for cooperative communication.	Contradicted. Both genders *decreased* backchannel usage instead.
(2) Filled Pauses	Men > Women Filled Pauses are features of informal speech and more typical of men's style	Partially supported.	Accommodation. Both men and women to *decrease* usage because both will adopt more formal style.	Partially supported. Accommodation but women *increased* filled pauses instead.
(3) Singlish Particles	Men > Women Particles are features of informal speech and more typical of men's style	Fully supported.	Accommodation. Both men and women to *decrease* usage because both will adopt more formal style.	Supported.

	Hypothesis	Results
(4) **Language Style Matching Scores**	FF, MM > MF Same-gender dyads would have higher LSM score than mixed-gender dyads.	Not supported. FF > MF > MM with significant difference between FF and MM, showing that FF style-match more than MM.

Considering the first three variables as presented in Table 20 in the previous subsection, the use of backchannels, filled pauses, and Singlish particles all showed significant effects with regards to gender, which means that Hypothesis (a) for all these variables are supported — i.e., women use more backchannels than men, and men use more filled pauses and Singlish particles than women. This indicates that the gendered patterns of speech found in the literature are also present in Singapore English, and will be further discussed in Section 5.1 (*Gendered Language in Singapore English*) of the General Discussion chapter.

For Hypothesis (b), however, the outcomes vary for these three variables. In backchannelling, Hypothesis 1b expected backchannels usage to increase in mixed-gender contexts compared to same-gender contexts, but this hypothesis is not supported as backchannels decreased instead. Hypothesis 2b for filled pauses was partially supported as the accommodation in mixed-gender dyads is affirmed, but instead of women decreasing their use of filled pauses, an increase was observed. On the other hand, Hypothesis 3b, which predicted Singlish particles usage to reduce for both men and women, is fully supported with both men and women converging their usage of Singlish particles with mutual decrease their use of particles in mixed-gender contexts.

Lastly, for Language Style Matching scores, Hypothesis 4 which expected same-gender dyads to have greater LSM scores than mixed-gender dyads is not supported. Same-gender dyads did not show higher LSM scores than mixed-gender dyads. Instead, the significant differences found were between the same-gender men and same-gender women pairs.

Chapter 5 General Discussion

5.1 Gendered Language in Singapore English

Section 2.2.4 (*The Dynamic Approach*) in the literature review raised the pitfalls of language and gender research being reductive and essentialist, categorising linguistic features as "men's language" and "women's language" (Butler, 1999; Hall, 2003; Talbot, 2019). However, one cannot deny that gender is still a major social category in which men and women use to perform and construct their identities (Eckert & McConnell-Ginet, 2003; Kiesling, 2007; Coates, 2013). During the course of research over the previous decades, certain linguistic features have been found to be more typical of feminine styles of speech, or more typical of masculine styles of speech, and sometimes neither. Since gender is a main effect for all the variables investigated in this thesis, this means that there is gender differentiation in the use of backchannels, filled pauses, and Singlish particles in dyadic interactions between friends in Singapore.

For backchannels, women were found to have a higher usage as compared to men, which is in accordance with the literature citing that supportive communication is a tendency of feminine speech that is present in both the western anglophone countries (e.g., Bilous & Krauss, 1988; Mulac & Bradac, 1995; Plug et al., 2021), as well as Asian countries like Japan (e.g., Kogure, 2003; Ptaszynski, Hasegawa & Masui, 2014).

Filled pauses and Singlish particles were used significantly more by male speakers than female speakers which is also consistent with the findings in the literature. Liberman (2014), Tottie (2011), and Shriberg (2001) also observed that men used more filled pauses in the Fisher corpus, British National Corpus, and Switchboard corpus; and Yuan et al.'s (2016) study of Chinese filled pauses 嗯 *en*/呃 *e* also showed the same patterns that men use more filled pauses than women. For Singlish particles, the study by Leimgruber et al. (2020) similarly showed that text messages sent by men include more usage of Singlish particles than women, which tells us that men's higher frequency of Singlish particle usage is reflected in both spoken interactions and computer-mediated text communication.

However, the case for Singlish particles unfolds more interestingly upon further inquiry. As far as comparing with other cultures and languages that share sentence-final particles (SFPs), this is where the patterns diverge, making SFPs in the case of Singapore English rather unique in comparison. Singlish particles are derived mainly from an amalgamation of various Chinese languages like Cantonese and other Southern Chinese varieties (Lim, 2007), so it is useful to compare Singlish particles with these languages and their frequency with regards to gender. In Cantonese, it has been consistently shown that women use *more* sentence-final particles than men (Chan, 1999; Zhang & Zhuang, 2020), and likewise for Mandarin Chinese in both Mainland China (Shin, 1998; Chan, 1998) and Taiwan (Lin, 2005; Baumel, 2020). This is frequently attributed to politeness strategies that is often employed by women through the use of sentence-final particles (Chan, 1998, 1999; Wamsley 2019, Baumel, 2020). As can be seen, the trends from these Chinese varieties are opposite from what we observe from the use of SFPs in Singapore in which men used more Singlish particles instead of women. I propose two possible explanations for the conflicting findings between SFPs in Singapore and the other regions — the role of SFPs in relation to gender and politeness, and the diglossia in Singapore English.

Firstly, as discussed, Cantonese and Mandarin speakers in the other SFP-containing speech communities (Hong Kong, China, and Taiwan) perceive the use of SFPs as polite while Singlish and its particle usage has been perceived as impolite and rude (Yoong, 2010). Though Stadler (2018) argued for the use of Singlish as a politeness strategy, she noted the lack of overt politeness markers in Singlish. However, what Singlish particles lack in overt impoliteness particles is made up for by its function in solidarity (Lai & Tan, 2023), which is relevant to the idea of Singapore's diglossia of with the H-variety of Singapore Standard English (SSE) and the L-variety of Singapore Colloquial English (SCE, or Singlish) (Gupta, 1989).

The L-variety of Singlish is associated with a sense of solidarity (Brown, 1999; Koh, 2009) that binds Singaporeans together (Cavallaro & Ng, 2009), and especially so for men, as explored by Lim (2009) in their use of extreme/crude Singlish as performance of masculinity. This is a common observation as the L-variety often carries with it certain covert prestige that men tend to use more of to construct their masculinity (Trudgill, 1974; 2000), which will be revisited in the next section in more detail. This distinction between H- and L-variety in Singapore English is an additional layer of complexity on top of the different perceptions of politeness of SFPs, which could account for these differences in communities observed above. In other words, the underlying narrative is the same — women tend to use more standard and polite forms in speech than men, only that SFPs are considered overtly *polite* and *feminine* in other SFP-containing languages/communities whereas SFPs are considered overtly *impolite* and *masculine* in Singapore English, and hence, women were observed to use more SFPs in other communities while men were observed to use more SFPs in Singapore.

Naturally, the caveat for these conclusions on speaker gender as a predictor of differences in language use is that they only work under traditional gender assumptions that men would use prototypically masculine features and women would use more prototypically feminine features, barring edge cases such as "sissies", "tomboys" (Hall, 2003), and the full spectrum of the queer community — which is understandably a limitation of both this thesis as well as the NSC data. Nevertheless, the conclusions here are useful in illuminating trends in masculine and feminine tendencies in speech and characterisation of speech for Singaporeans, specifically, the use of backchannelling is indexical of feminine styles of speech while the use of filled pauses and Singlish particles are indexical of masculine styles of speech.

5.2 Gendered Interactions and Social Distance

Apart from the features that arise from speaker gender as discussed above, this thesis also used backchannels, filled pauses, and Singlish particles as proxies to indicate the level of social distance between interlocutors, and in turn, how these features change in same-gender or mixed-gender conditions. Particularly, this section is interested in synthesising the results for filled pauses and Singlish particles, which were shown to have predicted the meaningful relationships between gendered interactions and social distance. On the other hand, the case of backchannels will instead be revisited in the next chapter in Section 5.3 (*Word Class of Variables*), addressing its limitations and deviation from theory and predictions in more detail.

5.2.1 Socialisation Strategies in Closer Social Distances

As indicated in Section 2.5.1 (*Research Questions and Hypotheses*), the thrust of this thesis is informed by theories of CAT, together with the Difference and Dynamic approaches in language and gender. The Hypotheses (b) made in this thesis is based on the premise that there will be a larger social distance between participants of a different gender, and closer social distance with participants of the same gender. Therefore, participants with closer social distances (same-gender dyads) are more likely to exhibit linguistic features that are *less formal* and have *more markers of solidarity* than participants with greater social distances (mixed-gender dyads) — that is, in the form of fewer backchannels, more filled pauses, and more Singlish particles.

As discussed in the previous section on gender differences in Singapore English, the variables of filled pauses and Singlish particles were shown to be indexical of masculine features of speech and considered features that indicate informality and solidarity within the literature. The results of both these two variables in same-gender contexts further showed that male pairs (MM) have significantly higher usage of these features compared to female pairs (FF), suggesting that the use of filled pauses and Singlish particles are particularly salient in men and highly relevant in the establishment of male in-group solidarities. In the view of the Dynamic approach, these two features present their performativity of masculinity in their own ways. Firstly, the more unconstrained displays of disfluencies (filled pauses) show the informal context of male-male interactions and the lack of attempts to be formal with each other (Schachter et al., 1991; Broen & Siegel, 1972) which is a particularly important aspect in male interactions (Kiesling, 2007). Secondly, the use of non-standard/L-variety speech features, like the use of Singlish particles, carries a covert prestige (Trudgill, 1974; 2000) that Singaporean men default to in their performance of their masculinities with other men, as a way of expressing toughness and machismo (Lim, 2009) during speech.

Furthermore, in line with the Difference approach, these behaviours observed in same-gender interactions indicate the differences in socialisation strategies between men and women within their respective in-groups, and that men and women in Singapore inhabit different cultural spheres that have different socialisation norms regarding language use in interaction. Men were socialised to use much more of the L-variety among themselves, whereas women are socialised to not use as much.

5.2.2 Socialisation Strategies in Larger Social Distances

On the other hand, in mixed-gender interactions where social distances between interlocutors are larger, both these variables saw female and male speakers converge to a point of minimal difference for backchannels and particles used (see Figure 7 for filled pauses and Figure 11 for Singlish particles). This change in linguistic behaviour indicates a clear trend that the speakers are negotiating a cross-gender (and hence cross-cultural) boundary with increased social distance — a boundary which resulted in men decreasing the use of filled pauses and Singlish particles, and women increased use of filled pauses or reduced the use of Singlish particles. In other words, the gender differentiation in filled pauses and Singlish particles usage between men and women in same-gender contexts were attenuated in mixed-gender contexts, as shown by the validation of Hypothesis (b) for these two variables.

We can see the effect social distance has on the use of Singlish particles where both men and women have shown convergence with decreased usage, especially for the case of Singlish particles, which is the solidarity marker for Singaporeans (and thus, both genders). This is supported by the main effect of dyad type for Singlish particles, in which mixed-gender dyads use significantly fewer particles than same-gender dyads. This shows that in the pairs of the same gender, the close social distance with same-gender partners reflected high solidarity and thus higher Singlish particles usage. Whereas for mixed-gender dyads, the increased social distance due to cross-gender communication reflected lowered use of Singlish particles, a clear case of psychological convergence between men and women in interaction (see Section 5.2.3 below on *Linguistic Convergence and Psychological Convergence*). Using a more formal register and speech style is a common strategy for interactions with greater social distances (Giles & Ogay, 2007), and especially so for men who have more social motivations to switch away from their in-group socialisation practices to more standard and polite forms (Goffman, 1977; Kiesling, 2007), as women are usually the users of the more standard and polite forms of speech (Holmes, 1995; Mills, 2003).

5.2.3 Linguistic Convergence and Psychological Convergence

This subsection seeks to explain the slight differences in trends observed in filled pauses and Singlish particles. Although hypothesised similarly to Singlish particles, filled pauses are proposed to be more controlled and less frequent in formal contexts with larger social distances, hence it is expected that both male and female speakers would decrease in use in mixed-gender contexts (Hypothesis 2b). However, instead of a decrease the use of filled pauses, results show that women *increased* their use of filled pauses when talking with men.

In Section 2.3.5 (*Speech Complementarity*), I covered the differences between *linguistic convergence* and *psychological convergence*. Linguistic convergence is simply an accommodation of linguistic features in relation to the two interlocutors' different levels of usage, where they would increase or decrease their use of said feature respectively to match with each other. Psychological convergence, on the other hand, is also a form of communication accommodation, but is based on the convergence of the two interlocutors towards a similar psychological state/objective and need not be reflected in convergence of linguistic features (Thakerar et al., 1982; Giles & Ogay, 2007; Dragojevic et al., 2016). Both Hypothesis 2b (filled pauses) and 3b (Singlish particles) are made with psychological convergence in mind — that men and women would err on the side of being formal with an opposite gender conversation partner and use less filled pauses and Singlish particles.

However, while psychological convergence was not shown for the use of filled pauses, linguistic convergence was observed instead. In the results in Figure 7, it can clearly be seen that men decrease and women increase their use of filled pauses in mixed-gender interactions to meet in the middle (Section 4.2.2 *Model Fitting Results: Filled Pauses*). This spotlights a very important aspect in CAT that this study (and future research) need to be careful with — toeing the line between linguistic convergence and psychological convergence, as they may not point trends in the same direction as each other.

Another reason for the slight differences in filled pauses and Singlish particles could also be the fact that, while both features are similar in that they signal the lack of formality during speech, filled pauses fundamentally differs from Singlish particles in word class and linguistic function. Singlish particles has the additional function such as being a solidarity marker that filled pauses do not, and also being a specific feature of a distinct language variety. Whatever the case, the awareness and distinction between linguistic convergence and psychological convergence is something paramount for any work involving CAT.

5.3 Word Class of Variables

This difference in word class is more relevant to the variable of backchannels. We saw that in the summary of results (Table 20 and Table 21) that the outcome on backchannels for Hypothesis (b) is different from filled pauses and Singlish particles. One possible explanation for the discrepancy between these two groups of results could just be due to the inherent nature of word classes of these variables.

Not all linguistic features are built the same and are of equal usefulness. "Backchannels" as a linguistic feature is considered a much more open class of words (Drummond & Hopper, 1993) compared to filled pauses and Singlish particles, which are comparatively more closed. The ambiguity of what is termed "backchannels" in the literature surrounding what constitutes a backchannel or not (see Section 3.6.2 Measuring Backchannelling). For example, the term *okay* that was originally referred to as a backchannel by Yngve (1970), but not other researchers that came after (c.f., Drummond & Hopper, 1993, Fellegy, 1995). In an analysis done where it was included, the term *okay* would be the second-most used backchannel in the NSC and the results of the linear mixed models were drastically changed (see Appendix A: *Backchannelling (with okay): Speaker Gender, Partner Gender, Dyad Type*). When included, the main effect of gender on backchannels is no longer present, and there are no statistical differences between any of the speakers in any of the dyad conditions. This means that depending on what researchers want to include as a backchannel in analyses, it would have a sizeable impact on the results, since backchannelling can contain a wide-ranging breadth of words and phrases in high frequency. Therefore, a similar argument could be made for the terms *yes* and *right*, which are considered to be backchannels, as they function similarly to *okay* like a discourse marker or a simple response of agreement instead of a backchannel.

Conversely, other phrases and non-speech sounds that were not considered in this thesis may also fall in the category of backchannels, such as *I see*, or *oh* and *ah*. Essentially, there could be an innumerable number of words, phrases, sounds, that are responses that do not attempt to take the conversational floor. Additionally, backchannelling and supportive communication need not be only done linguistically through lexical items, but also paralinguistically through head nods, gestures, and change in body posture (Fellegy, 1995), which were not captured in the NSC.

This means that some linguistic features, such as backchanneling, are subjected to more levels of interpretations and ambiguity that cannot be easily segmented and captured unless there is a more dedicated manual system of notation with greater complexity and granularity for the corpus. In other words, some variables, like backchannels require more paralinguistic and contextual notations, and are not as useful when it comes to corpus-based research.

5.4 Intersectional Factors with Gender

As discussed in Section 2.2, research on gender and language run the risk of being reductive and essentialist which is why the Dynamic approach espouses intersectionality theory as a more nuanced way for seeing how social factors come together to influence language use.

Though ethnicity (or more commonly, race) is one of the more significant variables in gender research, there were no interaction effects of ethnicity groups and speaker gender found in this study (see Section 4.5.1 *Considering Ethnicity*). Also, when comparing dyads in mixed-ethnicity with same-ethnicity contexts, results for all variables showed no differences between these two groups, except for backchannelling. However, the possible issue with backchannels and the open nature of its word class has been discussed above and should be considered possible point of inaccuracy for the variable.

This thesis does not suggest that factors outside of gender does not have an impact on language use, nor does it downplay the importance of the multiplicity in human experiences for people of different demographics. In fact, linear models with age as a factor have revealed that it has an effect on the dependent variables being measured, and age groups as a factor predicting differences in social distance and language and use is a topic worthy of further investigation in future research.

5.5 Language Style Matching and Cultural Scripts

In this thesis, I also utilised a supplementary measure, Language Style Matching, to look at how differently dyads of different gender pairings match each other in linguistic style. As concluded in Section 4.4 (*Language Style Matching Scores*), the results show that same-gender female dyads style-match significantly more than same-gender male dyads.

Results refuted Hypothesis 4, which means that the outcomes of LSM score analyses does not align with the premise that same-gender dyads that are closer in social distance style-match more than mixed-gender dyads. Instead, results show a narrative that is congruent to the literature regarding speaking styles of men and women — a more collaborative style of speech among women and a more competitive speech style among men (Coates, 2012) that could explain the disparity between the high LSM scores for women's talk and comparatively low LSM scores for men's talk.

This ties in with a crucial point offered by Cameron (1997) when she commented on recognisable scenarios that exemplify gender stereotypes and language in popular culture:

> "Is it because we have actually witnessed these scenarios occurring in real life, or is it because we can so readily supply the cultural script that makes them meaningful and 'typical'? One argument for the latter possibility is that if you *reverse* the genders in Tannen's anecdotes, it is still possible to supply a script which makes sense of the alleged gender difference." (p.48).

In the above quote, Cameron suggests a retrospective aspect to gender-based justifications in linguistic differences. There seems to exist a reasonable sociocultural explanation or academic theory to fit any phenomenon observed in results. Therefore, the essential question then becomes: In a given geographical region, and cultural and relationship context, how do the results of a study fit into existing theories of the particular field? For instance, in the LSM example above, the postulated hypothesis is based on a theory of CAT (Giles & Ogay, 2007), but the results showed more agreement with interactional style theory proposed in language and gender (Cameron, 1997; Davies, 2003; Coates, 2012). Furthermore, if for example, mixed-gender dyads style-match more than same-gender dyads, it could be concluded that in mixed-gender interactions, a mutual stylistic convergence took place due to speaker's desire for social attractiveness (Giles et al., 1991) and to be well-liked by the conversational partner of another gender (Goffman, 1977). Going back to Cameron's comment above, through the many ways results could go, it is possible to fit a cultural script with theoretical weight to explain the outcomes.

However, whether or not these cultural scripts and theories are rooted in gender biases, their importance and relevance is to explain the observed trends, and ultimately, involve further research to test out these observations in more isolated, focused studies. This allows a more definitive conclusion of the effect of gender on language, which in turn reflects the importance of the role of gender in a speech communities' interactional dynamics and social construction of a speaker's identities using speech. Hence, in the next chapter, I will discuss what are the limitations of my study, and the possibilities for future research to improve my conclusions of gendered interaction in Singapore.

Chapter 6 Limitations and Future Research

6.1 Unmatched Participants

In my analyses, the participants drawn from Part 3 of the NSC are all unique, meaning that none of the participants attempted the study more than once. While this means that there is independence of the data points, this also means that the comparisons across same-gender and mixed-gender dyads are unmatched, that is, the same participant did not have the chance to interact in both same-gendered and mixed-gendered conditions. This means that the trendlines of the graphs shown in the Results chapter is not strictly the trendlines of the participants who spoke in one condition and then the other, where it can *actually* be concluded that participants increased/decreased their use of the linguistic feature.

Therefore, it is not fully observable whether differences are due to the difference in dyad condition or idiosyncrasies of participants in a particular dyad type. For example, a situation in which a certain participant is just a man who uses a lot of filled pauses, thus inflating the amount filled pauses of whichever group the man is in. Despite this, with the large number of participants considered for both dyad types in this thesis, these idiosyncratic effects of individual participants is attenuated and would not result in too drastic of differences. Additionally, the linear mixed effects model was able to account for random effects on the pair-level at the very least. Hence, this issue of unmatched participants is not major, but is still less than ideal and could be misleading if the trendlines are viewed as accommodation effects of the individuals across dyad type conditions.

6.2 Gender Salience

Another side effect of corpus-based observational study is that gender differences was usually not the primary impetus for the data collection. Unlike studies such as Reid et al.'s (2003), which included a pre-study brief to prime gender salience in half of the participants before the commencement of their conversational discussion, the NSC is not created with gender differences in mind, and hence, gender salience in participants was not a consideration. This means that in the same-gender and mixed-gender interactions of the NSC, the awareness of the speakers' and their interlocutors' gender were never brought to the forefront of their consciousness for any linguistic change observed to be a direct result of gender-primed cognition and behaviour. Ideally, to find out if a certain factor elicits changes in language use, experimental design should dictate all factors to be held constant except for the factor in question to prevent any confounding that could detract from the efficacy of the study.

While it could be argued that there is an advantage for unprimed studies to be more naturalistic and more reflective of real-life interactions, the downside is that one cannot fully ascertain whether gender is the main factor moderating these linguistic features in the background noise of various other social factors. While the results of the analyses showed that there were no interaction effects between gender and ethnicity and gender and age, the presence of main effects for both ethnicity and age show that they have at least some correlation with the variables being investigated. Hence, the findings in this thesis could be benefit from further testing with a more focused study that takes advantage of gender salience to better highlight and analyse the relationship between gender and language use in relation to formality, politeness, and social distance.

6.3 Considering Temporal Dimensions of CAT

Similar to this thesis, many studies investigating gender differences utilise CAT through cross-categorical comparisons of mixed-gender and same-gender dyads (e.g. McMillan, 1977; Mulac, 1988; Reid 1995; Hilte, 2022). However, communication accommodation is a temporal process in which interlocutors behave and react in accordance with what was said and how it was said in an interaction (Hannah & Murachver, 2007, Dragojevic et al., 2016), so there is an added value in looking within each interaction to observe accommodation effects across the duration of the conversation. For example, a study from Pfiester (2009) reported that men decreased their use of backchannels over time with strangers while women increased their backchannel usage, and Hannah and Murachver (2007) saw men's and women's speech tend towards their respective gendered patterns over time.

Dragojevic et al. (2016) commented that, "…one's stance is dynamic and has the potential to shift on an ongoing basis throughout the encounter (see Genesee & Bourhis, 1982), as interlocutors react and respond to perceptions of each other's behaviors, needs, and motives." (p.45), which shows the ever-changing nature of speaker stance throughout a conversation that causes communication accommodation to be a very variable and flexible process. However, since the analyses in this thesis involves a methodology that is atemporal, wherein transcripts are taken in its entirety with count occurrences and rates of the linguistic features in question, results did not capture time-based changes that occur throughout conversation. Therefore, for improvement, certain approaches can be adopted, such as Roger and Nesshoever (1987) who split the conversation in the first half and second half, or Tan (2017) who took the first minute and penultimate minute of the conversation for comparison to compare changes within a conversation. This temporal aspect of CAT can provide an additional dimension of granularity to the results on language accommodation.

6.4 Swearing and Corpus Limitations

Initially, one other dependent variable that this thesis was interested in exploring is the use of swear words by men and women in different gender dyads. Swear words are very clear indicators of informality and expressions of solidarity, especially for younger Singaporeans (Tan et al., 2021), and thus, swearing would serve as a useful data point to answer the research question of this thesis (see Section 2.4.4 *Swearing*). According to studies involving swearing over the years by Timothy Jay (1992; 2000; Jay & Janschewitz, 2008), 10 of the most frequent swear words comprise 80% of swearing data (Jay, 2009) — *fuck*, *shit*, *hell*, *damn*, *goddamn*, *Jesus Christ*, *ass*, *oh my god*, *bitch*, and *sucks*.

However, in the recording guide during the collection of the NSC, participants were explicitly told to refrain from using vulgarities. In addition, even when participants did use swear words during the recording, these expletives — *fuck*, *nabeh*, *wahlan*, *whore*, *slut*, *kimak*, *bitch*, *asshole* — were manually flagged and subsequently censored before the NSC was published to the public, and hence, the data for these swear words are unavailable. Unfortunately, this means that an analysis on swearing cannot be carried out as there would be significant gaps in the database for any accurate representation of swearing to be obtained.

Although the post-processing censorship was not perfect (277 tokens of *fuck* and 85 tokens of *bitch* were still present), an initial attempt was made, working with the remaining swear words — *damn*, *shit*, *hell*, as well as euphemisms of *fuck,* such as *freak*/*frick* (Adams, 1999). However, the gaps caused by the censorship of major swear words rendered this attempt methodologically unsound and cannot provide a holistic, accurate look on swear words to draw meaningful conclusions on language change in gendered interactions.

6.5 Future Directions

This thesis has taken an extensive look at language use in interactions between men and women in Singapore and how they differ in same-gender and mixed-gender dyads. One thing that particularly stood out and worthy of further research is the use of Singlish particles in same-gender and mixed-gender dyads. This research revealed that Singlish, the L-variety of Singapore English, seems to lie in an interesting intersection of politeness, solidarity, and expression of masculinity. This is to be expected as they are discourse particles that do various pragmatic work. While there has been research done in said areas, the complexity of Singlish particles needs to be further explored due to some of its conflicting results in the respective areas. For politeness, there are researchers that claim Singlish to be polite (Stadler, 2018), rude (Yoong, 2010), and both polite and impolite (Lai & Tan, 2023). For solidarity, the general consensus is that Singlish is a solidarity marker (Brown, 1999; Koh, 2009), but this fact is not reflected in solidarity ratings when tested (Cavallaro & Ng, 2009) due to warped perceptions of Singlish (Rubdy, 2001). For masculinity, while Singlish itself is not necessarily a gendered variety of Singapore English, more "extreme" or "crude" forms of Singlish are used and associated with men (Lim, 2009). There seems to be more than meets the eye regarding the role of Singlish when it comes to these various functions it serves for Singaporeans, and how it changes when these areas of politeness, solidarity, and masculinity intersect.

Additionally, research looking at swearing would also be especially helpful to further ascertain findings in this study and investigate relationship between social distance and dyadic interactions between genders, since the data in the NSC was not suited to investigate this feature. This is because like Singlish particles, they are solidarity markers that are also considered impolite and expressions of masculinity; and like filled pauses, they occur less frequently in more formal contexts. This makes swear words another good proxy to gauge social distance between interlocutors and how their usage changes in interactions of different gender dyads.

Furthermore, different methodologies can also be considered for further probing on this topic, such as Critical Discourse Analysis (CDA) or conversation analysis (CA), to gain a more qualitative look at how the use of Singlish and Singlish particles reflect current cultural discourses and the pragmatic work they do in context. This is also especially crucial when we want to consider non-traditional, non-heteronormative expressions of gender, sexuality, and sexual orientation, which corpus-based studies such as this thesis usually cannot account for.

Chapter 7 Conclusion

In this thesis, I examined linguistic features used by men and women in Singapore in two different contexts — same-gender and mixed-gender dyads. This thesis is interested in the effects of speaker gender and dyad type (and therefore partner gender) on three linguistic features that serve as indicators of social distance — backchannels, filled pauses, and Singlish particles. By using linear mixed effects modelling (LMMs), it is found that speaker gender is a significant predictor for all three variables, showing that women use more backchannels than men, and men use more filled pauses and Singlish particles than women, aligning with the hypotheses and previous studies. More crucially, using Communication Accommodation Theory (CAT) to look at trends from same-gender to mixed-gender dyads, results show that the use of backchannels between men and women diverged, which was not expected, while use of filled pauses and Singlish particles converged as expected. The convergence patterns of filled pauses and Singlish particles usage show that the larger social distance between men and women was bridged through the phenomenon of both linguistic convergence (adoption of the interlocutor's frequency of filled pauses) and psychological convergence (reduced use of Singlish particles to be more formal/standard with the opposite gender).

An additional variable, Language Style Matching (LSM) scores, is measured against the participant pair types — male-male (MM), female-female (FF), and male-female (MF). For this variable, FF pairs score significantly than MM pairs, meaning that women match each in linguistic style more than men do, showing agreement with theory of cooperative talk between women and competitive talk between men. Their lack of significant differences with MF pairs shows that there were attempts by both genders to accommodate their speech styles to each other.

The study here revealed the weaknesses of open-class word categories, such as backchannels, in large-scale corpus studies, and conversely solidified the use of closed-class word categories, like culture-specific discourse markers such as Singlish particles. However, the use and function of Singlish particles as it relates to themes of politeness, solidarity, and masculinity are topics that require more work and further research.

Nevertheless, even as studies on gender and language differences need to be treated with more caution and caveats, it is still important for them to serve as reflections to our heavily gendered world and lead us to a closer understanding of interactions between men and women, such as the speakers' tendencies in language use, the features that are indexical to masculine and feminine styles of speech, and the strategies they employ to bridge communication with different individuals of varying social distances and social identities.

References

Acton, E. K. (2011). On Gender Differences in the Distribution of um and uh. *University of Pennsylvania Working Papers in Linguistics*, *17*(2), Article 2.

Adams, M. (1999). Another Effing Euphemism. *American Speech*, *74*(1), 110–112.

Alsagoff, L. (2010). English in Singapore: Culture, capital and identity in linguistic variation. *World Englishes*, *29*(3), 336–348. https://doi.org/10.1111/j.1467-971X.2010.01658.x

Anderson, C., Bjorkman, B., Denis, D., Doner, J., Grant, M., Sanders, N., & Taniguchi, A. (2022). 10.5 Variationist methods and concepts. In *Essentials of Linguistics* (2nd Edition). eCampusOntario. https://ecampusontario.pressbooks.pub/essentialsoflinguistics2/chapter/10-5-variationist-methods-and-concepts/

Anthony, L. (2020). *Antconc (Version 3.5.9)* [Computer software]. Waseda University, Tokyo. https://www.laurenceanthony.net/software

Baker, R., & Hazan, V. (2011). DiapixUK: Task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior Research Methods*, *43*(3), 761–770.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Baumel, J. (2020). *Gendered Use of Sentence-Final Particles in Taiwan Mandarin: A Corpus Inquiry* [Bachelor's Thesis]. Harvard College.

Baxter, J. (2006). *Speaking Out: The Female Voice in Public Contexts*. Palgrave Macmillan.

Bayard, D. (1995). *Kiwitalk: Sociolinguistics and New Zealand society*. Dunmore Press.

Bierstetel, S. J., Farrell, A. K., Briskin, J. L., Harvey, M. W., Gable, S. L., Ha, T., Ickes, W., Lin, W.-F., Orina, M. M., Saxbe, D., Simpson, J. A., Ta, V. P., & Slatcher, R. B. (2020). Associations between language style matching and relationship commitment and satisfaction: An integrative data analysis. *Journal of Social and Personal Relationships*, *37*(8–9), 2459–2481. https://doi.org/10.1177/0265407520923754

Bilous, F. R., & Krauss, R. M. (1988). Dominance and accommodation in the conversational behaviours of same- and mixed-gender dyads. *Language & Communication*, *8*(3), 183–194. https://doi.org/10.1016/0271-5309(88)90016-X

Blechner, M. (1998). Maleness and Masculinity. *Contemporary Psychoanalysis*, *34*(4), 597–613. https://doi.org/10.1080/00107530.1998.10747024

Boersma, P., & Weenink, D. (2023). *Praat: Doing phonetics by computer (Version 6.3.10)* [Computer Software]. http://www.praat.org/

Boo, A., Lee, J., & Tan, Y.-Y. (2023). Particle stacking in Singlish – New data from the National Speech Corpus. *Lingua*, *287*, 103513. https://doi.org/10.1016/j.lingua.2023.103513

Bořil, T., & Skarnitzl, R. (2016). Tools rPraat and mPraat. In P. Sojka, A. Horák, I. Kopeček, & K. Pala (Eds.), *Text, Speech, and Dialogue* (pp. 367–374). Springer International Publishing. https://doi.org/10.1007/978-3-319-45510-5_42

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role, and Gender. *Language and Speech*, *44*(2), 123–147. https://doi.org/10.1177/00238309010440020101

Botha, W. (2018). A social network approach to particles in Singapore English. *World Englishes*, *37*(2), 261–281. https://doi.org/10.1111/weng.12250

Broen, P. A., & Siegel, G. M. (1972). Variations in Normal Speech Disfluencies. *Language and Speech*, *15*(3), 219–231. https://doi.org/10.1177/002383097201500302

Brown, A. (1999). *Singapore English in a Nutshell: An Alphabetical Description of Its Features*. Federal Publications.

Brown, P., & Levinson, S. C. (1978). Universals in language use: Politeness phenomena. In E. N. Goody (Ed.), *Questions and politenes: Strategies in social interaction* (pp. 56–311). Cambridge University Press.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language use*. Cambridge University Press.

Bucholtz, M. (2014). The Feminist Foundations of Language, Gender, and Sexuality Research. In S. Ehrlich, M. Meyerhoff, & J. Holmes (Eds.), *The Handbook of Language, Gender, and Sexuality* (2nd ed., pp. 23–47). John Wiley & Sons.

Burleson, B., & Kunkel, A. (2006). Revisiting the Different Cultures Thesis: An Assessment of Sex Differences and Similarities in Supportive Communication. *Sex Differences and Similarities in Communication 2nd Ed*.

Butler, J. (1988). Performative acts and gender constitution: An essay in phenomenology and feminist theory. *Theatre Journal*, *40*(4), 519–531.

Butler, J. (1993). *Bodies That Matter*. Routledge.

Butler, J. (1999). *Gender Trouble: Feminism and the Subversion of Identity*. Routledge.

Byrne, D. (1971). *The attraction paradigm*. Academic.

Cameron, D. (1995). Rethinking language and gender studies: Some issues for the 1990s. In S. Mills (Ed.), *Language and Gender: Interdisciplinary Perspectives* (pp. 31–44). Longman.

Cameron, D. (1997). Performing Gender Identity: Young Men's Talk and the Construction of Heterosexual Masculinity. In S. Johnson & U. H. Meinhof (Eds.), *Language and Masculinity*. Blackwell.

Cameron, D. (2020). Language and gender: Mainstreaming and the persistence of patriarchy. *International Journal of the Sociology of Language*, *2020*(263), 25–30. https://doi.org/10.1515/ijsl-2020-2078

Cameron, D., McAlinden, F., & O'Leary, K. (1989). Lakoff in context: The social and linguistic functions of tag questions. In J. Coates & D. Cameron (Eds.), *Women in their speech communities* (pp. 74–93). Longman. https://web.stanford.edu/~eckert/PDF/CameronTags.pdf

Canary, D. J., & Hause, K. S. (1993). Is there any reason to research sex differences in communication? *Communication Quarterly*, *41*, 129–144.

Cavallaro, F., & Ng, B. C. (2009). Between status and solidarity in Singapore. *World Englishes*, *28*(2), 143–159.

Cavallaro, F., Ng, B. C., & Seilhamer, M. F. (2014). *Singapore Colloquial English: Issues of prestige and identity*. https://doi.org/10.1111/weng.12096

Chan, M. K. M. (1998). Gender differences in the Chinese language: A preliminary report. In H. Lin (Ed.), *Ninth North American Conference on Chinese Linguistics (NACCL-9)* (Vol. 2, pp. 35–52). GSIL Publications, University of California.

Chan, M. K. M. (1999). *Sentence-Final Particles in Cantonese: A Gender-Linked Survey and Study*. 87–101. https://www.semanticscholar.org/paper/Sentence-Final-Particles-in-Cantonese-%3A-A-Survey-Chan/06bdcfc4e56ab2b2466152ed0460d44344eaf545

Chan, M. K. M., & Lin, Y. (2019). Chinese language and gender research. In *The Routledge Handbook of Chinese Applied Linguistics*. Routledge.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*(6), 893–910. https://doi.org/10.1037/0022-3514.76.6.893

Chung, C., & Pennebaker, J. (2007). The Psychological Functions of Function Words. In *Social communication* (pp. 343–359). Psychology Press.

Cieri, C., Graff, D., Kimball, O., Miller, D., & Walker, K. (2004). *Fisher English Training Speech Part 1 Transcripts LDC2004T19*. Linguistic Data Consortium.

Cieri, C., Graff, D., Kimball, O., Miller, D., & Walker, K. (2005). *Fisher English Training Part 2, Transcripts LDC2005T19*. Linguistic Data Consortium.

Clark, H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*, 73–111. https://doi.org/10.1016/S0010-0277(02)00017-3

Coates, J. (1996). *Women Talk: Conversation Between Women Friends*. Blackwell.

Coates, J. (1997). The construction of a collaborative floor in women's friendly talk. In T. Givon (Ed.), *Conversation: Cognitive, Communicative and Social Perspectives* (pp. 55–89). John Benjamins.

Coates, J. (2003). *Men Talk: Stories in the Making of Masculinities*. Blackwell.

Coates, J. (2012). Gender and discourse analysis. In M. Handford & J. P. Gee (Eds.), *The Routledge Handbook of Discourse Analysis*. Routledge.

Coates, J. (2013). *Women, Men and Everyday Talk*. Palgrave Macmillan UK. https://doi.org/10.1057/9781137314949

Coates, J. (Ed.). (2015). *Women, Men and Language: A Sociolinguistic Account of Gender Differences in Language* (3rd ed.). Routledge. https://doi.org/10.4324/9781315645612

Cooper, B. (2015). Intersectionality. In L. Disch & M. Hawkesworth (Eds.), *The Oxford Handbook of Feminist Theory* (pp. 385–406). Oxford University Press.

Crenshaw, K. (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*, *1989*(1). https://chicagounbound.uchicago.edu/uclf/vol1989/iss1/8

Davies, J. (2003). Expressions of Gender: An Analysis of Pupils' Gendered Discourse Styles in Small Group Classroom Discussions. *Discourse & Society - DISCOURSE SOCIETY*, *14*, 115–132. https://doi.org/10.1177/0957926503014002853

Davis, G., & Preves, S. (2017). Intersex and the Social Construction of Sex. *Contexts*, *16*(1), 80–80. https://doi.org/10.1177/1536504217696082

de Beauvoir, S. (1949). *Le Deuxième Sexe*. Gallimard.

Deborah, S. (1987). *Discourse Markers*. Cambridge University Press.

Delaherche, E., Chetouani, M., Mahdhaoui, A., Saint-Georges, C., Viaux, S., & Cohen, D. (2012). Interpersonal Synchrony: A Survey of Evaluation Methods across Disciplines. *IEEE Transactions on Affective Computing*, *3*(3), 349–365. https://doi.org/10.1109/T-AFFC.2012.12

Dixon, J. A., & Foster, D. H. (1998). Gender, Social Context, and Backchannel Responses. *The Journal of Social Psychology*, *138*(1), 134–136. https://doi.org/10.1080/00224549809600364

Dragojevic, M., Gasiorek, J., & Giles, H. (2016). Accommodative Strategies as Core of the Theory. In H. Giles (Ed.), *Communication Accommodation Theory: Negotiating Personal Relationships and Social Identities across Contexts* (pp. 36–59). Cambridge University Press. https://doi.org/10.1017/CBO9781316226537.003

Drummond, K., & Hopper, R. (1993). Back Channels Revisited: Acknowledgment Tokens and Speakership Incipiency. *Research on Language and Social Interaction*, *26*(2), 157–177. https://doi.org/10.1207/s15327973rlsi2602_3

Duncan, S., & Fiske, D. W. (1977). *Face to face interaction*. Erlbaum.

Duncan, S., & Niederehe, G. (1974). On signalling that it's your turn to speak. *Journal of Experimental Social Psychology*, *10*(3), 234–247.

Eckert, P., & McConnell-Ginet, S. (1992). Communities of practice: Where language, gender, and power all live. In K. Hall, M. Bucholtz, & B. Moonwomon (Eds.), *Locating Power: Proceedings of the Second Berkeley Women and language Conference* (pp. 89–99). University of California Press.

Eckert, P., & McConnell-Ginet, S. (2003). *Language and Gender*. Cambridge University Press. https://doi.org/10.1017/CBO9780511791147

Elhami, A. (2020). Communication Accommodation Theory: A Brief Review of the Literature. *Journal of Advances in Education and Philosophy*, *4*, 192–200. https://doi.org/10.36348/jaep.2020.v04i05.002

Endendijk, J. J., van Baar, A. L., & Deković, M. (2020). He is a Stud, She is a Slut! A Meta-Analysis on the Continued Existence of Sexual Double Standards. *Personality and Social Psychology Review*, *24*(2), 163–190. https://doi.org/10.1177/1088868319891310

Entwistle, J. (2000). *The Fashioned Body: Fashion, Dress and Modern Social Theory*. Polity.

Farris, C. (1995). A semeiotic analysis of sajiao as a gender marked communication style in Chinese. In J. Marshall & F. Y. L. Chiu (Eds.), *Unbound Taiwan: Closeups from a Distance* (pp. 1–29). Center for East Asian Studies.

Feldman, R. (2003). Infant-mother and infant-father synchrony: The coregulation of positive arousal. *Infant Mental Health Journal*, *24*(1), 1–23. https://doi.org/10.1002/imhj.10041

Fellegy, A. M. (1995). Patterns and Functions of Minimal Response. *American Speech*, *70*(2), 186–199. https://doi.org/10.2307/455815

Fishman, P. (1980). Conversational insecurity. In H. Giles, W. P. Robinson, & P. M. Smith (Eds.), *Language: Social psychological perspectives* (pp. 127–132). Pergamon.

Fishman, P. M. (1978). Interaction: The Work Women Do. *Social Problems*, *25*(4), 397–406. https://doi.org/10.2307/800492

Fitzpatrick, M., Mulac, A., & Dindia, K. (1995). Gender-Preferential Language Use in Spouse and Stranger Interaction. *Journal of Language and Social Psychology*, *14*(1–2), 18–39. https://doi.org/10.1177/0261927X95141002

Fixmer-Oraiz, N., & Wood, J. T. (2019). *Gendered Lives: Communication, Gender, & Culture* (13th ed.). Cengage.

Frances, S. J. (1979). Sex differences in nonverbal behavior. *Sex Roles*, *5*, 519–535.

Gallois, C., Ogay, T., & Giles, H. (2005). Communciation Accommodation Theory: A look back and a look ahead. In W. B. Gudykunst, *Theorizing about intercultural communication* (pp. 121–148). Sage.

Genesee, F., & Bourhis, R. (1982). The Social Psychological Significance of Code Switching in Cross-Cultural Communication. *Journal of Language and Social Psychology*, *1*(1), 1–27. https://doi.org/10.1177/0261927X8200100102

Giles, H. (1973). Accent Mobility: A model and some data. *Anthropological Linguistics*, *15*(2), 87–109.

Giles, H. (1978). Linguistic differentiation between ethnic groups. In H. Tajfel, *Differentiation between social groups: Studies in the social psychology of intergroup relations* (pp. 361–393). Academic Press.

Giles, H. (2016). Communication Accommodation Theory. In *The International Encyclopedia of Communication Theory and Philosophy* (pp. 1–7). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118766804.wbiect056

Giles, H., Coupland, J., & Coupland, N. (1991). Accommodation theory: Communication, contexts and consequence. In H. Giles, N. Coupland, & J. Coupland, *Contexts of accommodation: Developments in applied sociolinguistics* (pp. 1–68). Cambridge University Press.

Giles, H., & Ogay, T. (2007). Communication Accommodation Theory. In *Explaining communication: Contemporary theories and exemplars* (pp. 293–310). Lawrence Erlbaum Associates Publishers.

Giles, H., & Smith, P. M. (1979). Accommodation theory: Optimal levels ofconvergence. In H. Giles & R. N. St. Clair (Eds.), *Language and social psychology* (pp. 45–65). Blackwell.

Giles, H., Taylor, D., & Bourhis, R. (1973). Towards a Theory of Interpersonal Accommodation through Language: Some Canadian Data. *Language in Society*, *2*, 177–192. https://doi.org/10.1017/S0047404500000701

Godfrey, J. J., Holliman, E. C., & McDaniel, J. (1992). SWITCHBOARD: telephone speech corpus for research and development. *Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing*, *1*, 517–520.

Goffman, E. (1976). Gender Display. In E. Goffman (Ed.), *Gender advertisements: Communication and Culture* (pp. 69–77). Palgrave Macmillan.

Goffman, E. (1977). The Arrangement between the Sexes. *Theory and Society*, *4*(3), 301–331.

Goh, R. B. H. (2016). The anatomy of Singlish: Globalisation, multiculturalism and the construction of the 'local' in Singapore. *Journal of Multilingual and Multicultural Development*, *37*(8), 748–758. https://doi.org/10.1080/01434632.2015.1133628

Gonzales, A. L., Hancock, J. T., & Pennebaker, J. W. (2010). Language Style Matching as a Predictor of Social Dynamics in Small Groups. *Communication Research*, *37*(1), 3–19. https://doi.org/10.1177/0093650209351468

Goodwin, M. H. (1980). Directive-response speech sequences in girls' and boys' task activities. In S. McConnell-Ginet, R. Borker, & N. Furman, *Women and Language in Literature and Society* (pp. 157–173). Praeger.

Gries, S. (2015). The most under-used statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora*, *10*, 95–125. https://doi.org/10.3366/cor.2015.0068

Gudykunst, W. B. (1995). Anxiety/uncertainty management (AUM) theory: Current status. In R. L. Wiseman (Ed.), *Intercultural communication theory* (pp. 8–58). Thousand Oaks, CA.

Gupta, A. F. (1989). Singapore Colloquial English and Standard English. *Singapore Journal of Education*, *10*(2), 33–39. https://doi.org/10.1080/02188798908547659

Gupta, A. F. (1992). The pragmatic particles of Singapore colloquial English. *Journal of Pragmatics*, *18*(1), 31–57. https://doi.org/10.1016/0378-2166(92)90106-L

Gupta, A. F. (1994). *The Step-Tongue: Children's English in Singapore*. Multilingual Matters.

Gupta, A. F. (2006). Epistemic modalities and the discourse particles of Singapore. In K. Fischer (Ed.), *Approaches to Discourse Particles* (pp. 243–263). Elsevier.

Güvendir, E. (2015). Why are males inclined to use strong swear words more than females? An evolutionary explanation based on male intergroup aggressiveness | Elsevier Enhanced Reader. *Language Sciences*, *50*, 133s139. https://doi.org/10.1016/j.langsci.2015.02.003

Hall, J. A., & Roter, D. L. (2002). Do patients talk differently to male and female physicians?: A meta-analytic review. *Patient Education and Counseling*, *48*(3), 217–224. https://doi.org/10.1016/S0738-3991(02)00174-X

Hall, K. (2003). Exceptional Speakers: Contested and Problematized Gender Identities. In M. Meyerhoff & J. Holmes (Eds.), *Handbook of Language and Gender* (pp. 352–380). Basil Blackwell.

Hall, K., Borba, R., & Hiramoto, M. (2020). *Language and Gender* (pp. 892–913). https://doi.org/10.1002/9781118786093.iela0143

Hancock, A. B., & Rubin, B. A. (2015). Influence of Communication Partner's Gender on Language. *Journal of Language and Social Psychology*, *34*(1), 46–64. https://doi.org/10.1177/0261927X14533197

Hannah, A., & Murachver, T. (2007). Gender Preferential Responses to Speech. *Journal of Language and Social Psychology*, *26*(3), 274–290. https://doi.org/10.1177/0261927X06303457

Hare-Mustin, R. T., & Marecek, J. (1988). The Meaning of Difference. *American Psychologist*, *43*(6), 455–464.

Heinz, B. (2003). Backchannel responses as strategic responses in bilingual speakers' conversations. *Journal of Pragmatics*, *35*(7), 1113–1142. https://doi.org/10.1016/S0378-2166(02)00190-X

Hilte, L., Vandekerckhove, R., & Daelemans, W. (2022). Linguistic Accommodation in Teenagers' Social Media Writing: Convergence Patterns in Mixed-gender Conversations. *Journal of Quantitative Linguistics*, *29*(2), 241–268. https://doi.org/10.1080/09296174.2020.1807853

Ho, C., & Chionh, F. J. P. (2009). Do Singaporean youth use language differently? Gender and language use in personal weblogs. *2009 HICE Proceedings*, 80–91.

Hoffman, R., Hattie, J., & Borders, L. (2005). Personal Definitions of Masculinity and Femininity as an Aspect of Gender Self-Concept. *The Journal of Humanistic Counseling, Education and Development*, *44*. https://doi.org/10.1002/j.2164-490X.2005.tb00057.x

Hoffman, S., Blass, A.-K., & Mukherjee, J. (2014). Canonical Tag Questions in Asian Englishes: Forms, Functions, and Frequencies in Hong Kong English, Indian English, and Singapore English. In M. Filppula, J. Klemola, & D. Sharma (Eds.), *The Oxford Handbook of World Englishes* (pp. 697–714). Oxford Handbooks. https://doi.org/10.1093/oxfordhb/9780199777716.013.025

Hogg, M. A. (1985). Masculine and Feminine Speech in Dyads and Groups: A Study of Speech Style and Gender Salience. *Journal of Language and Social Psychology*, *4*(2), 99–112. https://doi.org/10.1177/0261927X8500400202

Holmes, J. (1982). The functions of tag questions. *English Language Research Journal*, *3*, 40–65.

Holmes, J. (1984). Hedging your bets and sitting on the fence: Some evidence for hedges as support structures. *Te Reo*, *27*.

Holmes, J. (1990). Hedges and boosters in women's and men's speech. *Language & Communication*, *10*(3), 185–205. https://doi.org/10.1016/0271-5309(90)90002-S

Holmes, J. (1995). *Women, Men and Politeness*. Longman.

Holmes, J., & Meyerhoff, M. (2003). *The Handbook of Language and Gender*. Blackwell.

Hong, W. (1997). Gender Differences in Chinese Request Patterns. *Journal of Chinese Linguistics*, *25*(2), 193–210.

IMDA. (2023). *National Speech Corpus*. Infocomm Media Development Authority. https://www.imda.gov.sg/how-we-can-help/national-speech-corpus

Inoue, M. (2002). Gender, Language, and Modernity: Toward an Effective History of Japanese Women's Language. *American Ethnologist*, *29*(2), 392–422.

Ireland, M. E., & Pennebaker, J. W. (2010). Language style matching in writing: Synchrony in essays, correspondence, and poetry. *Journal of Personality and Social Psychology*, *99*(3), 549–571. https://doi.org/10.1037/a0020386

Ireland, M., Slatcher, R., Eastwick, P., Scissors, L., Finkel, E., & Pennebaker, J. (2011). Language Style Matching Predicts Relationship Initiation and Stability. *Psychological Science*, *22*, 39–44. https://doi.org/10.1177/0956797610392928

Jacob, C., Guéguen, N., Martin, A., & Boulbry, G. (2011). Retail salespeople's mimicry of customers: Effects on consumer behavior. *Journal of Retailing and Consumer Services*, *18*(5), 381–388. https://doi.org/10.1016/j.jretconser.2010.11.006

James, D., & Drakich, J. (1993). Understanding gender differences in amount of talk: A critical review of research. In *Gender and conversational interaction* (pp. 281–312). Oxford University Press.

Jay, T. (1992). *Cursing in America*. John Benjamins Publishing Company.

Jay, T. (2000). *Why We Curse*. John Benjamins Publishing Company.

Jay, T. (2009). The Utility and Ubiquity of Taboo Words. *Perspectives on Psychological Science*, *4*(2), 153–161. https://doi.org/10.1111/j.1745-6924.2009.01115.x

Jay, T., & Janschewitz, K. (2008). The pragmatics of swearing. *Journal of Politeness Research. Language, Behaviour, Culture*, *4*, 267–288. https://doi.org/10.1515/JPLR.2008.013

John, P., & Weber, H. (1980). *English in Singapore and Malaysia*. Oxford University Press.

Kachru, B. B. (1985). Standard, Codification and Sociolinguistic Realism: The English Language in the Outer Circle. In R. Quirk & H. Widdowson (Eds.), *English in the World: Teaching and Learning the Language and Literatures* (pp. 11–30). Cambridge University Press.

Kaplan, A. (2016). *Women Talk More than Men...And Other Myths about Language Explained*. Cambridge University Press.

Kiesling, S. (2007). Men, Masculinities, and Language. *Language and Linguistics Compass*, *1*(6), 653–673. https://doi.org/10.1111/j.1749-818X.2007.00035.x

Kiesling, S. F. (1998). Men's Identities and Sociolinguistic Variation: The Case of Fraternity Men. *Journal of Sociolinguistics*, *2*(1), 69–99. https://doi.org/10.1111/1467-9481.00031

Kogure, M. (2003). *Gender differences in the use of backchannels: Do Japanese men and women accommodate to each other? - ProQuest* [Ph.D Dissertation, The University of Arizona]. https://www.proquest.com/openview/6ebcc3073f97bffdf981822af4d65269/1?pq-origsite=gscholar&cbl=18750&diss=y

Koh, J. X., Mislan, A., Khoo, K., Ang, B., Ang, W., Ng, C., & Tan, Y.-Y. (2019). Building the Singapore English National Speech Corpus. *Interspeech 2019*, 321–325. https://doi.org/10.21437/Interspeech.2019-1525

Koh, Z. A. A. (2009). *A Bourdieuvian Analysis of the Use of Singlish by Youths in Singapore* [Master's Thesis]. National University of Singapore.

Kulesza, W., Dolinski, D., Huisman, A., & Majewski, R. (2014). The Echo Effect: The Power of Verbal Mimicry to Influence Prosocial Behavior. *Journal of Language and Social Psychology*, *33*(2), 183–201. https://doi.org/10.1177/0261927X13506906

L. (2010). English in Singapore: Culture, capital and identity in linguistic variation. *World Englishes*, *29*(3), 336–348.

Lai, J. H. (2019). *"Aiyo! Why like that?" The case of gender and Singlish interjections in text messaging* [Undergraduate Thesis, Nanyang Technological University]. http://hdl.handle.net/10356/76688

Lai, Y.-S., & Tan, Y.-Y. (2023). Singlish particles and politeness work in face-threatening speech acts. *Asian Englishes*, *25*(1), 146–166. https://doi.org/10.1080/13488678.2022.2132130

Lakoff, R. (1973). Language and Woman's Place. *Language in Society*, *2*(1), 45–80.

Lakoff, R. T. (1975). *Language and Woman's Place*. Harper & Row.

Laserna, C. M., Seih, Y.-T., & Pennebaker, J. W. (2014). *Um* . . . Who Like Says *You Know*: Filler Word Use as a Function of Age, Gender, and Personality. *Journal of Language and Social Psychology*, *33*(3), 328–338. https://doi.org/10.1177/0261927X14526993

Leaper, C., & Ayres, M. M. (2007). A Meta-Analytic Review of Gender Variations in Adults' Language Use: Talkativeness, Affiliative Speech, and Assertive Speech. *Personality and Social Psychology Review*, *11*(4), 328–363. https://doi.org/10.1177/1088868307302221

Leaper, C., & Robnett, R. (2011). Women Are More Likely Than Men to Use Tentative Language, Aren't They? A Meta-Analysis Testing for Gender Differences and Moderators. *Psychology of Women Quarterly*, *35*, 129–142. https://doi.org/10.1177/0361684310392728

Lee, S. K. (2018). *A Nanosyntactic Approach to Sentence-Final Particles in Singlish: A Cartographic Perspective* [Bachelor's Thesis, National University of Singapore]. https://scholarbank.nus.edu.sg/handle/10635/143512

Leimgruber, J., Lim, J. J., Gonzales, W. D. W., & Hiramoto, M. (2020). Ethnic and gender variation in the use of Colloquial Singapore English discourse particles. *English Language and Linguistics*, *25*(3), 1–20. https://doi.org/10.1017/S1360674320000453

Lenth, R. V. (2023). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (R package version 1.8.4-1) [Computer software]. https://CRAN.R-project.org/package=emmeans

Ler, S. L. V. (2005). *An in-depth study of discourse in Singapore English* [Doctoral Thesis, National University of Singapore]. https://scholarbank.nus.edu.sg/handle/10635/14757

Ler, S. L. V. (2006). A relevance-theoretic approach to discourse particles in Singapore English. In F. Kerstin (Ed.), *Approaches to Discourse Particles* (pp. 149–166). Elsevier.

Liberman, M. (2014). Language Log» More on UM and UH. *Language Log*. https://languagelog.ldc.upenn.edu/nll/?p=13713

Lim, C. Y. R. (2009). *Negotiating gender identity through the use of Singlish* [Undergraduate Thesis]. Nanyang Technological University.

Lim, L. (2000). Ethnic group differences aligned? Intonation patterns of Chinese, Indian and Malay Singaporean English. In A. Brown, D. Deterding, & E. L. Low (Eds.), *The English Language in Singapore: Research on Pronunciation* (pp. 10–21). Singapore Association for Applied Linguistics.

Lim, L. (2007). Mergers and acquisitions: On the ages and origins of Singapore English particles. *World Englishes*, *26*(4), 446–473. https://doi.org/10.1111/j.1467-971X.2007.00522.x

Lin, H. H. (2005). *Contextualizing linguistic politeness in Chinese—A socio-pragmatic approach with examples from persuasive sales talk in taiwan mandarin* [Doctoral Thesis]. Ohio State University.

MacGeorge, E. L., Graves, A. R., Feng, B., Gillihan, S. J., & Burleson, B. R. (2004). The Myth of Gender Cultures: Similarities Outweigh Differences in Men's and Women's Provision of and Responses to Supportive Communication. *Sex Roles*, *50*(3), 143–175. https://doi.org/10.1023/B:SERS.0000015549.88984.8d

Maltz, D., & Borker, R. (1982). A Cultural approach to male-female miscommunication. In J. J. Gumperz (Ed.), *Language and social identity* (pp. 196–216). Cambridge University Press.

Maschler, Y., & Schiffrin, D. (2015). Discourse Markers Language, Meaning, and Context. In *The Handbook of Discourse Analysis* (pp. 189–221). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118584194.ch9

McConnell-Ginet, S. (1975). Our father tongue: Essays in linguistic politics. *Diacritics*, *5*(4), 44–56.

McMillan, J. R., Clifton, A. K., McGrath, D., & Gale, W. S. (1977). Women's language: Uncertainty or interpersonal sensitivity and emotionality? *Sex Roles*, *3*(6), 545–559. https://doi.org/10.1007/BF00287838

Meister, H., Landwehr, M., Pyschny, V., Walger, M., & Wedel, H. von. (2009). The perception of prosody and speaker gender in normal-hearing listeners and cochlear implant recipients. *International Journal of Audiology*, *48*(1), 38–48. https://doi.org/10.1080/14992020802293539

Mills, S. (2003). *Gender and Politeness*. Cambridge University Press.

Montepare, J. M., & Vega, C. (1988). Women's Vocal Reactions to Intimate and Casual Male Friends. *Personality & Social Psychology Bulletin*, *14*(1), 103–113. https://doi.org/10.1177/0146167288141011

Mott, H., & Petrie, H. (1995). Workplace Interactions: Women's Linguistic Behavior. *Journal of Language and Social Psychology*, *14*(3), 324–336. https://doi.org/10.1177/0261927X95143005

Mulac, A., & Bradac, J. J. (1995). Women's style in problem solving interaction: Powerless, or simply feminine? In P. J. Kalbfleisch & M. J. Cody (Eds.), *Gender, power, and communication in human relationships* (pp. 83–104). Lawrence Erlbaum.

Mulac, A., Bradac, J. J., & Gibbons, P. (2001). Empirical support for the gender-as-culture hypothesis.: An intercultural analysis of male/female language differences. *Human Communication Research*, *27*(1), 121–152. https://doi.org/10.1111/j.1468-2958.2001.tb00778.x

Mulac, A., Erlandson, K. T., Farrar, W. J., Hallett, J. S., Molloy, J. L., & Prescott, M. E. (1998). "Uh-huh. What's That All About?": Differing Interpretations of Conversational Backchannels and Questions as Sources of Miscommunication Across Gender Boundaries. *Communication Research*, *25*(6), 641–668. https://doi.org/10.1177/009365098025006004

Mulac, A., Wiemann, J. M., Widenmann, S. J., & Gibson, T. W. (1988). Male/female language differences and effects in same-sex and mixed-sex dyads: The gender-linked language effect. *Communication Monographs*, *55*(4), 315–335. https://doi.org/10.1080/03637758809376175

Namy, L. L., Nygaard, L. C., & Sauerteig, D. (2002). Gender Differences in Vocal Accommodation: The Role of Perception. *Journal of Language and Social Psychology*, *21*(4), 422–432. https://doi.org/10.1177/026192702237958

Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender Differences in Language Use: An Analysis of 14,000 Text Samples. *Discourse Processes*, *45*(3), 211–236. https://doi.org/10.1080/01638530802073712

Niederhoffer, K. G., & Pennebaker, J. W. (2002). Linguistic Style Matching in Social Interaction. *Journal of Language and Social Psychology*, *21*(4), 337–360. https://doi.org/10.1177/026192702237953

Oakley, A. (1972). *Sex, Gender and Society*. Temple Smith.

Ochs, E. (1990). Indexicality and socialization. In G. Herdt, J. W. Stigler, & R. A. Schweder (Eds.), *Cultural Psychology: Essays on Comparative Human Development* (pp. 287–308). Cambridge University Press. https://doi.org/10.1017/CBO9781139173728.009

Ochs, E. (1992). Indexing Gender. In A. Duranti & C. Goodwin (Eds.), *Rethinking Context: Language as an Interactive Phenomenon* (pp. 335–358). Cambridge University Press.

Okamoto, S. (1997). Social context, linguistic ideology, and indexical expressions in Japanese. *Journal of Pragmatics*, *28*, 795–817.

Okamoto, S. (2013). Variability in societal norms for Japanese women's speech: Implications for linguistic politeness. *Multilingua - Journal of Cross-Cultural and Interlanguage Communication*, *32*(2), 203–223. https://doi.org/10.1515/multi-2013-0010

Okamoto, S., & Sato, S. (1992). Less feminine speech among young Japanese females. In K. Hall, M. Bucholtz, & B. Moonwomon (Eds.), *Locating Power: Proccedings of the 2nd Berkeley Women & Language Conference* (pp. 478–488). University of California Press.

Onnela, J.-P., Waber, B. N., Pentland, A., Schnorf, S., & Lazer, D. (2014). Using sociometers to quantify social interaction patterns. *Scientific Reports*, *4*(1), 1–9.

Paechter, C. (2003). Masculinities and femininities as communities of practice. *Women's Studies International Forum*, *26*(1), 69–77. https://doi.org/10.1016/S0277-5395(02)00356-4

Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015). *Linguistic inquiry and word count: LIWC2015*. Pennebaker Conglomerates. www.LIWC.net

Pfiester, R. A. (2009). *Stereotype threat in mixed-sex dyadic communication* [Ph.D Dissertation]. Univerisity of Texas.

Pickering, M. J., & Garrod, S. (2004). The interactive-alignment model: Developments and refinements—Response. *Behavioral and Brain Sciences*, *27*(2), 212–225. https://doi.org/10.1017/S0140525X04450055

Platt, J. (1987). Communicative functions of particles in Singapore English. In R. Steele & T. Threadgold (Eds.), *Language topics* (Vol. 1, pp. 391–401). Benjamins.

Plug, I., Radboud University, NL, Stommel, W., Radboud University, NL, Lucassen, P. L. B. J., Radboud Univ. Medical Centre, NL, olde Hartman, T. C., Radboud Univ. Medical Centre, NL, Van Dulmen, S., Radboud Inst. for Health Sciences, NL, Das, E., & Radboud University, NL. (2021). Do women and men use language differently in spoken face-to-face interaction? A scoping review. *Review of Communication Research*, *9*, 43–79. https://doi.org/10.12840/ISSN.2255-4165.026

Precht, K. (2002). *Gender differences in affect, evidentiality, and hedging in American conversation* [Paper]. Annual Meeting of the Linguistic Society of America, San Francisco.

Ptaszynski, M., Hasegawa, D., & Masui, F. (2014). Women Like Backchannel, But Men Finish Earlier: Pattern Based Language Modeling of Conversations Reveals Gender and Social Distance Differences. *In Samsung HLT Young Researchers Symposium*

Colocated with 9th International Conference on Natural Language Processing (PolTAL 2014), paper ID: 02.

R Core Team. (2022). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Reid, J. (1995). A study of gender differences in minimal responses. *Journal of Pragmatics*, *24*(5), 489–512. https://doi.org/10.1016/0378-2166(94)00066-N

Reid, S. A., Keerie, N., & Palomares, N. A. (2003). Language, Gender Salience, and Social Influence. *Journal of Language and Social Psychology*, *22*(2), 210–233. https://doi.org/10.1177/0261927X03022002004

Richardson, B. H., McCulloch, K. C., Taylor, P. J., & Wall, H. J. (2019). The cooperation link: Power and context moderate verbal mimicry. *Journal of Experimental Psychology: Applied*, *25*(1), 62–76. https://doi.org/10.1037/xap0000200

Rochon, E., Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (2000). Quantitative Analysis of Aphasic Sentence Production: Further Development and New Data. *Brain and Language*, *72*(3), 193–218. https://doi.org/10.1006/brln.1999.2285

Roger, D., & Nesshoever, W. (1987). Individual differences in dyadic conversational strategies: A further study. *British Journal of Social Psychology*, *26*(3), 247–255. https://doi.org/10.1111/j.2044-8309.1987.tb00786.x

Romaine, S. (2003). Variation in Language and Gender. In J. Holmes & M. Meyerhoff (Eds.), *The Handbook of Language and Gender* (pp. 98–118). Blackwell Publishing Ltd. https://doi.org/10.1002/9780470756942.ch4

RStudio Team. (2020). *RStudio: Integrated Development for R* [Computer software]. RStudio, PBC, Boston, MA. http://www.rstudio.com/

Rubdy, R. (2001). Creative destruction: Sinagpore's Speak Good English movement. *World Englishes*, *20*(3), 341–355. https://doi.org/doi.org/10.1111/1467-971X.00219

Sax, L. (2002). How common is intersex? A response to Anne Fausto-Sterling. *Journal of Sex Research*, *39*(3), 174–178. https://doi.org/10.1080/00224490209552139

Schachter, S., Christenfeld, N., Ravina, B., & Bilous, F. (1991). Speech Disfluency and the Structure of Knowledge. *Journal of Personality and Social Psychology*, *60*(3), 362–367.

Shields, S. A. (2008). Gender: An Intersectionality Perspective. *Sex Roles*, *59*(5–6), 301–311. https://doi.org/10.1007/s11199-008-9501-8

Shih, Y.-H. (1984). SHIH, YU-HUEI (施玉惠). 1984. 從社會逄言學觀點探討中文男女兩性語言的差異 [A sociolinguistic study of gender differences in male and female's language in Chinese) Jiaoxue yu Yanjiu 6:207-229. *教学與研究 (Teaching and Research)*, *6*, 207‘229.

Shriberg, E. (2001). To "errrr" is human: Ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association*, *31*(1), 153–169.

Siddiqui, S. (2014). *Language, Gender and Power: The Politics of Representation and Hegemony in South Asia*. Oxford University Press.

Silverstein, M. (1979). Language structure and linguistic ideology. Chicago: Chicago Linguistic Society. In P. R. Clyne, W. Hanks, & C. L. Hofbauer (Eds.), *The elements: A parasession on linguistic units and levels* (pp. 193–247). Chicago Linguistic Society.

Singapore Department of Statistics. (2021). *Population Trends 2021*. https://www.singstat.gov.sg/publications/population/population-trends

Soh, Y. Q., Lee, J., & Tan, Y.-Y. (2022). Ethnicity and Tone Production on Singlish Particles. *Languages*, *7*(3), Article 3. https://doi.org/10.3390/languages7030243

Spence, J. T., & Buckner, C. (1995). Masculinity and Femininity: Defining the Undefinable. In *Gender, Power, and Communication in Human Relationships*. Routledge.

Spender, D. (1985). *Man Made Language*. Routledge & Kegan Paul.

Stadler, S. (2018). Conventionalized politeness in Singapore Colloquial English. *World Englishes*, *37*(2), 307–322. https://doi.org/10.1111/weng.12314

Stupka, R. (2011). Communication Accommodation in Mixed Gender Dyads. *Oshkosh Scholar*, *VI*, 64–78.

Swaab, R. I., Maddux, W. W., & Sinaceur, M. (2011). Early words that work: When and how virtual linguistic mimicry facilitates negotiation outcomes. *Journal of Experimental Social Psychology*, *47*(3), 616–621. https://doi.org/10.1016/j.jesp.2011.01.005

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel, *The Social Psychology of Intergroup Relations* (pp. 33–53). Wadsworth.

Takahashi, M. (2014). A comparative study of tag questions in four Asian Englishes from a corpus-based approach. *Asian Englishes*, *16*(2), 101–124. https://doi.org/10.1080/13488678.2014.908529

Talbot, M. (2019). *Language and Gender*. John Wiley & Sons.

Tan, G. S. T. (2017). *Tuning in to your partner's speech patterns and its impact on task completion – A Linguistic Perspective of the Communication Accommodation Theory (CAT)* [Undergraduate Thesis]. Nanyang Technological University.

Tan, S. J. W., Nurhilda, M., Lee, D. J., & Nurul, A. K. (2021). The Influence of Social Swearing on In-group Dynamics of Singaporean Youth. *Pioneer Road*, *1*, 141–148.

Tannen, D. (1990). *You Just Don't Understand*. William Morrow.

Tannen, D. (1993). *Gender and conversational interaction*. Oxford University Press.

Tannen, D. (2017). *You're the Only One I Can Tell: Inside the Language of Women's Friendships* (First Edition). Ballantine Books.

Teo, T. S. (2013). *Swearing in Singapore: Doing it the multilingual way* [Undergraduate Thesis]. Nanyang Technological University.

Thorne, B., & Henley, N. (1975). *Language and Sex: Difference and Dominance*. Newbury House Publishers, Inc.

Tottie, G. (2011). Uh and Um as sociolinguistic markers in British English. *International Journal of Corpus Linguistics*, *16*, 173–197. https://doi.org/10.1075/ijcl.16.2.02tot

Trudgill, P. (1974). *The Social Differentiation of English in Norwich*. Cambridge University Press.

Trudgill, P. (2000). *Sociolinguistics: An Introduction to Language and Society*. Penguin UK.

van Baaren, R. B., Holland, R. W., Steenaert, B., & van Knippenberg, A. (2003). Mimicry for money: Behavioral consequences of imitation. *Journal of Experimental Social Psychology*, *39*(4), 393–398. https://doi.org/10.1016/S0022-1031(03)00014-3

Van Den Berg, M., & Xu, D. (2010). *Urbanization and the Restructuring of Speech Communities in China and Europe*. Cambridge Scholars Publishing.

Waara, E., & Shaw, P. (2006). Male and Female Witnesses' Speech in Swedish Criminal Trials. *HERMES - Journal of Language and Communication in Business*, *36*, Article 36. https://doi.org/10.7146/hjlcb.v19i36.25842

Wamsley, J. C. (2019). Gendered Usage of Sentence-Final Particles in Mandarin Chinese. *IULC Working Papers*, *19*(1), Article 1. https://scholarworks.iu.edu/journals/index.php/iulcwp/article/view/27100

Wang, H.-C., & Fussell, S. (2010). Groups in groups: Conversational similarity in online multicultural multiparty brainstorming. *Proceedings of the 2010 ACM Conference on*

Computer Supported Cooperative Work, 351–360. https://doi.org/10.1145/1718918.1718980

Warren, P. (2016). *Uptalk: The Phenomenon of Rising Intonation*. Cambridge University Press.

Wee, L. (2004). Reduplication and discourse particles. In L. Lim (Ed.), *Singapore English: A grammatical description,* (pp. 105–126). John Benjamins Publishing Company.

West, C., & Zimmerman, D. H. (1987). Doing Gender. *Gender and Society*, *1*(2), 125–151.

Wiegman, R. (2012). *Wiegman, Robyn. 2012. Object Lessons. Durham and London: Duke University Press.* Duke University Press.

Winter, B., & Bürkner, P.-C. (2021). Poisson regression for linguists: A tutorial introduction to modelling count data with brms. *Language and Linguistics Compass*, *15*(11), e12439. https://doi.org/10.1111/lnc3.12439

Wong, J. (2014). *The Culture of Singapore English*. Cambridge University Press.

Xu, D. (2015, April 1). *Gender Differences in Chinese Speech Communities*. The Oxford Handbook of Chinese Linguistics. https://doi.org/10.1093/oxfordhb/9780199856336.013.0033

Yngve, V. (1970). On getting a word in edgewise. *Papers from the Sixth Regional Meeting [of the] Chicago Linguistic Society*, p.568.

Yoong, D. S. C. (2010). Standard English and Singlish: The Clash of Language Values in Contemporary Singapore. In Y. Treis & R. De Busser (Eds.), *Selected Papers from the 2009 Conference of the Australian Linguistic Society*.

Yuan, J., Xu, X., Lai, W., & Liberman, M. (2016). Pauses and pause fillers in Mandarin monologue speech: The effects of sex and proficiency. *Speech Prosody 2016*, 1167–1170. https://doi.org/10.21437/SpeechProsody.2016-240

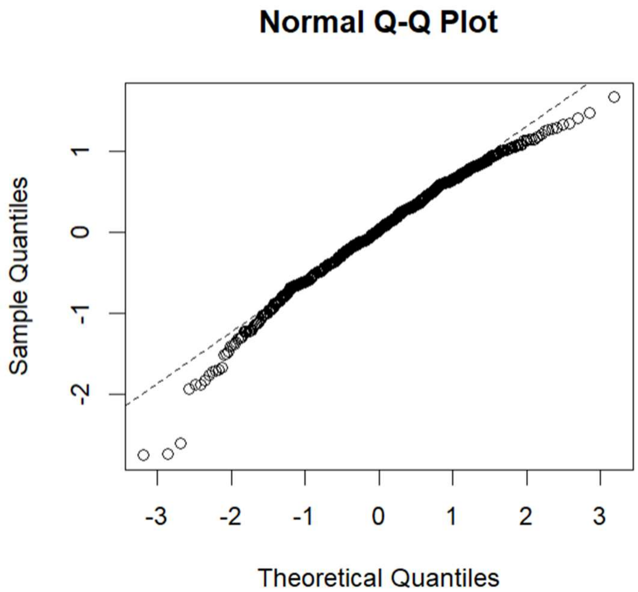
Zhang, A. (1995). *Another look at the sajiao phenomenon* [Master's Thesis]. Ohio State University.

Zhang, L., & Zhuang, Z. (2020, December). Gender difference in the occurrence of Cantonese sentence-final particles across lifespan. *Proceedings of the International Conference on Gender, Language and Education (ICGLE)*. International Conference on Gender, Language and Education (ICGLE), Hong Kong, China. https://repository.eduhk.hk/en/publications/gender-difference-in-the-occurrence-of-cantonese-sentence-final-p

Appendix A: All Tables and Figures for Linear Modelling Results extracted from R

Backchannelling: Speaker Gender, Partner Gender, Dyad Type

Quantile-Quantile plots to check for normality assumptions



ANOVA result table for backchannels per million words

```
Type III Analysis of Variance Table with Satterthwaite's method
                      Sum Sq Mean Sq NumDF  DenDF F value  Pr(>F)  
Gender                3.9427  3.9427     1 686.43  6.6506 0.01012 *
Partner.Gender        0.0405  0.0405     1 686.43  0.0683 0.79384  
Gender:Partner.Gender 0.4620  0.4620     1 344.00  0.7793 0.37798  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table of Estimated Marginal Means (EMMs) values

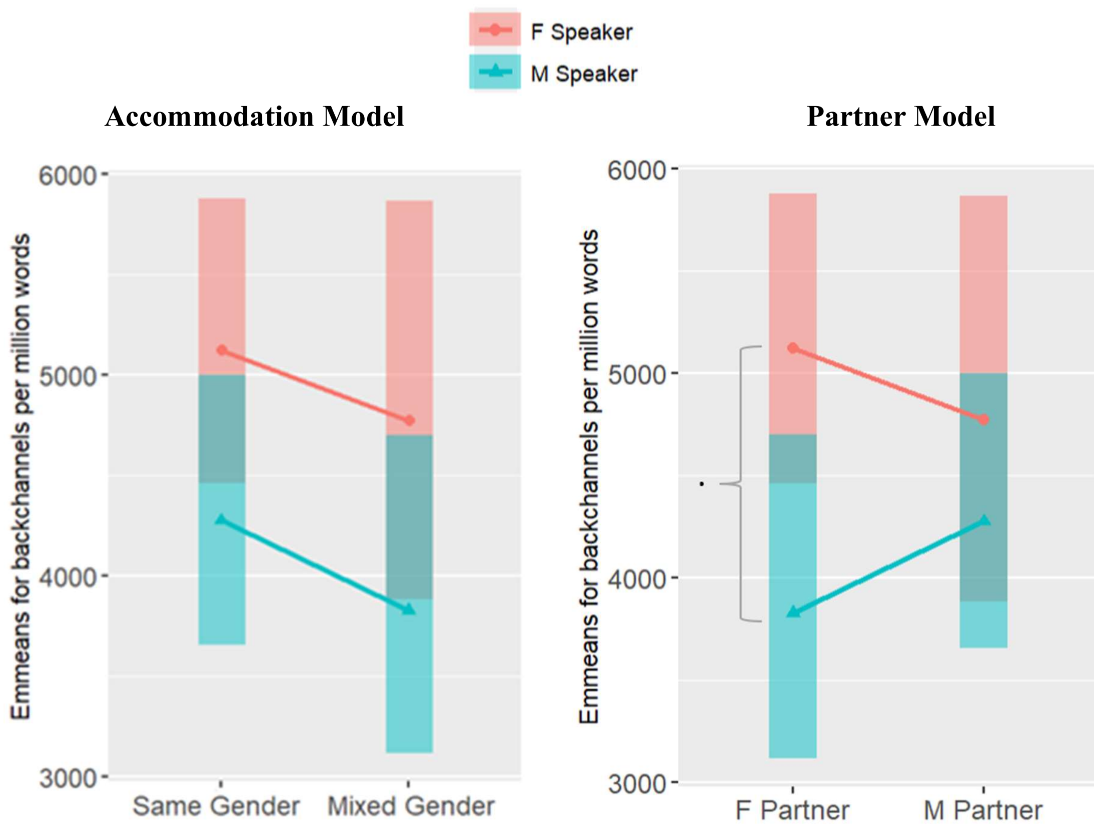
```
 Gender    Partner.Gender response  SE  df lower.CL upper.CL
 F Speaker F Partner          5121 359 344     4462     5877
 M Speaker F Partner          3826 402 590     3113     4702
 F Speaker M Partner          4773 501 590     3883     5865
 M Speaker M Partner          4275 340 344     3657     4998

Degrees-of-freedom method: kenward-roger 
Confidence level used: 0.95 
Intervals are back-transformed from the log scale
```

Table of pairwise contrasts

contrast	estimate	SE	df	t.ratio	p.value
F Speaker F Partner - M Speaker F Partner	1295	538	344	2.405	0.0781
F Speaker F Partner - F Speaker M Partner	348	616	344	0.565	0.9424
F Speaker F Partner - M Speaker M Partner	845	494	344	1.712	0.3192
M Speaker F Partner - F Speaker M Partner	-947	498	590	-1.902	0.2283
M Speaker F Partner - M Speaker M Partner	-449	526	344	-0.855	0.8280
F Speaker M Partner - M Speaker M Partner	497	605	344	0.822	0.8442

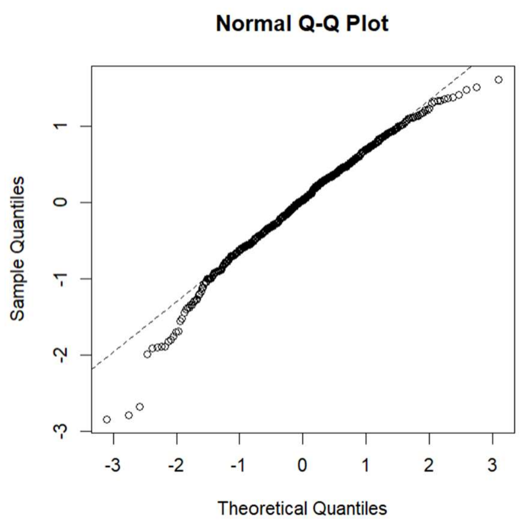
Degrees-of-freedom method: inherited from kenward-roger when re-gridding
P value adjustment: tukey method for comparing a family of 4 estimates

Plots of EMMs for Accommodation Model and Partner Model



Backchannelling: Ethnicity

Quantile-Quantile plots to check for normality assumptions



ANOVA result table for backchannels per million words

```
Type III Analysis of Variance Table with Satterthwaite's method
               Sum Sq Mean Sq NumDF  DenDF F value Pr(>F)
Pair.Ethnicity 1.0115 0.50576     2 259.78  0.7559 0.4706
```

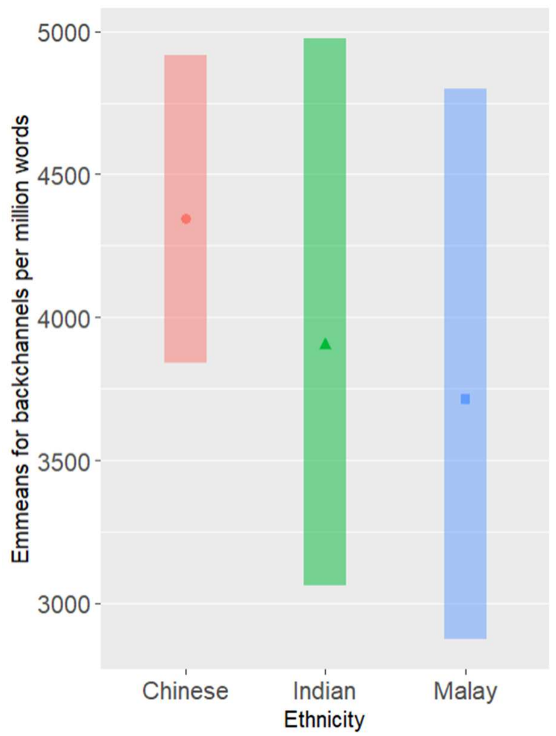
Table of Estimated Marginal Means (EMMs) values

```
 Pair.Ethnicity response  SE  df lower.CL upper.CL
 Chinese            4345 273 275     3839     4918
 Indian             3903 481 267     3062     4974
 Malay              3713 483 258     2874     4797

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95
Intervals are back-transformed from the log scale
```

Table of pairwise contrasts

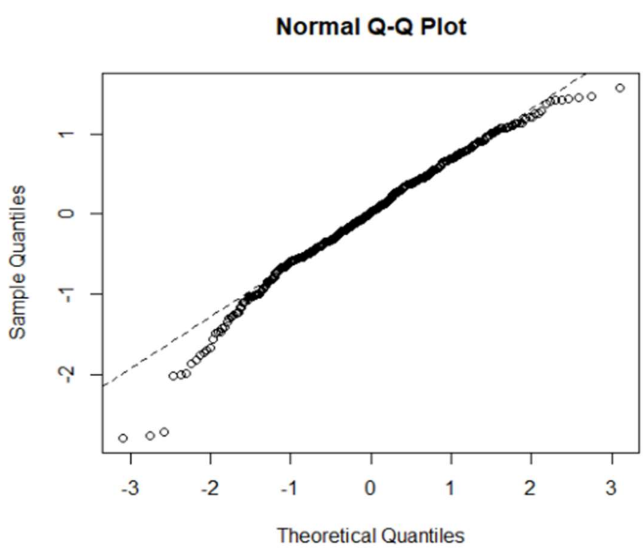
```
 contrast         estimate  SE  df t.ratio p.value
 Chinese - Indian      443 553 267   0.800  0.7032
 Chinese - Malay       632 555 258   1.139  0.4911
 Indian - Malay        189 682 258   0.278  0.9584

Degrees-of-freedom method: inherited from kenward-roger when re-gridding
P value adjustment: tukey method for comparing a family of 3 estimates
```

Plots of EMMs for the three ethnicity pairings



Backchannelling: Ethnicity and Speaker Gender

Quantile-Quantile plots to check for normality assumptions



ANOVA result table for backchannels per million words

```
Type III Analysis of Variance Table with Satterthwaite's method
                     Sum Sq Mean Sq NumDF  DenDF F value  Pr(>F)  
Pair.Ethnicity      0.20810 0.10405     2 283.27  0.1561 0.85557  
Gender              3.03921 3.03921     1 502.62  4.5588 0.03323 *
Pair.Ethnicity:Gender 0.13391 0.06696   2 503.14  0.1004 0.90446  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table of Estimated Marginal Means (EMMs) values

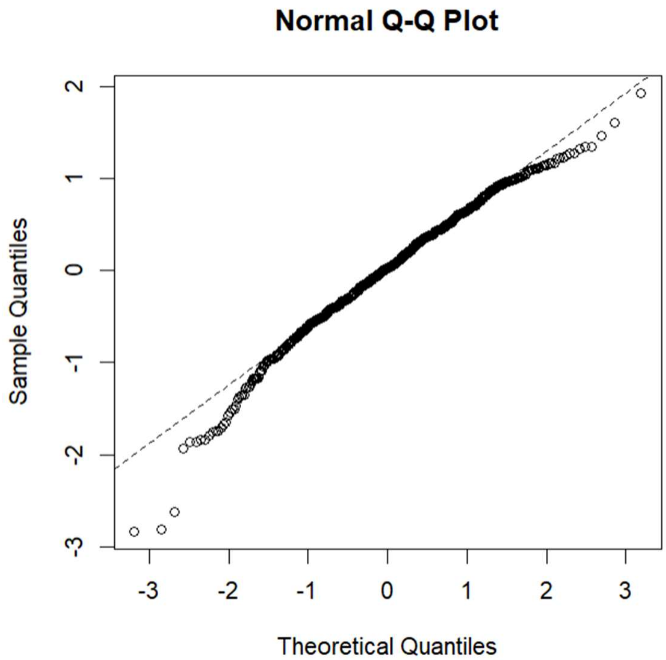
```
 Pair.Ethnicity Gender    response  SE  df lower.CL upper.CL
 Chinese        F Speaker     4657 340 338     4034     5377
 Indian         F Speaker     4869 960 449     3305     7174
 Malay          F Speaker     4254 877 410     2837     6379
 Chinese        M Speaker     3736 385 447     3052     4575
 Indian         M Speaker     3499 505 329     2634     4648
 Malay          M Speaker     3459 533 322     2554     4684

Degrees-of-freedom method: kenward-roger 
Confidence level used: 0.95 
Intervals are back-transformed from the log scale
```

Table of pairwise contrasts

```
 contrast                              estimate   SE  df t.ratio p.value
 Chinese F Speaker - Indian F Speaker    -212.0 1019 338  -0.208  0.9999
 Chinese F Speaker - Malay F Speaker      402.9  940 338   0.428  0.9982
 Chinese F Speaker - Chinese M Speaker    920.5  484 338   1.901  0.4034
 Chinese F Speaker - Indian M Speaker    1158.1  609 329   1.901  0.4031
 Chinese F Speaker - Malay M Speaker     1197.9  632 322   1.894  0.4077
 Indian F Speaker - Malay F Speaker       614.9 1300 410   0.473  0.9970
 Indian F Speaker - Chinese M Speaker    1132.5 1034 447   1.095  0.8834
 Indian F Speaker - Indian M Speaker     1370.1 1032 329   1.328  0.7693
 Indian F Speaker - Malay M Speaker      1409.9 1098 322   1.284  0.7938
 Malay F Speaker - Chinese M Speaker      517.6  957 410   0.541  0.9945
 Malay F Speaker - Indian M Speaker       755.2 1012 329   0.746  0.9759
 Malay F Speaker - Malay M Speaker        795.0  977 322   0.814  0.9648
 Chinese M Speaker - Indian M Speaker     237.6  635 329   0.374  0.9990
 Chinese M Speaker - Malay M Speaker      277.4  658 322   0.422  0.9983
 Indian M Speaker - Malay M Speaker        39.8  735 322   0.054  1.0000

Degrees-of-freedom method: inherited from kenward-roger when re-gridding 
P value adjustment: tukey method for comparing a family of 6 estimates
```

Backchannelling: Mixed Ethnicity vs. Same Ethnicity

Quantile-Quantile plots to check for normality assumptions



ANOVA result table for backchannels per million words

```
Type III Analysis of Variance Table with Satterthwaite's method
               Sum Sq Mean Sq NumDF  DenDF F value     Pr(>F)    
Pair.Ethnicity 10.126  10.126     1 583.13  16.518 0.00005478 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

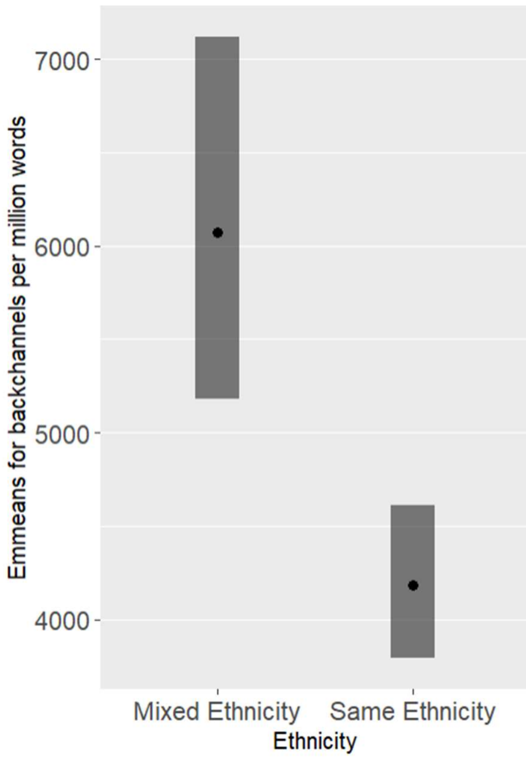
Table of Estimated Marginal Means (EMMs) values

```
$emmeans
 Pair.Ethnicity  response  SE  df lower.CL upper.CL
 Mixed Ethnicity     6072 491 511     5179     7118
 Same Ethnicity      4183 208 393     3792     4613

Degrees-of-freedom method: kenward-roger 
Confidence level used: 0.95 
Intervals are back-transformed from the log scale 
```

Table of pairwise contrasts

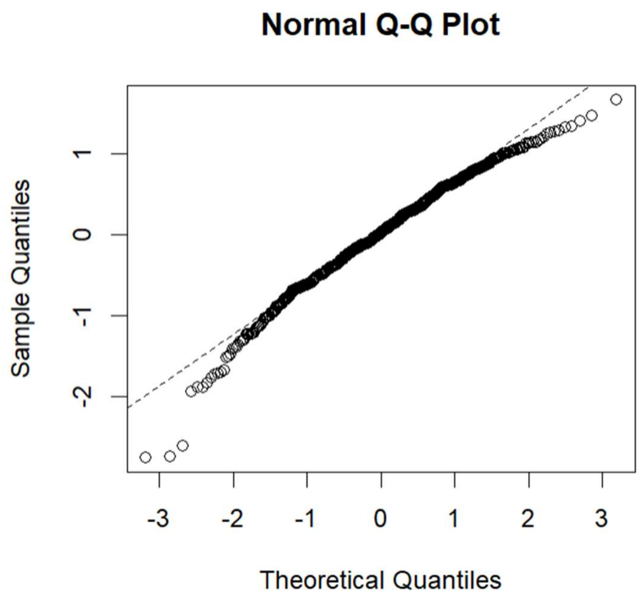
```
 contrast                          estimate  SE  df t.ratio p.value
 Mixed Ethnicity - Same Ethnicity     1889 520 393   3.635  0.0003

Degrees-of-freedom method: inherited from kenward-roger when re-gridding 
```

Plots of EMMs for the three ethnicity pairings



Backchannelling: Mixed Ethnicity vs. Same Ethnicity and Speaker Gender

Quantile-Quantile plots to check for normality assumptions



ANOVA result table for backchannels per million words

```
Type III Analysis of Variance Table with Satterthwaite's method
                       Sum Sq Mean Sq NumDF  DenDF F value     Pr(>F)
Pair.Ethnicity        11.2470 11.2470     1 580.42 18.4808 0.00002013 ***
Gender                 3.3793  3.3793     1 677.72  5.5528    0.01873 *
Pair.Ethnicity:Gender  0.1290  0.1290     1 685.59  0.2120    0.64538
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table of Estimated Marginal Means (EMMs) values

```
 Pair.Ethnicity  Gender    response  SE  df lower.CL upper.CL
 Mixed Ethnicity F Speaker     6619 733 624     5325     8227
 Same Ethnicity  F Speaker     4648 291 519     4110     5256
 Mixed Ethnicity M Speaker     5609 610 587     4530     6946
 Same Ethnicity  M Speaker     3638 258 557     3165     4182

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95
Intervals are back-transformed from the log scale
```

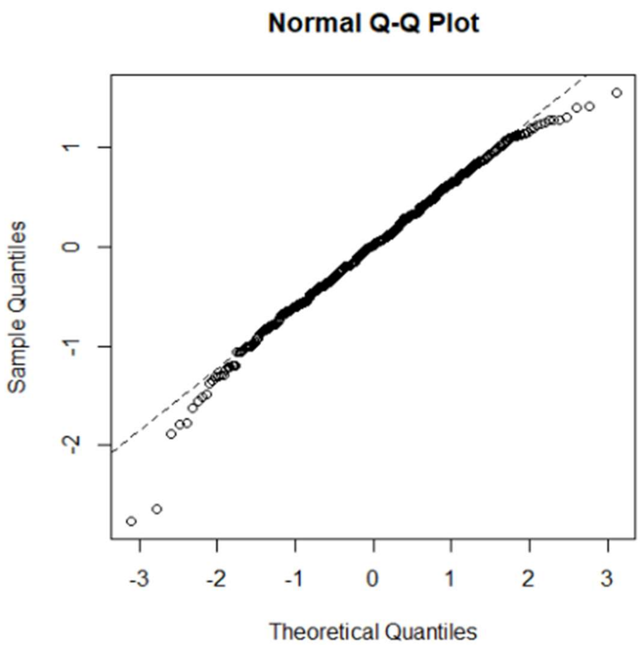
Table of pairwise contrasts

contrast	estimate	SE	df	t.ratio	p.value
Mixed Ethnicity F Speaker - Same Ethnicity F Speaker	1971	774	519	2.546	0.0543
Mixed Ethnicity F Speaker - Mixed Ethnicity M Speaker	1010	918	587	1.100	0.6899
Mixed Ethnicity F Speaker - Same Ethnicity M Speaker	2981	769	557	3.874	0.0007
Same Ethnicity F Speaker - Mixed Ethnicity M Speaker	-961	667	519	-1.442	0.4737
Same Ethnicity F Speaker - Same Ethnicity M Speaker	1010	366	519	2.760	0.0304
Mixed Ethnicity M Speaker - Same Ethnicity M Speaker	1971	654	557	3.013	0.0144

Degrees-of-freedom method: inherited from kenward-roger when re-gridding
P value adjustment: tukey method for comparing a family of 4 estimates

Backchannelling: Age Groups

Quantile-Quantile plots to check for normality assumptions



ANOVA result table for backchannels per million words

```
Type III Analysis of Variance Table with Satterthwaite's method
            Sum Sq Mean Sq NumDF DenDF F value   Pr(>F)
Age.Pairing 6.6832  2.2277     3   261  4.0111 0.008151 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

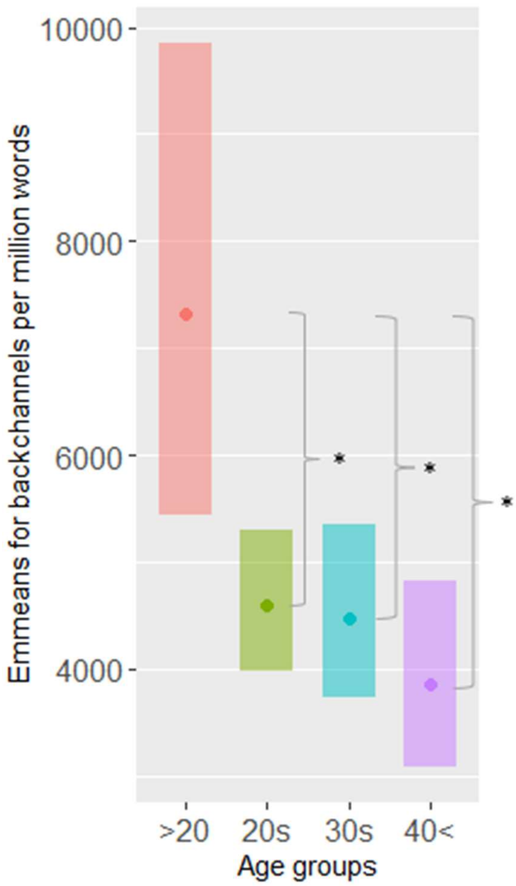
Table of Estimated Marginal Means (EMMs) values

```
 Age.Pairing response   SE  df lower.CL upper.CL
 >20            7320 1103 261     5440     9850
 20s            4588  332 261     3978     5291
 30s            4463  409 261     3726     5346
 40<            3856  436 261     3086     4817

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95
Intervals are back-transformed from the log scale
```

Table of pairwise contrasts

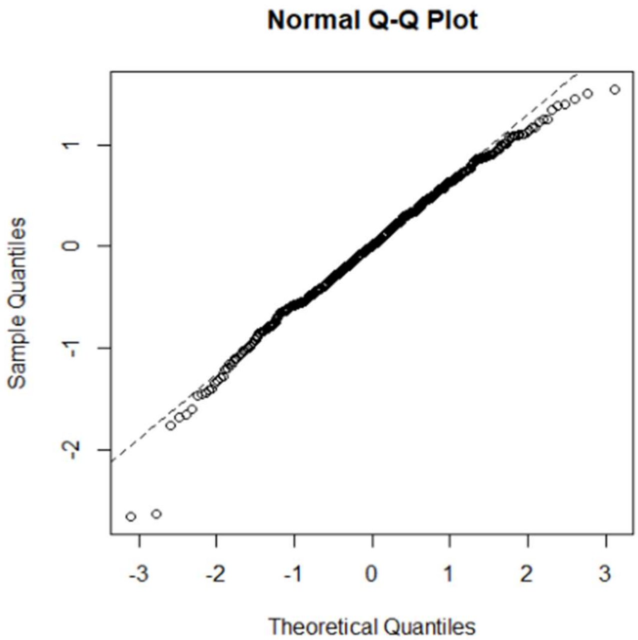
```
 contrast  estimate   SE  df t.ratio p.value
 >20 - 20s     2733 1152 261   2.372  0.0852
 >20 - 30s     2857 1177 261   2.428  0.0744
 >20 - 40<     3465 1186 261   2.920  0.0198
 20s - 30s      124  527 261   0.236  0.9954
 20s - 40<      732  548 261   1.335  0.5412
 30s - 40<      608  598 261   1.016  0.7401

Degrees-of-freedom method: inherited from kenward-roger when re-gridding
P value adjustment: tukey method for comparing a family of 4 estimates
```

Plots of EMMs for the age groupings



Backchannelling: Age Groups and Speaker Gender

Quantile-Quantile plots to check for normality assumptions



ANOVA result table for backchannels per million words

Type III Analysis of Variance Table with Satterthwaite's method

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)	
Age.Pairing	6.6003	2.2001	3	263.36	4.0581	0.007649	**
Gender	3.8537	3.8537	1	504.65	7.1082	0.007919	**
Age.Pairing:Gender	2.9589	0.9863	3	506.46	1.8192	0.142646	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Table of Estimated Marginal Means (EMMs) values

Age.Pairing	Gender	response	SE	df	lower.CL	upper.CL
>20	F Speaker	7854	1723	389	5102	12089
20s	F Speaker	4644	402	336	3917	5506
30s	F Speaker	5483	647	378	4347	6915
40<	F Speaker	4912	755	361	3630	6647
>20	M Speaker	6975	1297	340	4838	10056
20s	M Speaker	4492	492	412	3621	5572
30s	M Speaker	3570	436	392	2807	4539
40<	M Speaker	3086	457	351	2306	4130

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95
Intervals are back-transformed from the log scale

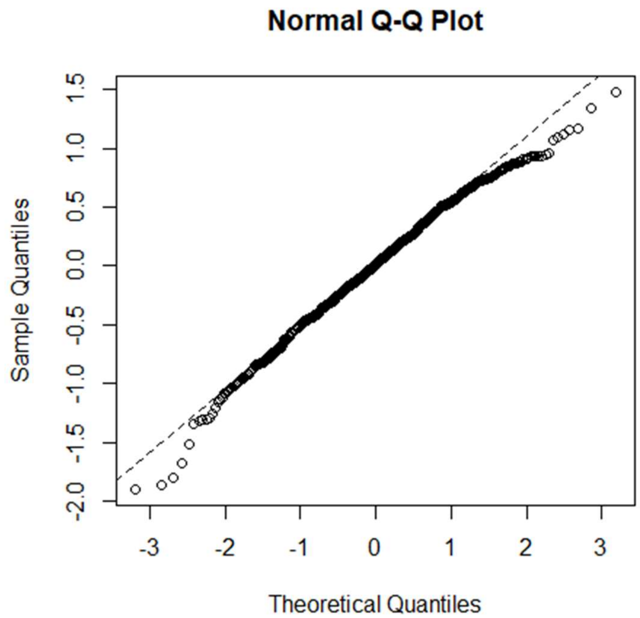
Table of pairwise contrasts

contrast	estimate	SE	df	t.ratio	p.value
>20 F Speaker - 20s F Speaker	3210	1769	336	1.814	0.6107
>20 F Speaker - 30s F Speaker	2371	1840	378	1.288	0.9028
>20 F Speaker - 40< F Speaker	2942	1881	361	1.564	0.7716
>20 F Speaker - >20 M Speaker	879	2030	340	0.433	0.9999
>20 F Speaker - 20s M Speaker	3362	1792	389	1.876	0.5682
>20 F Speaker - 30s M Speaker	4284	1777	389	2.410	0.2391
>20 F Speaker - 40< M Speaker	4768	1782	351	2.675	0.1338
20s F Speaker - 30s F Speaker	-839	762	336	-1.101	0.9564
20s F Speaker - 40< F Speaker	-268	856	336	-0.313	1.0000
20s F Speaker - >20 M Speaker	-2331	1358	336	-1.716	0.6767
20s F Speaker - 20s M Speaker	152	594	336	0.256	1.0000
20s F Speaker - 30s M Speaker	1074	593	336	1.811	0.6133
20s F Speaker - 40< M Speaker	1558	609	336	2.560	0.1747
30s F Speaker - 40< F Speaker	571	995	361	0.574	0.9992
30s F Speaker - >20 M Speaker	-1492	1450	340	-1.029	0.9697
30s F Speaker - 20s M Speaker	991	813	378	1.218	0.9262
30s F Speaker - 30s M Speaker	1913	723	378	2.647	0.1428
30s F Speaker - 40< M Speaker	2397	792	351	3.025	0.0536
40< F Speaker - >20 M Speaker	-2063	1501	340	-1.374	0.8685
40< F Speaker - 20s M Speaker	420	902	361	0.466	0.9998
40< F Speaker - 30s M Speaker	1342	872	361	1.539	0.7860
40< F Speaker - 40< M Speaker	1826	838	351	2.178	0.3675
>20 M Speaker - 20s M Speaker	2483	1388	340	1.789	0.6278
>20 M Speaker - 30s M Speaker	3405	1369	340	2.488	0.2043
>20 M Speaker - 40< M Speaker	3889	1376	340	2.827	0.0917
20s M Speaker - 30s M Speaker	922	658	392	1.402	0.8561
20s M Speaker - 40< M Speaker	1406	672	351	2.093	0.4213
30s M Speaker - 40< M Speaker	484	632	351	0.766	0.9947

Degrees-of-freedom method: inherited from kenward-roger when re-gridding
P value adjustment: tukey method for comparing a family of 8 estimates

Backchannelling (with *okay*): Speaker Gender, Partner Gender, Dyad Type

Quantile-Quantile plots to check for normality assumptions



ANOVA result table for backchannels per million words

Type III Analysis of Variance Table with Satterthwaite's method

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Gender	1.03689	1.03689	1	689.74	2.5750	0.1090
Partner.Gender	0.09048	0.09048	1	689.74	0.2247	0.6356
Gender:Partner.Gender	0.00092	0.00092	1	344.00	0.0023	0.9619

Table of Estimated Marginal Means (EMMs) values

Gender	Partner.Gender	response	SE	df	lower.CL	upper.CL
F Speaker	F Partner	7756	472	344	6881	8742
M Speaker	F Partner	6949	624	572	5825	8290
F Speaker	M Partner	7968	716	572	6679	9505
M Speaker	M Partner	7200	497	344	6287	8247

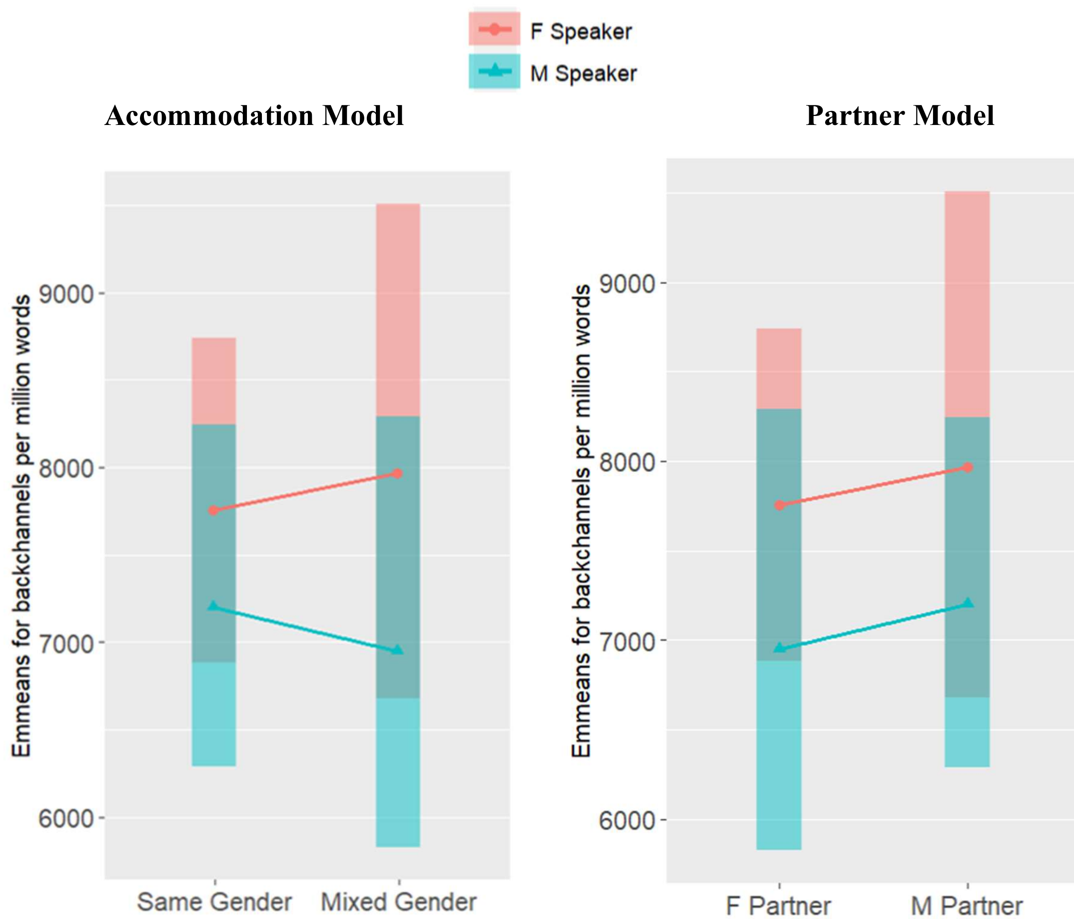
Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95
Intervals are back-transformed from the log scale

Table of pairwise contrasts

contrast	estimate	SE	df	t.ratio	p.value
F Speaker F Partner - M Speaker F Partner	807	783	344	1.031	0.7314
F Speaker F Partner - F Speaker M Partner	-212	857	344	-0.247	0.9947
F Speaker F Partner - M Speaker M Partner	556	685	344	0.811	0.8490
M Speaker F Partner - F Speaker M Partner	-1018	706	572	-1.442	0.4733
M Speaker F Partner - M Speaker M Partner	-251	798	344	-0.314	0.9892
F Speaker M Partner - M Speaker M Partner	768	871	344	0.881	0.8147

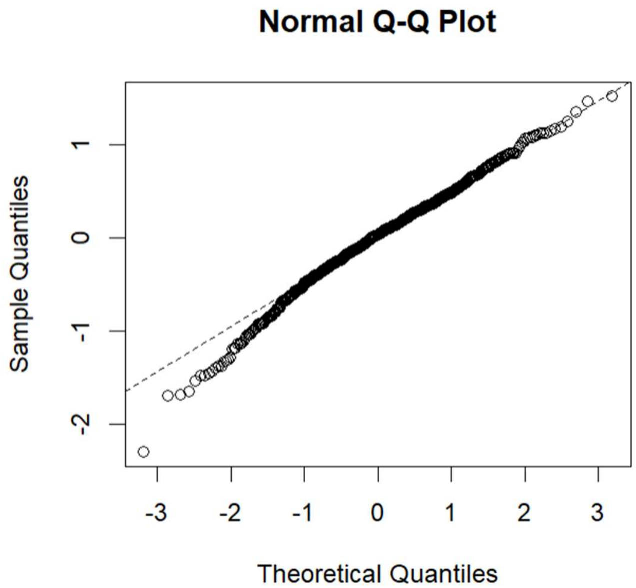
Degrees-of-freedom method: inherited from kenward-roger when re-gridding
P value adjustment: tukey method for comparing a family of 4 estimates

Plots of EMMs for Accommodation Model and Partner Model



Filled Pauses: Speaker Gender, Partner Gender, Dyad Type

Quantile-Quantile plots to check for normality assumptions



ANOVA result table for Filled Pauses per million words

```
Type III Analysis of Variance Table with Satterthwaite's method
                       Sum Sq Mean Sq NumDF  DenDF F value  Pr(>F)
Gender                1.49384 1.49384     1 684.47  3.7189 0.05421 .
Partner.Gender        1.42159 1.42159     1 684.47  3.5391 0.06036 .
Gender:Partner.Gender 0.00464 0.00464     1 344.00  0.0116 0.91444
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table of Estimated Marginal Means (EMMs) values

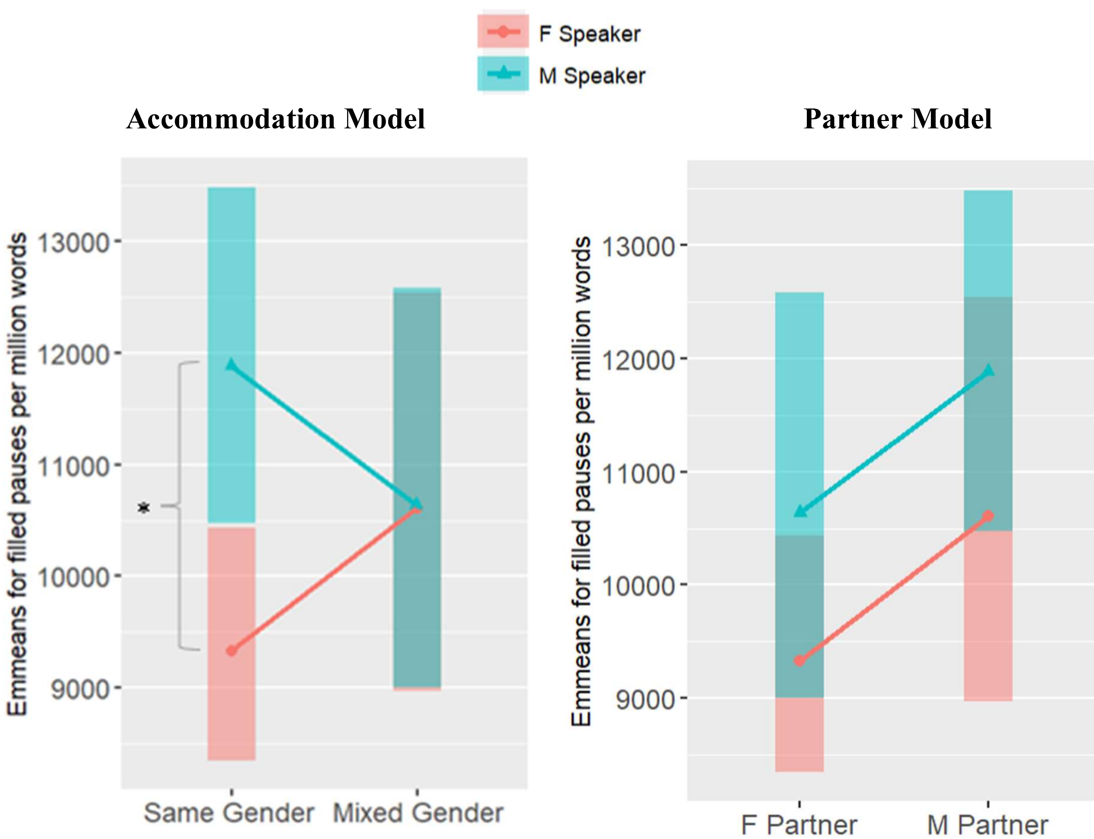
```
 Gender    Partner.Gender response  SE  df lower.CL upper.CL
 F Speaker F Partner          9330 528 344     8346    10429
 M Speaker F Partner         10641 908 596     8999    12582
 F Speaker M Partner         10609 905 596     8972    12545
 M Speaker M Partner         11886 763 344    10476    13486

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95
Intervals are back-transformed from the log scale
```

Table of pairwise contrasts

contrast	estimate	SE	df	t.ratio	p.value
F Speaker F Partner - M Speaker F Partner	-1311.2	1050	344	-1.248	0.5966
F Speaker F Partner - F Speaker M Partner	-1279.3	1048	344	-1.221	0.6142
F Speaker F Partner - M Speaker M Partner	-2555.9	928	344	-2.754	0.0314
M Speaker F Partner - F Speaker M Partner	31.9	998	596	0.032	1.0000
M Speaker F Partner - M Speaker M Partner	-1244.8	1186	344	-1.049	0.7204
F Speaker M Partner - M Speaker M Partner	-1276.6	1184	344	-1.078	0.7031

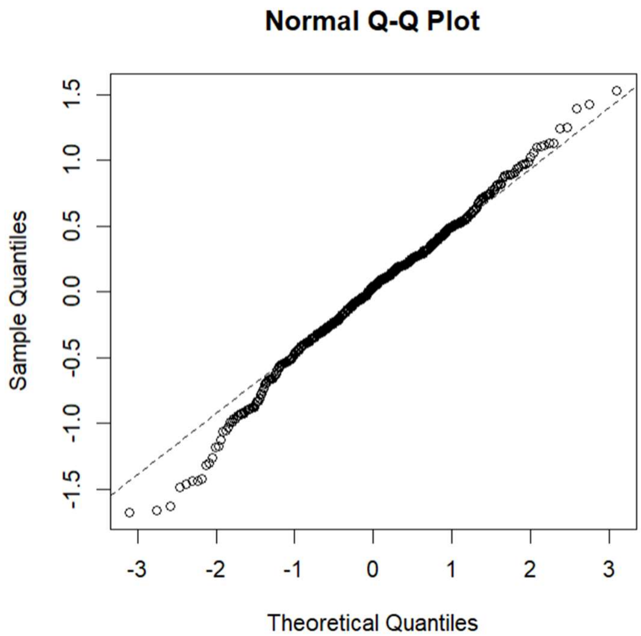
```
Degrees-of-freedom method: inherited from kenward-roger when re-gridding
P value adjustment: tukey method for comparing a family of 4 estimates
```

Plots of EMMs for Accommodation Model and Partner Model



Filled Pauses: Ethnicity

Quantile-Quantile plots to check for normality assumptions



ANOVA result table for Filled Pauses per million words

```
Type III Analysis of Variance Table with Satterthwaite's method
               Sum Sq Mean Sq NumDF  DenDF F value  Pr(>F)
Pair.Ethnicity 2.6535  1.3268     2 267.83  3.3785 0.03556 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

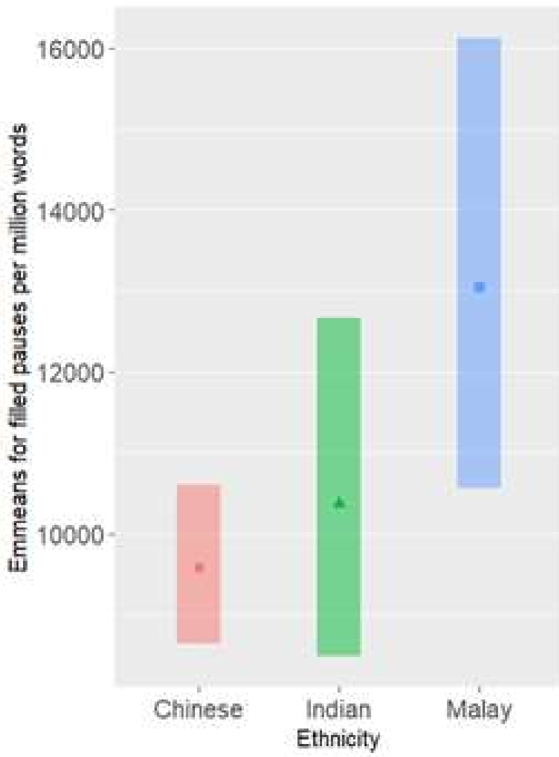
Table of Estimated Marginal Means (EMMs) values

```
 Pair.Ethnicity response   SE  df lower.CL upper.CL
 Chinese            9579  496 275     8650    10607
 Indian            10365 1053 268     8485    12660
 Malay             13048 1401 259    10561    16120

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95
Intervals are back-transformed from the log scale
```

Table of pairwise contrasts

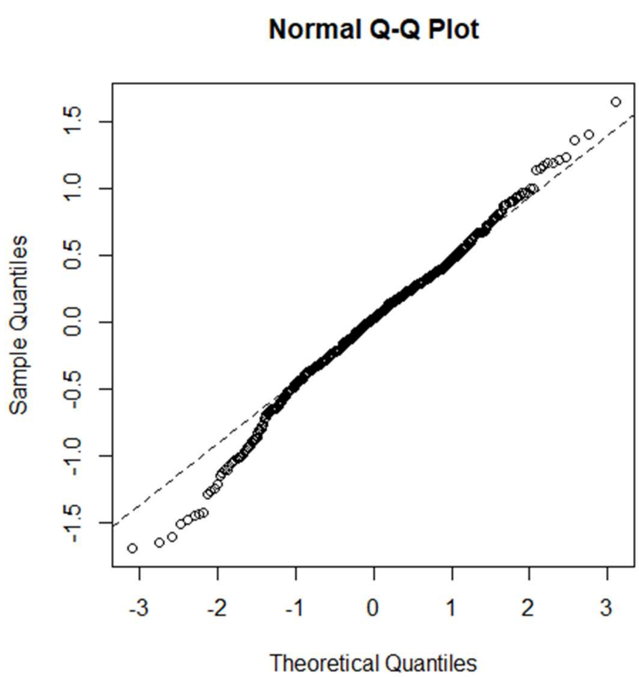
```
 contrast         estimate   SE  df t.ratio p.value
 Chinese - Indian     -786 1164 268  -0.675  0.7781
 Chinese - Malay     -3469 1486 259  -2.334  0.0529
 Indian - Malay      -2683 1753 259  -1.531  0.2782

Degrees-of-freedom method: inherited from kenward-roger when re-gridding
P value adjustment: tukey method for comparing a family of 3 estimates
```

Plots of EMMs for the three ethnicity pairings



Filled Pauses: Ethnicity and Speaker Gender

Quantile-Quantile plots to check for normality assumptions



ANOVA result table for Filled Pauses per million words

```
Type III Analysis of Variance Table with Satterthwaite's method
                      Sum Sq Mean Sq NumDF  DenDF F value  Pr(>F)
Pair.Ethnicity       1.94038 0.97019     2 288.66  2.4679 0.08655 .
Gender               1.02001 1.02001     1 507.29  2.5946 0.10785
Pair.Ethnicity:Gender 0.69127 0.34563     2 507.46  0.8792 0.41575
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table of Estimated Marginal Means (EMMs) values

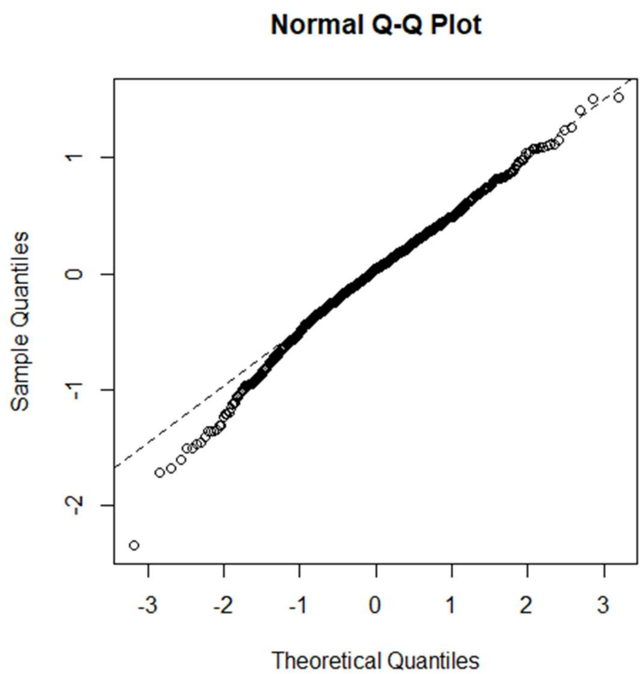
```
 Pair.Ethnicity Gender    response   SE  df lower.CL upper.CL
 Chinese        F Speaker     8831  524 341     7858     9924
 Indian         F Speaker     8910 1411 455     6526    12163
 Malay          F Speaker    13287 2208 420     9585    18419
 Chinese        M Speaker    11415  944 454     9702    13430
 Indian         M Speaker    11173 1311 333     8870    14075
 Malay          M Speaker    12919 1620 327    10095    16533

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95
Intervals are back-transformed from the log scale
```

Table of pairwise contrasts

```
 contrast                               estimate   SE  df t.ratio p.value
 Chinese F Speaker - Indian F Speaker      -78.7 1505 341  -0.052  1.0000
 Chinese F Speaker - Malay F Speaker     -4456.1 2269 341  -1.964  0.3654
 Chinese F Speaker - Chinese M Speaker   -2583.7 1015 341  -2.546  0.1140
 Chinese F Speaker - Indian M Speaker    -2342.5 1412 333  -1.659  0.5602
 Chinese F Speaker - Malay M Speaker     -4088.1 1702 327  -2.401  0.1588
 Indian F Speaker - Malay F Speaker      -4377.3 2620 420  -1.671  0.5520
 Indian F Speaker - Chinese M Speaker    -2505.0 1698 454  -1.475  0.6803
 Indian F Speaker - Indian M Speaker     -2263.8 1784 333  -1.269  0.8016
 Indian F Speaker - Malay M Speaker      -4009.4 2148 327  -1.866  0.4249
 Malay F Speaker - Chinese M Speaker      1872.4 2401 420   0.780  0.9708
 Malay F Speaker - Indian M Speaker       2113.6 2568 333   0.823  0.9631
 Malay F Speaker - Malay M Speaker         368.0 2561 327   0.144  1.0000
 Chinese M Speaker - Indian M Speaker      241.2 1616 333   0.149  1.0000
 Chinese M Speaker - Malay M Speaker     -1504.4 1875 327  -0.802  0.9670
 Indian M Speaker - Malay M Speaker      -1745.6 2084 327  -0.838  0.9603

Degrees-of-freedom method: inherited from kenward-roger when re-gridding
P value adjustment: tukey method for comparing a family of 6 estimates
```

Filled Pauses: Mixed Ethnicity vs. Same Ethnicity

Quantile-Quantile plots to check for normality assumptions



ANOVA result table for Filled Pauses per million words

```
Type III Analysis of Variance Table with Satterthwaite's method
                 Sum Sq Mean Sq NumDF  DenDF F value Pr(>F)
Pair.Ethnicity 0.18208 0.18208     1 596.27  0.4519 0.5017
```

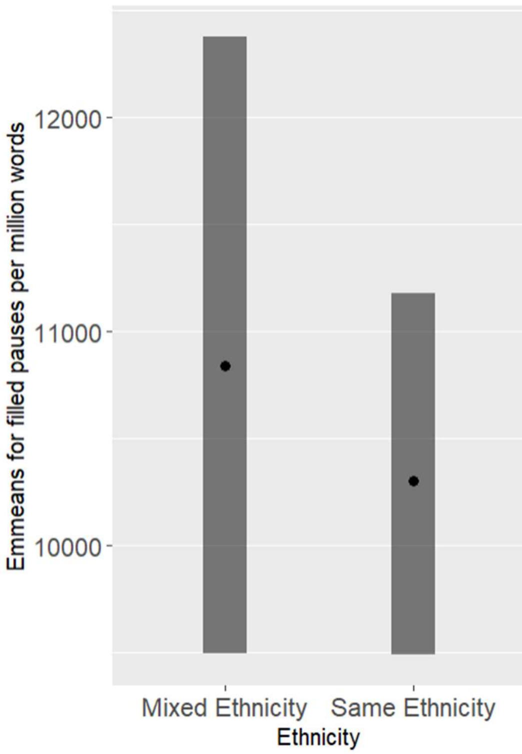
Table of Estimated Marginal Means (EMMs) values

```
 Pair.Ethnicity  response  SE  df lower.CL upper.CL
 Mixed Ethnicity    10842 731 519     9496    12378
 Same Ethnicity     10300 430 395     9489    11181

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95
Intervals are back-transformed from the log scale
```

Table of pairwise contrasts

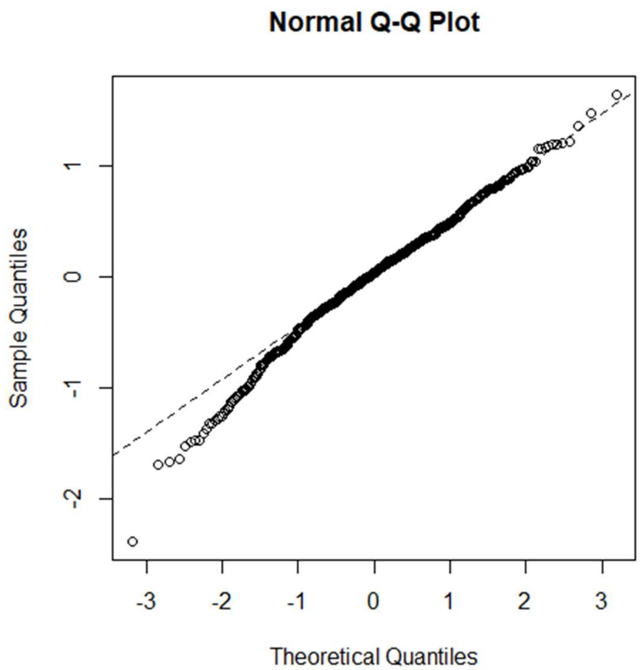
```
 contrast                          estimate  SE  df t.ratio p.value
 Mixed Ethnicity - Same Ethnicity       541 818 395   0.662  0.5083

Degrees-of-freedom method: inherited from kenward-roger when re-gridding
```

Plots of EMMs for the three ethnicity pairings



Filled Pauses: Mixed Ethnicity vs. Same Ethnicity and Speaker Gender

Quantile-Quantile plots to check for normality assumptions



ANOVA result table for Filled Pauses per million words

```
Type III Analysis of Variance Table with Satterthwaite's method
                     Sum Sq Mean Sq NumDF  DenDF F value  Pr(>F)
Pair.Ethnicity      0.09096 0.09096     1 591.05  0.2271 0.63386
Gender              0.16523 0.16523     1 682.66  0.4125 0.52090
Pair.Ethnicity:Gender 2.51858 2.51858   1 688.35  6.2883 0.01238 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table of Estimated Marginal Means (EMMs) values

```
 Pair.Ethnicity  Gender    response   SE  df lower.CL upper.CL
 Mixed Ethnicity F Speaker    11591 1063 632     9681    13878
 Same Ethnicity  F Speaker     9353  486 523     8444    10359
 Mixed Ethnicity M Speaker    10155  916 595     8506    12125
 Same Ethnicity  M Speaker    11708  689 563    10430    13142

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95
Intervals are back-transformed from the log scale
```

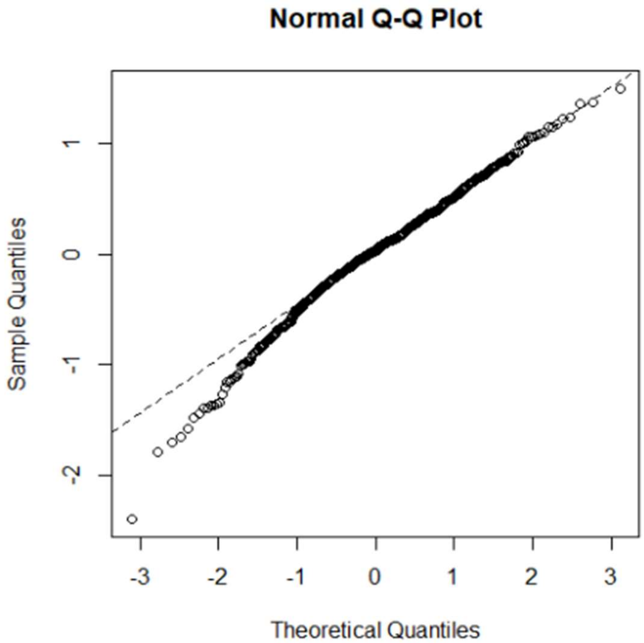
Table of pairwise contrasts

contrast	estimate	SE	df	t.ratio	p.value
Mixed Ethnicity F Speaker - Same Ethnicity F Speaker	2238	1143	523	1.959	0.2052
Mixed Ethnicity F Speaker - Mixed Ethnicity M Speaker	1436	1345	595	1.067	0.7095
Mixed Ethnicity F Speaker - Same Ethnicity M Speaker	-117	1246	563	-0.094	0.9997
Same Ethnicity F Speaker - Mixed Ethnicity M Speaker	-803	1020	523	-0.787	0.8605
Same Ethnicity F Speaker - Same Ethnicity M Speaker	-2355	791	523	-2.976	0.0161
Mixed Ethnicity M Speaker - Same Ethnicity M Speaker	-1552	1125	563	-1.380	0.5122

Degrees-of-freedom method: inherited from kenward-roger when re-gridding
P value adjustment: tukey method for comparing a family of 4 estimates

Filled Pauses: Age Groups

Quantile-Quantile plots to check for normality assumptions



ANOVA result table for Filled Pauses per million words

```
Type III Analysis of Variance Table with Satterthwaite's method
            Sum Sq Mean Sq NumDF DenDF F value  Pr(>F)
Age.Pairing 3.5874  1.1958     3   261  2.8444 0.03821 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table of Estimated Marginal Means (EMMs) values

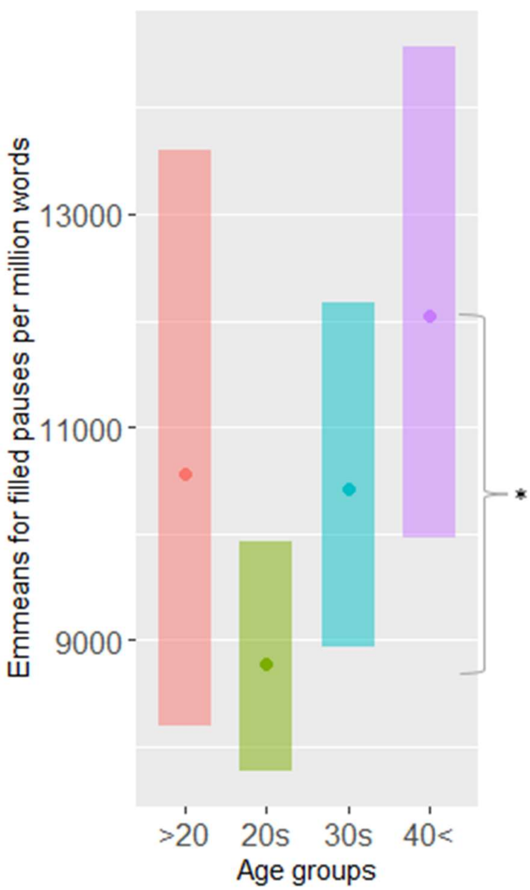
```
 Age.Pairing response   SE  df lower.CL upper.CL
 >20            10553 1361 261     8187    13603
 20s             8781  544 261     7773     9920
 30s            10424  817 261     8933    12164
 40<            12045 1165 261     9956    14571

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95
Intervals are back-transformed from the log scale
```

Table of pairwise contrasts

contrast	estimate	SE	df	t.ratio	p.value
>20 - 20s	1772	1465	261	1.209	0.6215
>20 - 30s	129	1587	261	0.081	0.9998
>20 - 40<	-1491	1791	261	-0.833	0.8389
20s - 30s	-1643	982	261	-1.673	0.3398
20s - 40<	-3263	1285	261	-2.539	0.0564
30s - 40<	-1620	1423	261	-1.139	0.6658

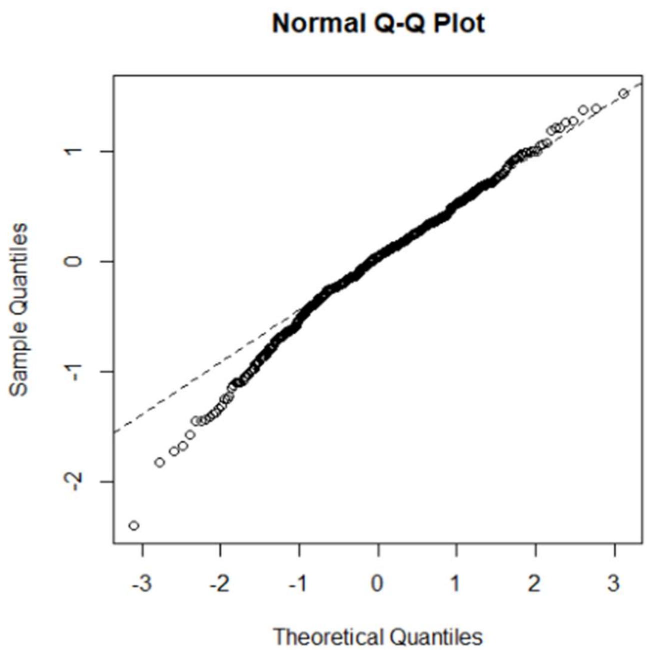
Degrees-of-freedom method: inherited from kenward-roger when re-gridding
P value adjustment: tukey method for comparing a family of 4 estimates

Plots of EMMs for the age groupings



Filled Pauses: Age Groups and Speaker Gender

Quantile-Quantile plots to check for normality assumptions



ANOVA result table for Filled Pauses per million words

Type III Analysis of Variance Table with Satterthwaite's method

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Age.Pairing	2.53622	0.84541	3	262.17	2.0064	0.1135
Gender	0.56244	0.56244	1	497.33	1.3348	0.2485
Age.Pairing:Gender	2.04993	0.68331	3	499.77	1.6217	0.1834

Table of Estimated Marginal Means (EMMs) values

Age.Pairing	Gender	response	SE	df	lower.CL	upper.CL
>20	F Speaker	10687	2005	384	7390	15456
20s	F Speaker	7749	572	334	6702	8960
30s	F Speaker	9775	987	376	8016	11921
40<	F Speaker	12443	1634	357	9610	16110
>20	M Speaker	10462	1660	337	7657	14295
20s	M Speaker	10889	1022	407	9054	13096
30s	M Speaker	11178	1169	389	9100	13729
40<	M Speaker	11689	1479	348	9115	14991

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95
Intervals are back-transformed from the log scale

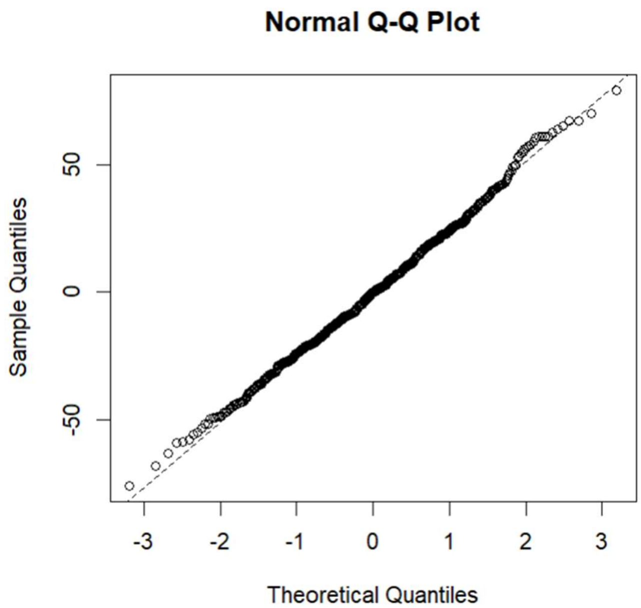
Table of pairwise contrasts

contrast	estimate	SE	df	t.ratio	p.value
>20 F Speaker - 20s F Speaker	2938	2085	334	1.409	0.8529
>20 F Speaker - 30s F Speaker	912	2235	376	0.408	0.9999
>20 F Speaker - 40< F Speaker	-1756	2587	357	-0.679	0.9975
>20 F Speaker - >20 M Speaker	225	2466	337	0.091	1.0000
>20 F Speaker - 20s M Speaker	-202	2251	384	-0.090	1.0000
>20 F Speaker - 30s M Speaker	-491	2321	384	-0.211	1.0000
>20 F Speaker - 40< M Speaker	-1002	2492	348	-0.402	0.9999
20s F Speaker - 30s F Speaker	-2026	1140	334	-1.776	0.6365
20s F Speaker - 40< F Speaker	-4694	1732	334	-2.711	0.1229
20s F Speaker - >20 M Speaker	-2713	1756	334	-1.545	0.7823
20s F Speaker - 20s M Speaker	-3140	1113	334	-2.821	0.0933
20s F Speaker - 30s M Speaker	-3429	1301	334	-2.635	0.1473
20s F Speaker - 40< M Speaker	-3940	1585	334	-2.485	0.2053
30s F Speaker - 40< F Speaker	-2668	1909	357	-1.398	0.8581
30s F Speaker - >20 M Speaker	-687	1931	337	-0.356	1.0000
30s F Speaker - 20s M Speaker	-1114	1421	376	-0.784	0.9938
30s F Speaker - 30s M Speaker	-1403	1424	376	-0.985	0.9763
30s F Speaker - 40< M Speaker	-1914	1778	348	-1.077	0.9612
40< F Speaker - >20 M Speaker	1981	2330	337	0.850	0.9900
40< F Speaker - 20s M Speaker	1554	1928	357	0.806	0.9927
40< F Speaker - 30s M Speaker	1265	2009	357	0.630	0.9985
40< F Speaker - 40< M Speaker	754	2095	348	0.360	1.0000
>20 M Speaker - 20s M Speaker	-427	1950	337	-0.219	1.0000
>20 M Speaker - 30s M Speaker	-716	2030	337	-0.352	1.0000
>20 M Speaker - 40< M Speaker	-1227	2223	337	-0.552	0.9993
20s M Speaker - 30s M Speaker	-288	1553	389	-0.186	1.0000
20s M Speaker - 40< M Speaker	-800	1798	348	-0.445	0.9998
30s M Speaker - 40< M Speaker	-512	1885	348	-0.271	1.0000

Degrees-of-freedom method: inherited from kenward-roger when re-gridding
P value adjustment: tukey method for comparing a family of 8 estimates

Singlish particles: Speaker Gender, Partner Gender, Dyad Type

Quantile-Quantile plots to check for normality assumptions



ANOVA result table for Singlish particles per million words

```
Type III Analysis of Variance Table with Satterthwaite's method
                      Sum Sq Mean Sq NumDF  DenDF F value   Pr(>F)
Gender                8934.9  8934.9     1 616.47  8.4681 0.003745 **
Partner.Gender        1533.6  1533.6     1 616.47  1.4534 0.228439
Gender:Partner.Gender 3419.1  3419.1     1 344.00  3.2405 0.072714 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table of Estimated Marginal Means (EMMs) values

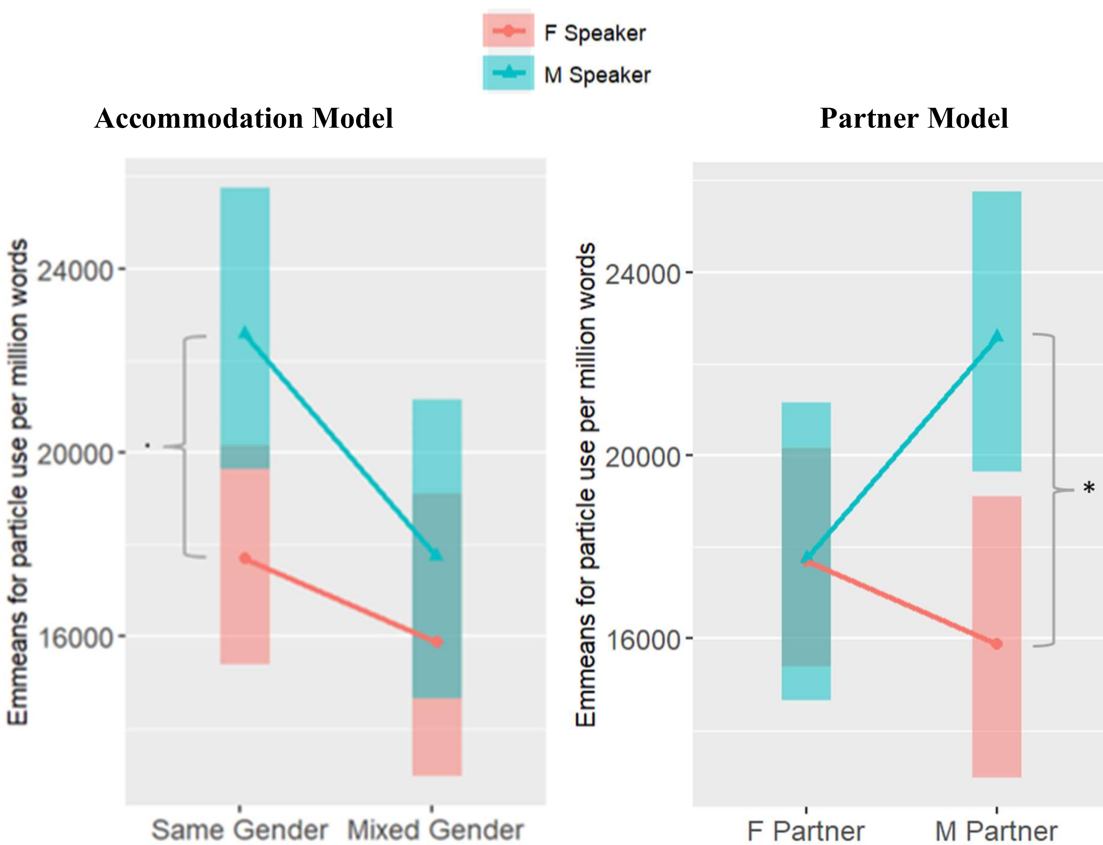
```
 Gender    Partner.Gender response   SE  df lower.CL upper.CL
 F Speaker F Partner         17693 1211 344    15391    20154
 M Speaker F Partner         17745 1655 462    14642    21145
 F Speaker M Partner         15889 1566 462    12961    19115
 M Speaker M Partner         22589 1551 344    19641    25743

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95
Intervals are back-transformed from the sqrt scale
```

Table of pairwise contrasts

contrast	estimate	SE	df	t.ratio	p.value
F Speaker F Partner - M Speaker F Partner	-52.2	2050	344	-0.025	1.0000
F Speaker F Partner - F Speaker M Partner	1803.4	1979	344	0.911	0.7989
F Speaker F Partner - M Speaker M Partner	-4896.2	1968	344	-2.488	0.0635
M Speaker F Partner - F Speaker M Partner	1855.6	1251	462	1.483	0.4486
M Speaker F Partner - M Speaker M Partner	-4844.0	2268	344	-2.136	0.1439
F Speaker M Partner - M Speaker M Partner	-6699.6	2204	344	-3.040	0.0135

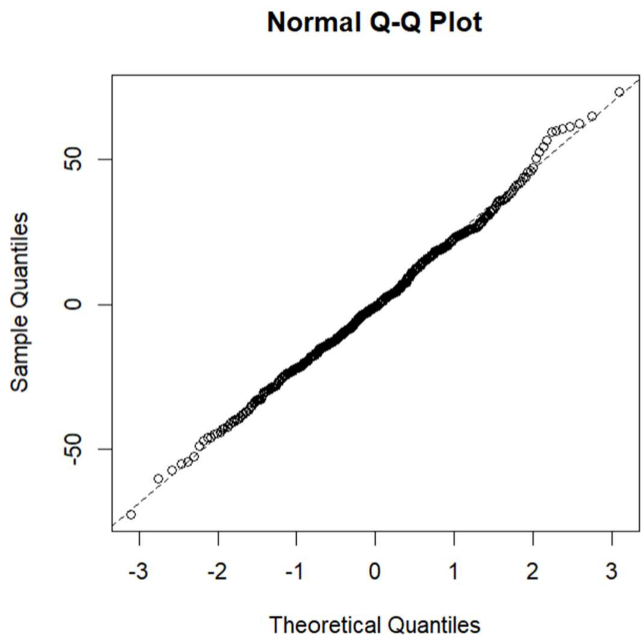
```
Degrees-of-freedom method: inherited from kenward-roger when re-gridding
P value adjustment: tukey method for comparing a family of 4 estimates
```

Plots of EMMs for Accommodation Model and Partner Model



Singlish particles: Ethnicity

Quantile-Quantile plots to check for normality assumptions



ANOVA result table for Singlish particles per million words

```
Type III Analysis of Variance Table with Satterthwaite's method
               Sum Sq Mean Sq NumDF  DenDF F value Pr(>F)
Pair.Ethnicity 1153.5  576.75     2 268.75  0.6129 0.5425
```

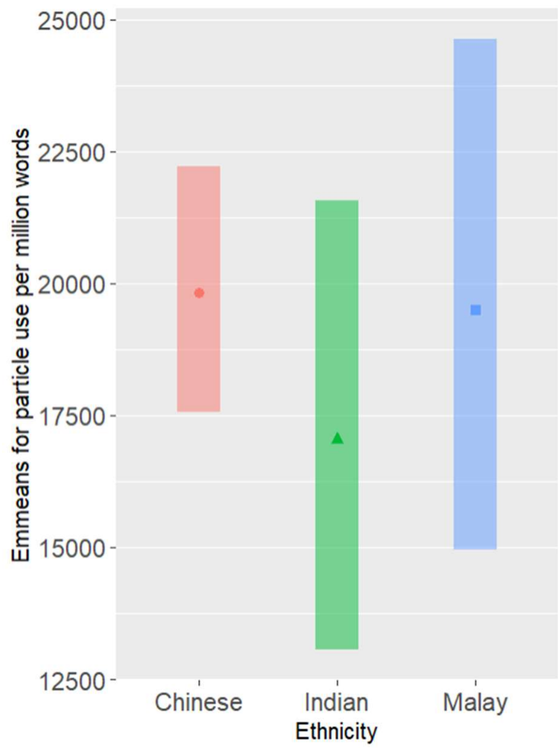
Table of Estimated Marginal Means (EMMs) values

```
 Pair.Ethnicity response   SE  df lower.CL upper.CL
 Chinese           19821 1181 276    17564    22214
 Indian            17045 2160 273    13058    21562
 Malay             19491 2457 268    14954    24629

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95
Intervals are back-transformed from the sqrt scale
```

Table of pairwise contrasts

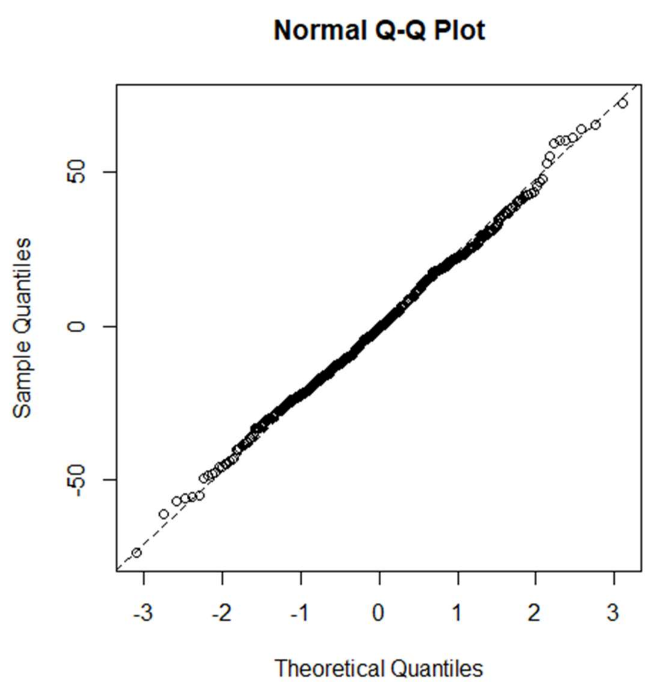
```
 contrast         estimate   SE  df t.ratio p.value
 Chinese - Indian     2776 2462 273   1.128  0.4979
 Chinese - Malay       329 2726 268   0.121  0.9920
 Indian - Malay      -2446 3271 268  -0.748  0.7352

Degrees-of-freedom method: inherited from kenward-roger when re-gridding
P value adjustment: tukey method for comparing a family of 3 estimates
```

Plots of EMMs for the three ethnicity pairings



Singlish particles: Ethnicity and Speaker Gender

Quantile-Quantile plots to check for normality assumptions



ANOVA result table for Singlish particles per million words

Type III Analysis of Variance Table with Satterthwaite's method

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)	
Pair.Ethnicity	2379.8	1189.9	2	284.40	1.2550	0.286641	
Gender	9197.6	9197.6	1	419.28	9.7008	0.001968	**
Pair.Ethnicity:Gender	3474.2	1737.1	2	418.42	1.8322	0.161348	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Table of Estimated Marginal Means (EMMs) values

Pair.Ethnicity	Gender	response	SE	df	lower.CL	upper.CL
Chinese	F Speaker	19335	1256	341	16944	21884
Indian	F Speaker	13765	2564	468	9187	19265
Malay	F Speaker	14692	2833	457	9652	20787
Chinese	M Speaker	20874	1644	475	17768	24230
Indian	M Speaker	18863	2461	341	14334	24014
Malay	M Speaker	22286	2864	340	17009	28276

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95
Intervals are back-transformed from the sqrt scale

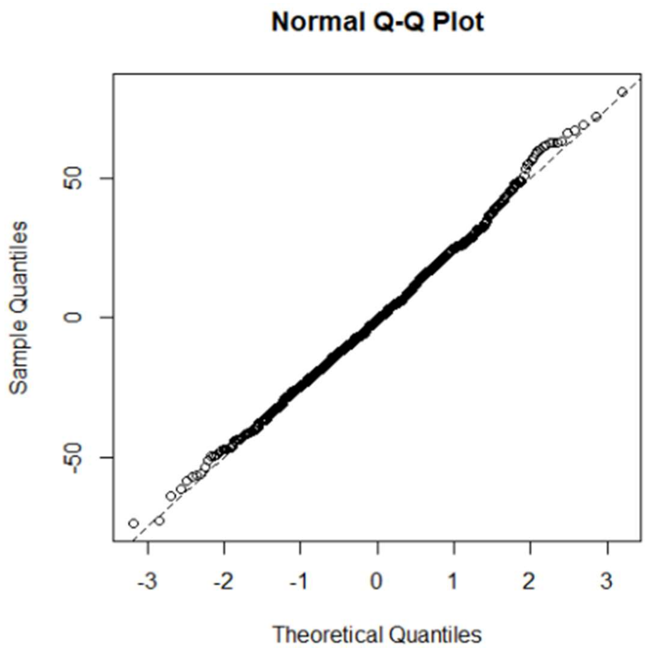
Table of pairwise contrasts

contrast	estimate	SE	df	t.ratio	p.value
Chinese F Speaker - Indian F Speaker	5570	2855	341	1.951	0.3731
Chinese F Speaker - Malay F Speaker	4643	3099	341	1.498	0.6656
Chinese F Speaker - Chinese M Speaker	-1539	1638	341	-0.939	0.9360
Chinese F Speaker - Indian M Speaker	472	2762	341	0.171	1.0000
Chinese F Speaker - Malay M Speaker	-2951	3127	340	-0.944	0.9348
Indian F Speaker - Malay F Speaker	-927	3821	457	-0.243	0.9999
Indian F Speaker - Chinese M Speaker	-7109	3046	468	-2.334	0.1827
Indian F Speaker - Indian M Speaker	-5099	2753	341	-1.852	0.4338
Indian F Speaker - Malay M Speaker	-8522	3844	340	-2.217	0.2327
Malay F Speaker - Chinese M Speaker	-6182	3276	457	-1.887	0.4114
Malay F Speaker - Indian M Speaker	-4171	3752	341	-1.112	0.8764
Malay F Speaker - Malay M Speaker	-7594	3169	340	-2.396	0.1604
Chinese M Speaker - Indian M Speaker	2010	2959	341	0.679	0.9841
Chinese M Speaker - Malay M Speaker	-1413	3302	340	-0.428	0.9982
Indian M Speaker - Malay M Speaker	-3423	3776	340	-0.907	0.9447

Degrees-of-freedom method: inherited from kenward-roger when re-gridding
P value adjustment: tukey method for comparing a family of 6 estimates

Singlish particles: Mixed Ethnicity vs. Same Ethnicity

Quantile-Quantile plots to check for normality assumptions



ANOVA result table for Singlish particles per million words

```
Type III Analysis of Variance Table with Satterthwaite's method
               Sum Sq Mean Sq NumDF  DenDF F value Pr(>F)
Pair.Ethnicity 7.9416  7.9416     1 691.18  0.0075 0.9311
```

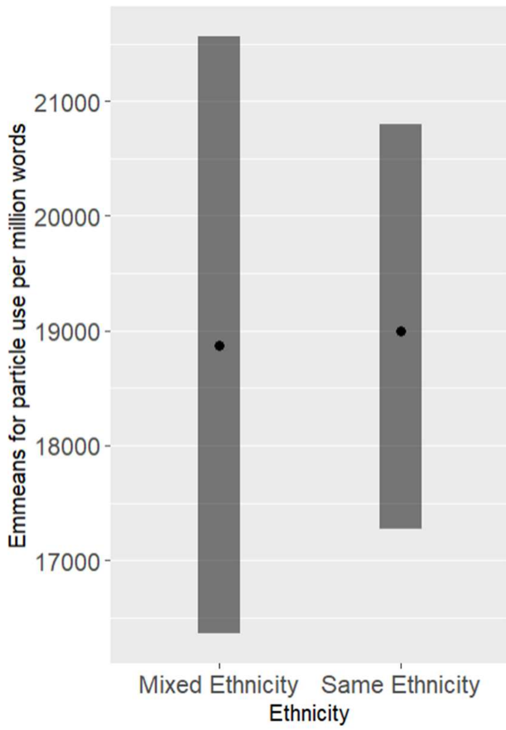
Table of Estimated Marginal Means (EMMs) values

```
 Pair.Ethnicity  response   SE  df lower.CL upper.CL
 Mixed Ethnicity    18872 1324 613    16362    21562
 Same Ethnicity     18994  895 410    17275    20795

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95
Intervals are back-transformed from the sqrt scale
```

Table of pairwise contrasts

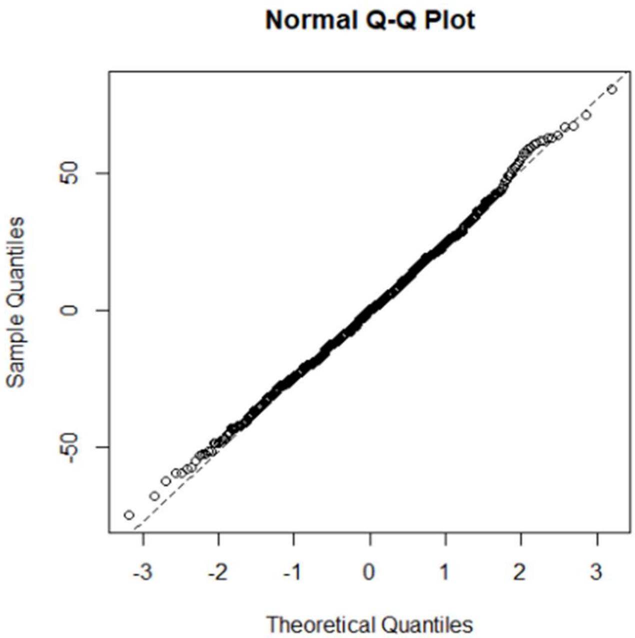
```
 contrast                          estimate   SE  df t.ratio p.value
 Mixed Ethnicity - Same Ethnicity      -122 1410 410  -0.086  0.9312

Degrees-of-freedom method: inherited from kenward-roger when re-gridding
```

Plots of EMMs for the three ethnicity pairings



Singlish particles: Mixed Ethnicity vs. Same Ethnicity and Speaker Gender

Quantile-Quantile plots to check for normality assumptions



ANOVA result table for Singlish particles per million words

```
Type III Analysis of Variance Table with Satterthwaite's method
                      Sum Sq Mean Sq NumDF  DenDF F value  Pr(>F)
Pair.Ethnicity          77.9    77.9     1 689.46  0.0737 0.78614
Gender                4510.4  4510.4     1 618.22  4.2666 0.03928 *
Pair.Ethnicity:Gender  210.4   210.4     1 600.74  0.1990 0.65566
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table of Estimated Marginal Means (EMMs) values

```
 Pair.Ethnicity  Gender    response   SE  df lower.CL upper.CL
 Mixed Ethnicity F Speaker    17833 1638 689    14763    21194
 Same Ethnicity  F Speaker    17679 1003 551    15763    19705
 Mixed Ethnicity M Speaker    19849 1737 680    16585    23406
 Same Ethnicity  M Speaker    20806 1200 609    18516    23229

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95
Intervals are back-transformed from the sqrt scale
```

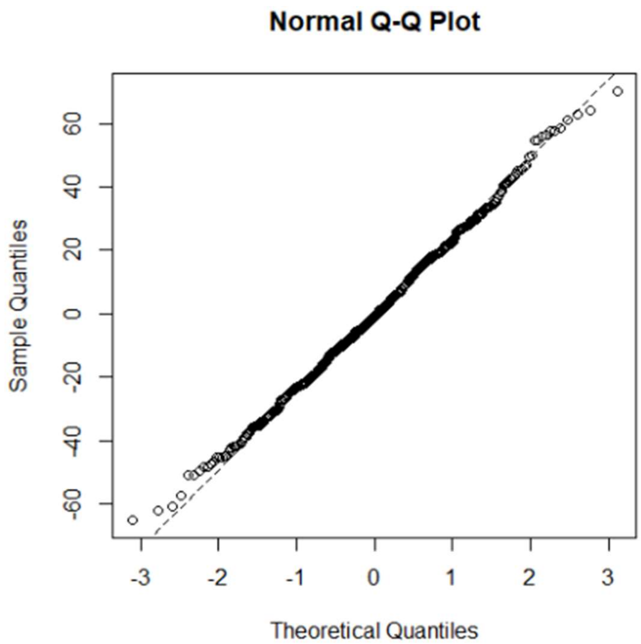
Table of pairwise contrasts

contrast	estimate	SE	df	t.ratio	p.value
Mixed Ethnicity F Speaker - Same Ethnicity F Speaker	155	1758	551	0.088	0.9998
Mixed Ethnicity F Speaker - Mixed Ethnicity M Speaker	-2016	2111	680	-0.955	0.7751
Mixed Ethnicity F Speaker - Same Ethnicity M Speaker	-2973	1896	609	-1.568	0.3977
Same Ethnicity F Speaker - Mixed Ethnicity M Speaker	-2170	1874	551	-1.158	0.6537
Same Ethnicity F Speaker - Same Ethnicity M Speaker	-3127	1276	551	-2.451	0.0690
Mixed Ethnicity M Speaker - Same Ethnicity M Speaker	-957	1971	609	-0.485	0.9623

Degrees-of-freedom method: inherited from kenward-roger when re-gridding
P value adjustment: tukey method for comparing a family of 4 estimates

Singlish particles: Age Groups

Quantile-Quantile plots to check for normality assumptions



ANOVA result table for Singlish particles per million words

Type III Analysis of Variance Table with Satterthwaite's method

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)	
Age.Pairing	12640	4213.4	3	261	4.4211	0.004719	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table of Estimated Marginal Means (EMMs) values

Age.Pairing	response	SE	df	lower.CL	upper.CL
>20	16695	2636	261	11907	22290
20s	16803	1271	261	14395	19398
30s	16945	1615	261	13913	20275
40<	25774	2457	261	21164	30839

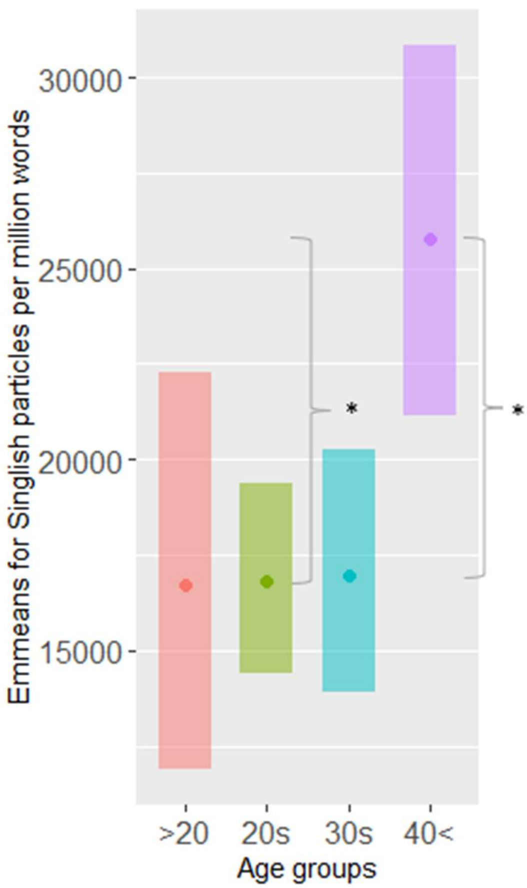
Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95
Intervals are back-transformed from the sqrt scale

Table of pairwise contrasts

contrast	estimate	SE	df	t.ratio	p.value
>20 - 20s	-108	2927	261	-0.037	1.0000
>20 - 30s	-250	3092	261	-0.081	0.9998
>20 - 40<	-9079	3604	261	-2.519	0.0593
20s - 30s	-141	2055	261	-0.069	0.9999
20s - 40<	-8971	2766	261	-3.243	0.0073
30s - 40<	-8830	2940	261	-3.003	0.0154

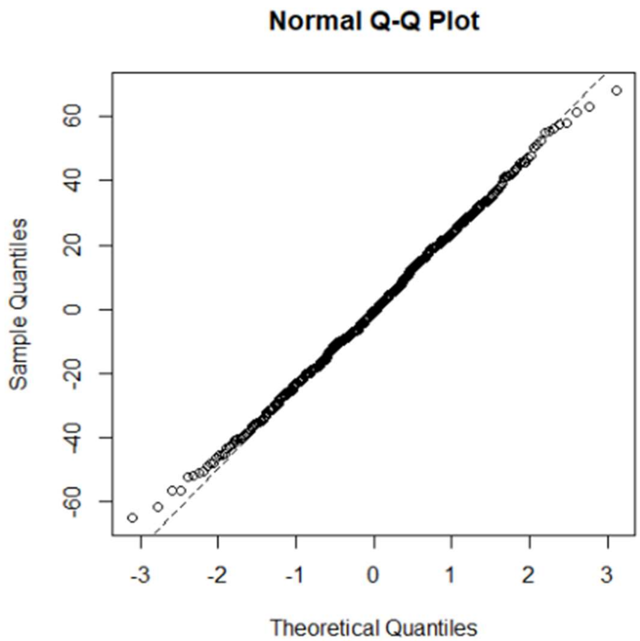
Degrees-of-freedom method: inherited from kenward-roger when re-gridding
P value adjustment: tukey method for comparing a family of 4 estimates

Plots of EMMs for the age groupings



Singlish particles: Age Groups and Speaker Gender

Quantile-Quantile plots to check for normality assumptions



ANOVA result table for Singlish particles per million words

```
Type III Analysis of Variance Table with Satterthwaite's method
                    Sum Sq Mean Sq NumDF  DenDF F value   Pr(>F)   
Age.Pairing        12752.3  4250.8     3 258.92  4.4458 0.004572 **
Gender              4704.4  4704.4     1 484.43  4.9202 0.027007 * 
Age.Pairing:Gender  5185.2  1728.4     3 478.15  1.8077 0.144854   
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table of Estimated Marginal Means (EMMs) values

```
 Age.Pairing Gender    response   SE  df lower.CL upper.CL
 >20         F Speaker    11849 2863 421     6889    18146
 20s         F Speaker    15672 1364 340    13104    18469
 30s         F Speaker    16824 1870 378    13348    20702
 40<         F Speaker    26063 3041 390    20427    32386
 >20         M Speaker    20508 3333 352    14476    27587
 20s         M Speaker    18843 1775 442    15516    22493
 30s         M Speaker    17076 1930 395    13492    21081
 40<         M Speaker    25510 2926 375    20081    31587

Degrees-of-freedom method: kenward-roger 
Confidence level used: 0.95 
Intervals are back-transformed from the sqrt scale
```

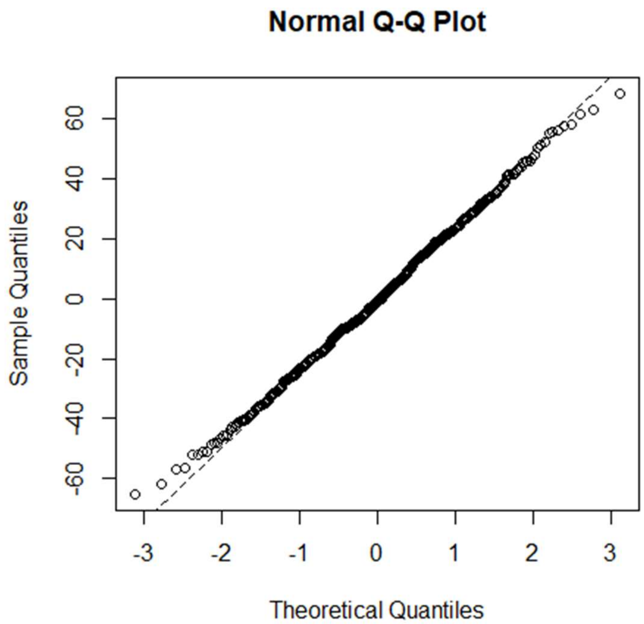
Table of pairwise contrasts

contrast	estimate	SE	df	t.ratio	p.value
>20 F Speaker - 20s F Speaker	-3823	3172	340	-1.205	0.9300
>20 F Speaker - 30s F Speaker	-4975	3420	378	-1.455	0.8307
>20 F Speaker - 40< F Speaker	-14215	4177	390	-3.403	0.0167
>20 F Speaker - >20 M Speaker	-8659	3568	352	-2.427	0.2315
>20 F Speaker - 20s M Speaker	-6994	3369	421	-2.076	0.4321
>20 F Speaker - 30s M Speaker	-5227	3453	395	-1.514	0.7998
>20 F Speaker - 40< M Speaker	-13661	4094	375	-3.337	0.0207
20s F Speaker - 30s F Speaker	-1152	2315	340	-0.498	0.9997
20s F Speaker - 40< F Speaker	-10392	3333	340	-3.118	0.0410
20s F Speaker - >20 M Speaker	-4836	3601	340	-1.343	0.8817
20s F Speaker - 20s M Speaker	-3171	1801	340	-1.761	0.6472
20s F Speaker - 30s M Speaker	-1404	2363	340	-0.594	0.9989
20s F Speaker - 40< M Speaker	-9838	3228	340	-3.048	0.0503
30s F Speaker - 40< F Speaker	-9239	3570	378	-2.588	0.1636
30s F Speaker - >20 M Speaker	-3684	3822	352	-0.964	0.9791
30s F Speaker - 20s M Speaker	-2019	2578	378	-0.783	0.9939
30s F Speaker - 30s M Speaker	-252	2057	378	-0.123	1.0000
30s F Speaker - 40< M Speaker	-8685	3472	375	-2.501	0.1980
40< F Speaker - >20 M Speaker	5556	4512	352	1.231	0.9221
40< F Speaker - 20s M Speaker	7220	3521	390	2.050	0.4493
40< F Speaker - 30s M Speaker	8987	3602	390	2.495	0.2005
40< F Speaker - 40< M Speaker	554	3461	375	0.160	1.0000
>20 M Speaker - 20s M Speaker	1664	3776	352	0.441	0.9999
>20 M Speaker - 30s M Speaker	3432	3851	352	0.891	0.9868
>20 M Speaker - 40< M Speaker	-5002	4435	352	-1.128	0.9504
20s M Speaker - 30s M Speaker	1767	2622	395	0.674	0.9976
20s M Speaker - 40< M Speaker	-6666	3422	375	-1.948	0.5186
30s M Speaker - 40< M Speaker	-8433	3505	375	-2.406	0.2413

Degrees-of-freedom method: inherited from kenward-roger when re-gridding
P value adjustment: tukey method for comparing a family of 8 estimates

LSM Scores

Quantile-Quantile plots to check for normality assumptions



ANOVA result table for LSM Scores

```
Analysis of Variance Table

Response: C.LSM
                Df   Sum Sq    Mean Sq F value  Pr(>F)
Gender.Pairing   2 0.005401 0.00270059  4.1794 0.01609 *
Residuals      344 0.222284 0.00064617
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

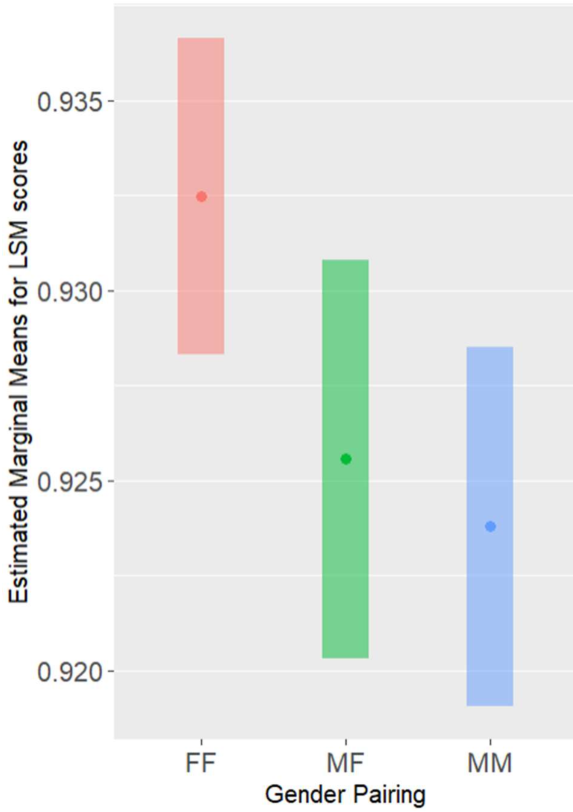
Table of Estimated Marginal Means (EMMs) values

```
 Gender.Pairing emmean       SE  df lower.CL upper.CL
 FF             0.9325 0.002118 344   0.9283   0.9366
 MF             0.9256 0.002665 344   0.9203   0.9308
 MM             0.9238 0.002402 344   0.9191   0.9285

Confidence level used: 0.95
```

Table of pairwise contrasts

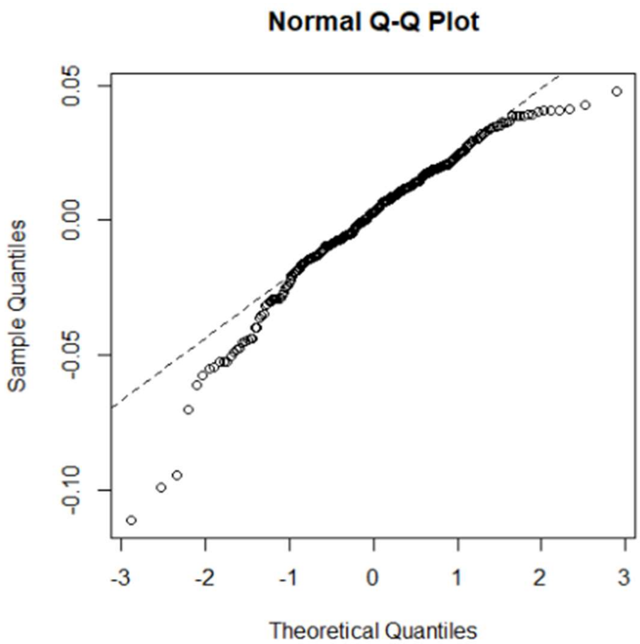
```
 contrast estimate      SE  df t.ratio p.value
 FF - MF   0.00691 0.00340 344   2.030  0.1066
 FF - MM   0.00868 0.00320 344   2.712  0.0192
 MF - MM   0.00177 0.00359 344   0.494  0.8740

P value adjustment: tukey method for comparing a family of 3 estimates
```

Plots of EMMs for the three gender pair types



LSM Scores: Ethnicity

Quantile-Quantile plots to check for normality assumptions



ANOVA result table for LSM Scores

```
Analysis of Variance Table

Response: C.LSM
                  Df    Sum Sq    Mean Sq F value  Pr(>F)
Pair.Ethnicity     2 0.004235 0.00211734   3.127 0.04554 *
Residuals        254 0.171990 0.00067713
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

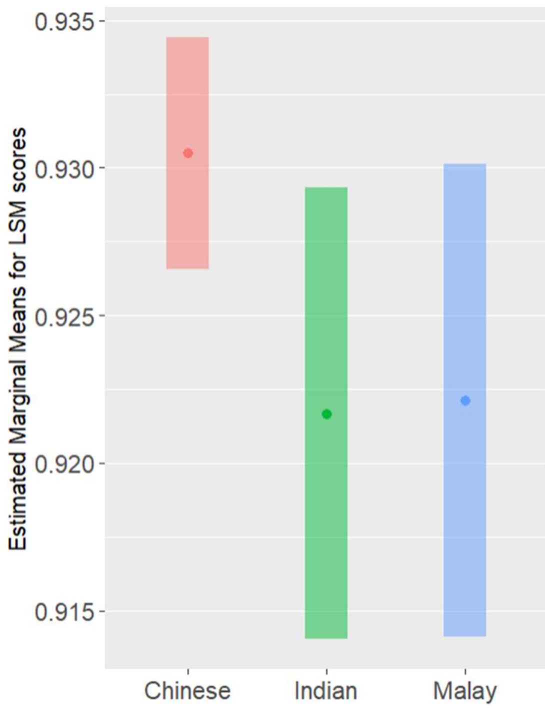
Table of Estimated Marginal Means (EMMs) values

```
 Pair.Ethnicity emmean       SE  df lower.CL upper.CL
 Chinese         0.9305 0.001990 254   0.9266   0.9344
 Indian          0.9217 0.003879 254   0.9140   0.9293
 Malay           0.9221 0.004064 254   0.9141   0.9301

Confidence level used: 0.95
```

Table of pairwise contrasts

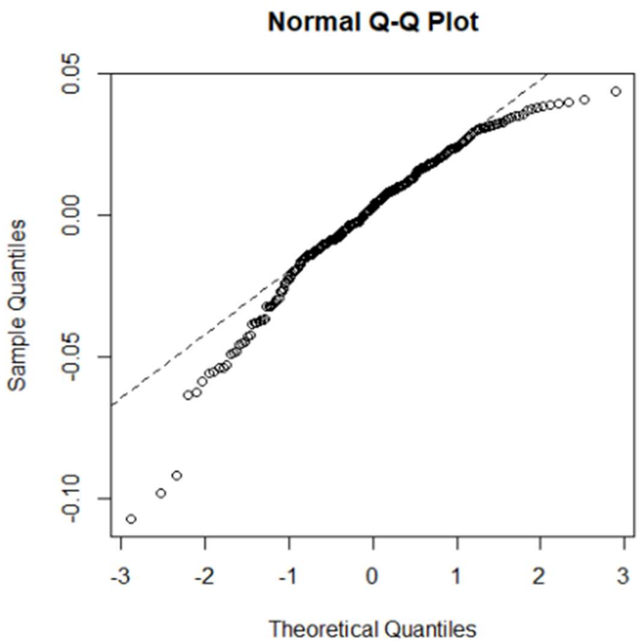
```
 contrast         estimate      SE  df t.ratio p.value
 Chinese - Indian  0.00882 0.00436 254   2.022  0.1089
 Chinese - Malay   0.00836 0.00452 254   1.847  0.1566
 Indian - Malay   -0.00046 0.00562 254  -0.082  0.9963

P value adjustment: tukey method for comparing a family of 3 estimates
```

Plots of EMMs for the three gender pair types



LSM Scores: Ethnicity and Speaker Gender

Quantile-Quantile plots to check for normality assumptions



ANOVA result table for LSM Scores

```
                               Df   Sum Sq    Mean Sq F value  Pr(>F)  
Pair.Ethnicity                  2 0.004235 0.00211734  3.1634 0.04399 *
Gender.Pairing                  2 0.001821 0.00091055  1.3604 0.25847  
Pair.Ethnicity:Gender.Pairing   4 0.004175 0.00104367  1.5593 0.18577  
Residuals                     248 0.165994 0.00066933                  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table of Estimated Marginal Means (EMMs) values

```
 Pair.Ethnicity Gender.Pairing emmean      SE  df lower.CL upper.CL
 Chinese        FF              0.932 0.00267 248    0.926    0.937
 Indian         FF              0.930 0.00862 248    0.913    0.947
 Malay          FF              0.936 0.00862 248    0.919    0.953
 Chinese        MF              0.928 0.00377 248    0.920    0.935
 Indian         MF              0.909 0.00780 248    0.893    0.924
 Malay          MF              0.926 0.00862 248    0.909    0.943
 Chinese        MM              0.931 0.00472 248    0.922    0.940
 Indian         MM              0.925 0.00517 248    0.914    0.935
 Malay          MM              0.915 0.00539 248    0.905    0.926

Confidence level used: 0.95
```

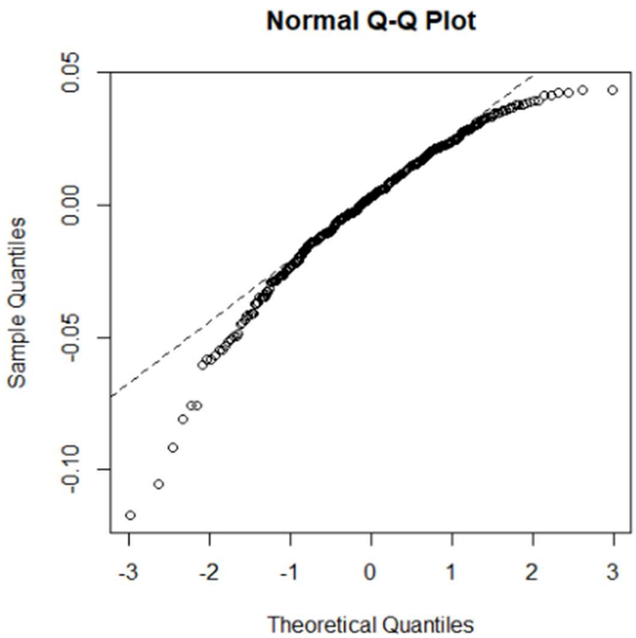
Table of pairwise contrasts

contrast	estimate	SE	df	t.ratio	p.value
Chinese FF - Indian FF	0.002046	0.00903	248	0.227	1.0000
Chinese FF - Malay FF	-0.004067	0.00903	248	-0.451	1.0000
Chinese FF - Chinese MF	0.003908	0.00462	248	0.846	0.9953
Chinese FF - Indian MF	0.022996	0.00824	248	2.789	0.1238
Chinese FF - Malay MF	0.005690	0.00903	248	0.630	0.9994
Chinese FF - Chinese MM	0.000692	0.00543	248	0.128	1.0000
Chinese FF - Indian MM	0.007168	0.00582	248	1.231	0.9489
Chinese FF - Malay MM	0.016394	0.00602	248	2.724	0.1448
Indian FF - Malay FF	-0.006114	0.01220	248	-0.501	0.9999
Indian FF - Chinese MF	0.001862	0.00941	248	0.198	1.0000
Indian FF - Indian MF	0.020950	0.01163	248	1.802	0.6812
Indian FF - Malay MF	0.003644	0.01220	248	0.299	1.0000
Indian FF - Chinese MM	-0.001354	0.00983	248	-0.138	1.0000
Indian FF - Indian MM	0.005122	0.01006	248	0.509	0.9999
Indian FF - Malay MM	0.014348	0.01017	248	1.411	0.8930
Malay FF - Chinese MF	0.007975	0.00941	248	0.847	0.9952
Malay FF - Indian MF	0.027063	0.01163	248	2.327	0.3299
Malay FF - Malay MF	0.009757	0.01220	248	0.800	0.9968
Malay FF - Chinese MM	0.004759	0.00983	248	0.484	0.9999
Malay FF - Indian MM	0.011235	0.01006	248	1.117	0.9712
Malay FF - Malay MM	0.020462	0.01017	248	2.012	0.5368
Chinese MF - Indian MF	0.019088	0.00867	248	2.203	0.4071
Chinese MF - Malay MF	0.001782	0.00941	248	0.189	1.0000
Chinese MF - Chinese MM	-0.003216	0.00605	248	-0.532	0.9998
Chinese MF - Indian MM	0.003260	0.00640	248	0.509	0.9999
Chinese MF - Malay MM	0.012486	0.00658	248	1.897	0.6168
Indian MF - Malay MF	-0.017306	0.01163	248	-1.488	0.8602
Indian MF - Chinese MM	-0.022304	0.00912	248	-2.446	0.2642
Indian MF - Indian MM	-0.015828	0.00936	248	-1.691	0.7517
Indian MF - Malay MM	-0.006602	0.00948	248	-0.696	0.9988
Malay MF - Chinese MM	-0.004998	0.00983	248	-0.508	0.9999
Malay MF - Indian MM	0.001478	0.01006	248	0.147	1.0000
Malay MF - Malay MM	0.010704	0.01017	248	1.052	0.9801
Chinese MM - Indian MM	0.006476	0.00701	248	0.924	0.9914
Chinese MM - Malay MM	0.015702	0.00717	248	2.190	0.4155
Indian MM - Malay MM	0.009226	0.00747	248	1.234	0.9482

P value adjustment: tukey method for comparing a family of 9 estimates

LSM Scores: Mixed Ethnicity vs. Same Ethnicity

Quantile-Quantile plots to check for normality assumptions



ANOVA result table for LSM Scores

```
Analysis of Variance Table

Response: C.LSM
                Df   Sum Sq    Mean Sq F value Pr(>F)
Pair.Ethnicity   1 0.000063 0.00006257  0.0948 0.7583
Residuals      345 0.227623 0.00065978
```

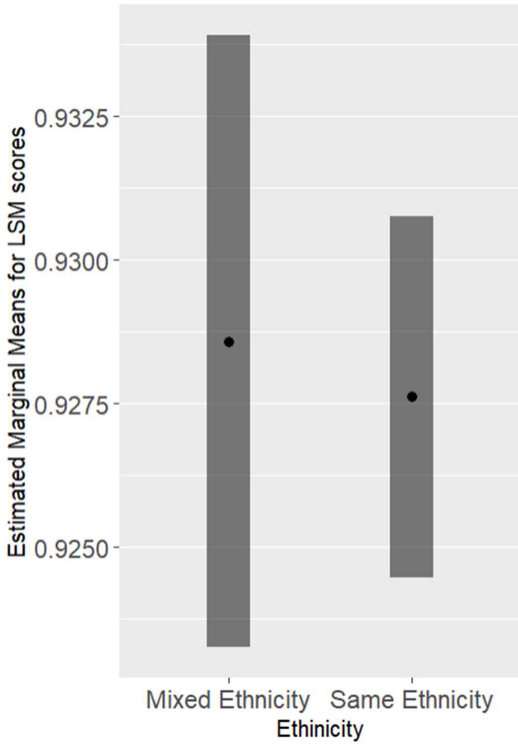
Table of Estimated Marginal Means (EMMs) values

```
 Pair.Ethnicity  emmean       SE  df lower.CL upper.CL
 Mixed Ethnicity 0.9286 0.002708 345   0.9233   0.9339
 Same Ethnicity  0.9276 0.001602 345   0.9245   0.9308

Confidence level used: 0.95
```

Table of pairwise contrasts

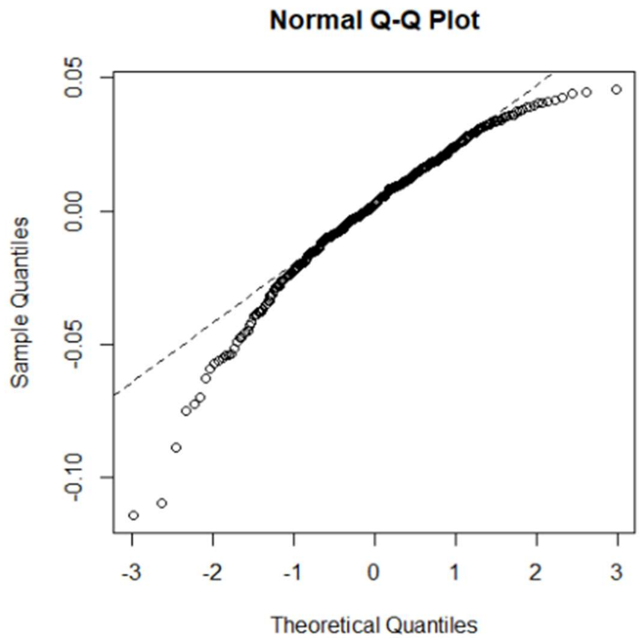
```
 contrast                          estimate      SE  df t.ratio p.value
 Mixed Ethnicity - Same Ethnicity 0.000969 0.00315 345   0.308  0.7583
```

Plots of EMMs for the three ethnicity pairings



LSM Scores: Mixed Ethnicity vs. Same Ethnicity and Speaker Gender

Quantile-Quantile plots to check for normality assumptions



ANOVA result table for LSM Scores

```
                               Df   Sum Sq    Mean Sq F value  Pr(>F)
Pair.Ethnicity                  1 0.000063 0.00006257  0.0963 0.75657
Gender.Pairing                  2 0.005522 0.00276119  4.2475 0.01506 *
Pair.Ethnicity:Gender.Pairing   2 0.000424 0.00021202  0.3261 0.72192
Residuals                     341 0.221676 0.00065008
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table of Estimated Marginal Means (EMMs) values

```
 Pair.Ethnicity  Gender.Pairing emmean      SE  df lower.CL upper.CL
 Mixed Ethnicity FF              0.935 0.00451 341    0.926    0.944
 Same Ethnicity  FF              0.932 0.00241 341    0.927    0.937
 Mixed Ethnicity MF              0.929 0.00520 341    0.919    0.939
 Same Ethnicity  MF              0.924 0.00311 341    0.918    0.931
 Mixed Ethnicity MM              0.923 0.00437 341    0.914    0.931
 Same Ethnicity  MM              0.924 0.00289 341    0.919    0.930

Confidence level used: 0.95
```

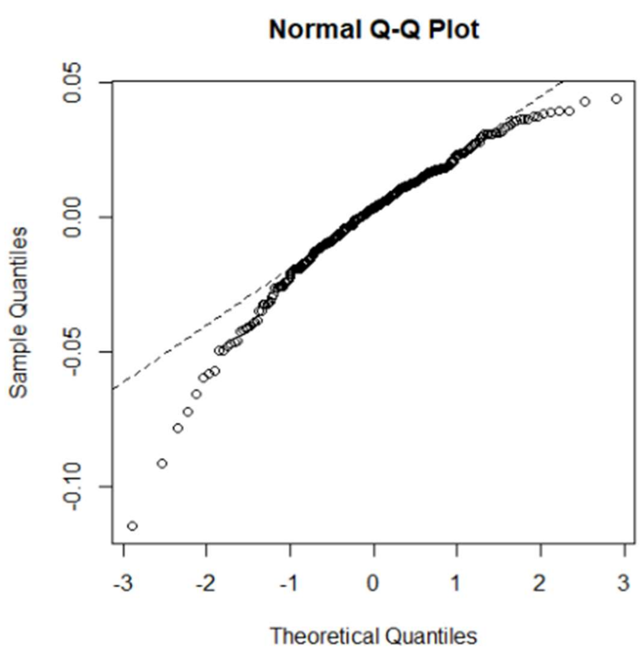
Table of pairwise contrasts

contrast	estimate	SE	df	t.ratio	p.value
Mixed Ethnicity FF - Same Ethnicity FF	0.002842	0.00511	341	0.556	0.9937
Mixed Ethnicity FF - Mixed Ethnicity MF	0.005870	0.00688	341	0.853	0.9572
Mixed Ethnicity FF - Same Ethnicity MF	0.010285	0.00548	341	1.877	0.4179
Mixed Ethnicity FF - Mixed Ethnicity MM	0.012025	0.00628	341	1.915	0.3948
Mixed Ethnicity FF - Same Ethnicity MM	0.010402	0.00535	341	1.943	0.3776
Same Ethnicity FF - Mixed Ethnicity MF	0.003029	0.00574	341	0.528	0.9950
Same Ethnicity FF - Same Ethnicity MF	0.007444	0.00394	341	1.890	0.4099
Same Ethnicity FF - Mixed Ethnicity MM	0.009183	0.00499	341	1.839	0.4419
Same Ethnicity FF - Same Ethnicity MM	0.007560	0.00376	341	2.011	0.3384
Mixed Ethnicity MF - Same Ethnicity MF	0.004415	0.00607	341	0.728	0.9784
Mixed Ethnicity MF - Mixed Ethnicity MM	0.006154	0.00680	341	0.905	0.9450
Mixed Ethnicity MF - Same Ethnicity MM	0.004532	0.00595	341	0.761	0.9737
Same Ethnicity MF - Mixed Ethnicity MM	0.001739	0.00537	341	0.324	0.9995
Same Ethnicity MF - Same Ethnicity MM	0.000117	0.00425	341	0.027	1.0000
Mixed Ethnicity MM - Same Ethnicity MM	-0.001623	0.00524	341	-0.310	0.9996

P value adjustment: tukey method for comparing a family of 6 estimates

LSM Scores: Age Groups

Quantile-Quantile plots to check for normality assumptions



ANOVA result table for LSM Scores

Anova Table (Type II tests)

Response: C.LSM

	Sum Sq	Df	F value	Pr(>F)
Age.Pairing	0.002724	3	1.4969	0.2158
Residuals	0.158343	261		

Table of Estimated Marginal Means (EMMs) values

Age.Pairing	emmean	SE	df	lower.CL	upper.CL
>20	0.9258	0.004740	261	0.9165	0.9351
20s	0.9325	0.002277	261	0.9280	0.9370
30s	0.9272	0.002883	261	0.9216	0.9329
40<	0.9249	0.003555	261	0.9179	0.9319

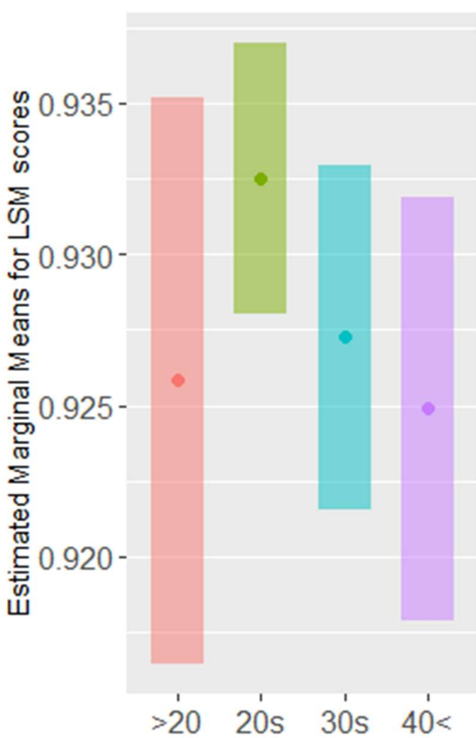
Confidence level used: 0.95

Table of pairwise contrasts

contrast	estimate	SE	df	t.ratio	p.value
>20 - 20s	-0.006665	0.00526	261	-1.267	0.5846
>20 - 30s	-0.001437	0.00555	261	-0.259	0.9939
>20 - 40<	0.000926	0.00593	261	0.156	0.9986
20s - 30s	0.005227	0.00367	261	1.423	0.4862
20s - 40<	0.007590	0.00422	261	1.798	0.2767
30s - 40<	0.002363	0.00458	261	0.516	0.9551

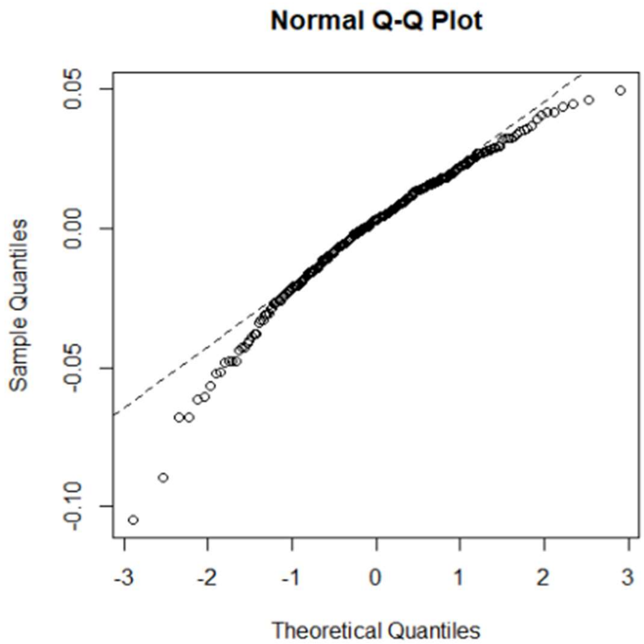
P value adjustment: tukey method for comparing a family of 4 estimates

Plots of EMMs for the age groupings



LSM Scores: Age Groups and Speaker Gender

Quantile-Quantile plots to check for normality assumptions



ANOVA result table for LSM Scores

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Age.Pairing	3	0.002724	0.00090815	1.5191	0.2100
Gender.Pairing	2	0.002064	0.00103195	1.7262	0.1801
Age.Pairing:Gender.Pairing	6	0.005029	0.00083817	1.4020	0.2142
Residuals	253	0.151250	0.00059782		

Table of Estimated Marginal Means (EMMs) values

Age.Pairing	Gender.Pairing	emmean	SE	df	lower.CL	upper.CL
>20	FF	0.937	0.00864	253	0.920	0.954
20s	FF	0.934	0.00316	253	0.928	0.941
30s	FF	0.926	0.00471	253	0.916	0.935
40<	FF	0.935	0.00576	253	0.923	0.946
>20	MF	0.932	0.00998	253	0.912	0.952
20s	MF	0.934	0.00462	253	0.924	0.943
30s	MF	0.925	0.00521	253	0.915	0.936
40<	MF	0.915	0.00773	253	0.900	0.930
>20	MM	0.916	0.00678	253	0.902	0.929
20s	MM	0.927	0.00454	253	0.919	0.936
30s	MM	0.931	0.00499	253	0.921	0.941
40<	MM	0.921	0.00547	253	0.910	0.932

Confidence level used: 0.95

Table of pairwise contrasts

contrast	estimate	SE	df	t.ratio	p.value
>20 FF - 20s FF	0.003038	0.00920	253	0.330	1.0000
>20 FF - 30s FF	0.011763	0.00984	253	1.195	0.9890
>20 FF - 40< FF	0.002587	0.01039	253	0.249	1.0000
>20 FF - >20 MF	0.005397	0.01320	253	0.409	1.0000
>20 FF - 20s MF	0.003866	0.00980	253	0.394	1.0000
>20 FF - 30s MF	0.012001	0.01009	253	1.189	0.9894
>20 FF - 40< MF	0.022297	0.01160	253	1.922	0.7441
>20 FF - >20 MM	0.021636	0.01099	253	1.969	0.7137
>20 FF - 20s MM	0.009961	0.00976	253	1.020	0.9971
>20 FF - 30s MM	0.006727	0.00998	253	0.674	0.9999
>20 FF - 40< MM	0.016625	0.01023	253	1.625	0.8983
20s FF - 30s FF	0.008725	0.00567	253	1.540	0.9280
20s FF - 40< FF	-0.000451	0.00657	253	-0.069	1.0000
20s FF - >20 MF	0.002359	0.01047	253	0.225	1.0000
20s FF - 20s MF	0.000828	0.00560	253	0.148	1.0000
20s FF - 30s MF	0.008963	0.00609	253	1.471	0.9471
20s FF - 40< MF	0.019259	0.00835	253	2.306	0.4750
20s FF - >20 MM	0.018598	0.00748	253	2.486	0.3534
20s FF - 20s MM	0.006923	0.00553	253	1.252	0.9840
20s FF - 30s MM	0.003689	0.00591	253	0.625	1.0000
20s FF - 40< MM	0.013587	0.00631	253	2.152	0.5857
30s FF - 40< FF	-0.009176	0.00744	253	-1.233	0.9858
30s FF - >20 MF	-0.006366	0.01104	253	-0.577	1.0000
30s FF - 20s MF	-0.007897	0.00659	253	-1.197	0.9888
30s FF - 30s MF	0.000238	0.00702	253	0.034	1.0000
30s FF - 40< MF	0.010534	0.00905	253	1.164	0.9911
30s FF - >20 MM	0.009873	0.00825	253	1.196	0.9889
30s FF - 20s MM	-0.001802	0.00654	253	-0.276	1.0000
30s FF - 30s MM	-0.005036	0.00686	253	-0.734	0.9999
30s FF - 40< MM	0.004863	0.00721	253	0.674	0.9999
40< FF - >20 MF	0.002810	0.01153	253	0.244	1.0000
40< FF - 20s MF	0.001279	0.00739	253	0.173	1.0000
40< FF - 30s MF	0.009415	0.00777	253	1.212	0.9877
40< FF - 40< MF	0.019710	0.00964	253	2.044	0.6628
40< FF - >20 MM	0.019049	0.00890	253	2.141	0.5942
40< FF - 20s MM	0.007374	0.00734	253	1.005	0.9975
40< FF - 30s MM	0.004141	0.00762	253	0.543	1.0000
40< FF - 40< MM	0.014039	0.00794	253	1.767	0.8341
>20 MF - 20s MF	-0.001531	0.01100	253	-0.139	1.0000
>20 MF - 30s MF	0.006604	0.01126	253	0.586	1.0000
>20 MF - 40< MF	0.016900	0.01263	253	1.339	0.9732
>20 MF - >20 MM	0.016239	0.01207	253	1.346	0.9721
>20 MF - 20s MM	0.004564	0.01097	253	0.416	1.0000
>20 MF - 30s MM	0.001331	0.01116	253	0.119	1.0000
>20 MF - 40< MM	0.011229	0.01138	253	0.987	0.9979
20s MF - 30s MF	0.008135	0.00697	253	1.168	0.9909
20s MF - 40< MF	0.018431	0.00901	253	2.046	0.6613
20s MF - >20 MM	0.017770	0.00821	253	2.166	0.5762
20s MF - 20s MM	0.006095	0.00648	253	0.941	0.9986
20s MF - 30s MM	0.002861	0.00680	253	0.421	1.0000
20s MF - 40< MM	0.012759	0.00716	253	1.782	0.8261
30s MF - 40< MF	0.010296	0.00933	253	1.104	0.9943
30s MF - >20 MM	0.009635	0.00855	253	1.126	0.9932
30s MF - 20s MM	-0.002040	0.00691	253	-0.295	1.0000
30s MF - 30s MM	-0.005274	0.00722	253	-0.731	0.9999
30s MF - 40< MM	0.004624	0.00755	253	0.612	1.0000
40< MF - >20 MM	-0.000661	0.01028	253	-0.064	1.0000
40< MF - 20s MM	-0.012336	0.00897	253	-1.376	0.9671
40< MF - 30s MM	-0.015570	0.00920	253	-1.692	0.8705
40< MF - 40< MM	-0.005672	0.00947	253	-0.599	1.0000
>20 MM - 20s MM	-0.011675	0.00816	253	-1.431	0.9564
>20 MM - 30s MM	-0.014909	0.00842	253	-1.771	0.8323
>20 MM - 40< MM	-0.005011	0.00871	253	-0.575	1.0000
20s MM - 30s MM	-0.003234	0.00675	253	-0.479	1.0000
20s MM - 40< MM	0.006664	0.00711	253	0.938	0.9987
30s MM - 40< MM	0.009898	0.00740	253	1.337	0.9734

P value adjustment: tukey method for comparing a family of 12 estimates