And then I *was like*, "No way!": A variationist study of *be like* in young Cape Town speech

Elizabeth Horn

University of Cape Town

Supervisor: Professor Rajend Mesthrie

Author Note

The financial assistance of the National Research Foundation (NRF), through the South Africa Research Chairs Initiative (SARChI) and Professor Rajend Mesthrie's Project on Language, Migration & Social Change (Grant no. 64805), towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

This dissertation is presented for the completion of the Degree of Master of Arts in the School for African and Gender studies, Anthropology and Linguistics.

Abstract

Over the last two decades, the adoption of the quotative *be like* has emerged as one of the prominent examples of ongoing historical changes in English. This has been documented internationally, but the degree to which this change has taken place in South African English remains understudied. This dissertation conducts an apparent-time study of quotative systems in Cape Town. In this study, 1791 quotatives, collected from 64 sociolinguistic interviews, are analysed to assess the use of *be like* in contrast with older quotative verbs, as well as how the use of different quotatives is constrained by local social and linguistic variables. The judgement sample consists of young participants of four races, White, Coloured, Indian and Black, and a 'control group' of older White speakers who represent the older norms. There is a focus on race in order to assess how social networks, and therefore language practices, may be deracialising, compared to the relative rigidities of a generation ago. Similarly, there is an emphasis on schooling in order to determine whether former model-C schools may be facilitating language practice change. A combination of statistical analyses in R, including ctrees and logistic regression, is used to determine the degree to which different social and linguistic variables influence quotative choice within the database. The results indicate that there is broad and prolific use of *be like* among the younger generation. Black and White participants use *be like* in the highest numbers, indicating that they may be leading the change. Similarly, speakers from former model-C schools use *be like* more than speakers from other schools, and best exhibit the *content of quote* and *tense and temporal reference* constraints on *be like*. Among them, there is a positive correlation between the use of *be like* and the expression of quotes containing internal monologue, as well as exclamations and non-lexicalised sounds. *Be like* is also positively correlated with the use of historical present tense.

Acknowledgements

I would like to express my profound gratitude to my supervisor, Professor Rajend Mesthrie. Without his valuable support and guidance, this dissertation would never have been written. I gratefully acknowledge, too, the financial support offered by the SARChI NRF grantholder-linked bursary (Grant no. 64805, Language, Migration & Social Change) and the De Villiers Smuts scholarship.

Thank you very much to Dr Daniel Ezra Johnson, Dr Alida Chevalier and Professor Thomas Hoffmann, as well as the SAMWOP 2019 attendees for hearing my queries, offering their help and providing feedback on using R for linguistics. I consider myself fortunate to have been able to receive advice from each of you.

The years spent writing this dissertation were full of unexpected events, from chronic illness diagnoses to a global pandemic. I would like to thank the doctors and surgeons who have aided my recovery. I am particularly indebted to Professor Siebert for being one of the few surgeons in South Africa who perform excision surgery for endometriosis and Dr Jessica Stanbridge for her compassion and encouragement in helping me find the correct psychiatric medication. I couldn't have finished this if not for the relief I found from chronic pain after our consultations.

Finally, thank you dearly to both my parents and my beloved partner, Nick. You have been there for me every step of the way and I am so excited to share my life with you. Thank you to Joan and Claire for your encouragement, as well as Fatima, the loveliest colleague and housemate anyone could wish for.

Plagiarism Declaration

1. I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.
2. I have used the Harvard convention for citation and referencing. Each contribution to, and quotation in, this dissertation from the work(s) of other people has been attributed, and has been cited and referenced. Any section taken from an internet source has been referenced to that source.
3. This dissertation is my own work, and is in my own words (except where I have attributed it to others).
4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.
5. I acknowledge that copying someone else's paper, or part of it, is wrong, and declare that this is my own work.

Signature Signed by candidate _____

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AAVE	African American Vernacular English
ASAE	Afrikaans South African English
BSAE	Black South African English
CSAE	Coloured South African English
FDET	Former Department of Education and Training
FMC	Former Model-C
HHT	Hypothetical, Habitual and Timeless
HP	Historical Present
ISAE	Indian South African English
SAE	South African English
WSAE	White South African English

1. Introduction

The quotative *be like* has, over the last three decades, risen to prominence within English quotative systems and been studied extensively in various parts of the world. The adoption of *be like* is considered one of the current, ongoing, global English language changes. In an effort to contribute to the documentation of *be like* in South Africa, this dissertation presents an analysis of the quotative systems of young speakers from Cape Town, with a particular focus on the use of *be like*, to better understand the ways that local speakers are participating in a global change. Made possible by Prof. Rajend Mesthrie's Database of South African Englishes (in progress), this dissertation additionally aspires to contribute to his research on crossover varieties of English and the deracialisation of South African Englishes.

1.1 Setting

This dissertation takes its understanding of GLOBALISATION from Blommaert (2010) who defines it as multiple processes taking place simultaneously with various speeds, scales and impacts. He further distinguishes between geopolitical and geocultural globalisation, regarding geopolitical globalisation as the older process which has been unfolding over many centuries and geocultural globalisation as a newer development resulting from "the emergence of new communication technologies, increasing and intensified global capitalist processes of accumulation and division of labour, and increased and intensified global inequalities resulting in new migration flows" (Blommaert, 2010:13-14). He sees super-diverse metropolitan areas being established as a result of geocultural globalisation and this thesis would argue that the proliferation of *be like*, too seems to be a result of the more recent geocultural globalisation

In Cape Town, the oldest city in South Africa, originally established by colonists as a refreshment station and now an international tourist destination, geocultural globalisation has occurred alongside the development of democracy as apartheid and sanctions ended, bringing South African speakers into greater contact with both one another and the international world. However, as noted by Mesthrie (2010), drastic economic inequalities have persisted within Cape Town, as with the rest of the country and many working class people of colour have yet to experience the relief and opportunities that all hoped would result from the end of apartheid. Like most South African cities and towns, racial segregation, though not enforced by law, can be seen in many suburbs and townships of Cape Town, entrenched by the apartheid architecture and town planning that have divided the peninsula city. Terms such as 'economic apartheid'

(Mesthrie, 2010) and 'post-apartheid apartheid' (used throughout the Fees Must Fall protests) have emerged as popular ways of describing the persistent inequalities.

Inevitably, linguistic practice has been shaped by such divided social conditions. This dissertation focuses specifically on varieties of spoken English within South Africa. Mesthrie (2010) summarises the way the political technology of apartheid resulted in the formation of distinct, racialised South African Englishes, each with their own characteristic accents and syntax. These were White South African English (WSAE), Coloured South African English (CSAE), Indian South African English (ISAE), Black South African English (BSAE) and Afrikaans South African English (ASAE). WSAE was spoken primarily by first language speakers, CSAE and ISAE were spoken by both first and second language speakers, whereas BSAE and ASAE were spoken primarily by second language speakers. DERACIALISATION refers to the process in which, as people of different races have begun to interact more, South African Englishes have become decreasingly delimited by race. Mesthrie's further work has documented the way that English dialects have become deracialised (Mesthrie, 2010) and crossover varieties (Mesthrie, 2017) have developed, particularly among young middle-class speakers, as social networks have become more racially diverse.

1.2 Theoretical Framework

The race labels, Black, Coloured, White and Indian will be referenced throughout this dissertation. These distinctions were established during apartheid and have received warranted criticism. South African ethnicities, cultures, heritages and ancestries are much more complex than the four terms imply. However, the decision has been made to use them here due to their use in the original descriptions of ethnic Englishes and the interest of this dissertation in the deracialisation of South African English (SAE). Race, as discussed in this project, is understood according to Falguni Sheth's (2009) view of race as a technology, rather than a natural or biological phenomenon in which people are grouped according to phenotype or ancestry. Fundamentally, Sheth views race as a social technology that is produced politically and marked by physical traits and signified through certain behaviours. However, crucially, in Sheth's conceptualisation of race, a group is targeted and outcast by society prior to the assignation of certain physical traits and behaviours as markers of that group. In her own words:

> I understand the concept of "race" to be a mode of division that is instantiated through political and legal institutions, rather than a term that consistently identifies certain stable and coherent populations. (Sheth, 2009:17)

Sheth (2009) argues that race functions not only to distinguish between various social groups, but to manage those groups, shaping the relationships formed between individual citizens, groups and powers such as the state. This certainly seems to apply in South Africa, where the racial categories Black, Indian, White and Coloured were established and enforced by the apartheid state in pursuit of fascism and White supremacy. These categories also take on particular meanings in South Africa, in which Black people were originally understood to be Bantu language speakers of African ancestry, White people to be of 17th-19th European settler ancestry, Indian people to be of 19th century Indian ancestry, and Coloured people were understood to be of indigenous Southern African (including among them both Bantu and Khoesan), European, Indonesian, Malaysian and Madagascan ancestries, who tended to have fewer ties to their older ancestries and heritages, sometimes as a result of their ancestors having been brought to South Africa as enslaved persons (Mesthrie, 2010; 2012). Mesthrie (2012) describes how these categories fail to describe South Africans' lived experiences, but also notes that participants within the Database of South African Englishes "showed little or no discomfort" (2012:375) with the terms and frequently self-identified with them.

> Thus the four official terms are highly problematic as political constructs, because they conceal a constant state of flux: overlapping and criss-crossing social histories, interests, associations, friendships and marriage patterns, which apartheid tried to override by rigidifying social space. (Mesthrie, 2012:374)

Whilst broad social categories such as race, class and gender are discussed in this dissertation, and used for the identification of trends within speech, the intention is not to treat such trends as deterministic or imply that various combinations of social categories reflect or encapsulate individual speakers' identities. Speaker identity is rather understood according to the framework proposed by Bucholz and Hall (2005), in which identity is perceived as a social phenomenon that emerges through and within linguistic and semiotic practices and interactions. Identity is not the focus of this study, although work examining the ways in which individuals use *be like* in the process of linguistically indexing various identities could prove interesting and fruitful for further documenting the ways *be like* is used in South Africa. As the

participants of this study were cisgender men and women, sex terms such as ‘male’ and ‘female’, may occasionally be used as a means of referencing speakers’ genders, in addition to ‘man’ and ‘woman’, during discussions of social variables.

Whilst class is an important factor in Mesthrie’s (2010) work, class is not explicitly discussed in this dissertation. Instead, speakers’ schooling will be used as a proxy for class. Specifically, this dissertation will distinguish between speakers who attended former model-C schools and those who did not. During apartheid, model-C and private schools had higher fees and were only attended by White students. Model-C schools charged lower fees than private schools and were offered a government subsidy near the end of apartheid and began admitting small numbers of Black students. Once apartheid ended, former model-C schools were no longer segregated and numbers of Black, Indian and Coloured students increased, but have continued to be primarily attended by students of middle-class backgrounds and “there is now a major social division between pupils at [ex-]model-C and private versus ‘township’ or DET (the former Department of Education and Training) schools.” (Mesthrie, 2010:6) This dissertation will distinguish between former model-C (FMC) schools on the one hand, and former Department of Education and Training (FDET) schools on the other hand.

Mesthrie’s (2010) work shows that fronted GOOSE vowels, initially used only by WSAE speakers, are now expressed by speakers of all races and use of the feature seems to have become a means of indexing youth and middle-class belonging. Mesthrie (2017) further observes that features of BSAE, specifically the expression of schwa as a full vowel and the neutralisation of vowel length, are no longer present in the English of young, Black speakers and a new collection of crossover varieties has emerged which reflect various degrees of the main features once associated with WSAE. This dissertation wishes to investigate the possibility that *be like* could also form one of these features originally associated with WSAE that has come to index youth and middle-class identities.

As this study aims to establish broadly the degree to which *be like* is present in South African quotative systems, it is conducted in a style similar to many of the social surveys used in what Eckert (2012) has dubbed the first wave of social variationist studies. In order to identify potential trends within the data, this dissertation will investigate whether there are any correlations between linguistic variables and the social variables of age, race, gender and schooling. The replicability offered by such studies allows for the contrasting of Cape Town quotative speech with international research on *be like*.

However, it also draws on later understandings of variation that were developed in the third wave of social variationist studies, which are characterised by a focus on style and the understanding of social meaning as flexible and dynamic, rather than fixed. Third wave variationism understands variation as the process through which variables acquire social meaning, in addition to its being a system that reflects social meaning. The intention of this study is not to identify static meanings for variables, but rather to observe broad trends within Cape Town quotative systems, whilst potentially gleaning some insight into possible sets of related meanings, which Eckert (2012) calls indexical fields, that may be associated with various quotatives.

Also useful is Eckert's description (echoing in some ways Silverstein's (2003) conception of indexical order) of how a variable can acquire social meaning:

> At some initial stage, a population may become salient, and a distinguishing feature of that population's speech may attract attention. Once recognized, that feature can be extracted from its linguistic surroundings and come, on its own, to index membership in that population. It can then be called up in ideological moves with respect to the population, invoking ways of belonging to, or characteristics or stances associated with, that population. (Eckert, 2012:94)

Silverstein's (2003) indexical order describes the way in which what he calls an 'nth order indexical' can acquire an "ethno-metapragmatically driven native interpretation" (Silverstein, 2003:212), thus becoming an 'n+1st order indexical'. In other words, a feature that is correlated with specific semantic functions or speakers expressing certain socio-demographic identities, can become enregistered, acquiring specific implied, inferred and/or relational meanings understood within local ideologies. It is through this process that different registers come into being and variables acquire social meaning.

1.3 Research Aims

Whilst the intrinsic relationship between variation and social meaning may not be the central focus of this dissertation, the aim of the following research is to establish an initial understanding of the social variation of *be like* and other quotatives in Cape Town, South Africa, after which further investigation into the indexical fields of various quotatives can be

made. Additionally, the research aspires to contribute to an understanding of how South African Englishes are changing in a post-apartheid context, how they are impacted by global and local language varieties, and which linguistic resources are available for the expression of linguistic, ethnic and regional identities. The objective of this research is to discover the extent to which *be like* has been adopted into South African speech in Cape Town, among Black, Coloured, White and Indian speakers, and to determine the effects of both social and linguistic variables on quotative choice.

The research questions are as follows:

1. To what extent is *be like* used by young people of various races and genders in Cape Town?
2. What are the variable constraints on *be like* and do they differ between young people of different races or genders?
3. Are former model-C school disseminators and accelerators of the adoption of *be like*?

2. Literature Review

The diffusion of the innovative quotative, *be like*, is a well-known example of ongoing language change in varieties of English across the globe. What follows is a review of the literature that has been published on *be like* in the last 30 years. Specifically, this chapter will explore the context that gave rise to the use of *like* as a quotative, the linguistic and social constraints, and change to narrative styles, that have accompanied its widespread adoption, and the degree to which its use varies across varieties.

2.1 Quotatives

A QUOTATIVE is a grammatical item or construction used to introduce what Tannen (1986) calls 'constructed dialogue'. Constructed dialogue may refer to quoted speech, thoughts or sounds. The term is preferred because speakers don't quote themselves or others verbatim. Rather, they construct dialogue according to their narrative intentions and (unreliable) memories. Tannen (1986) compares the quotes employed in conversation to the dialogue found in novels and plays, saying that the one is no less constructed than the other.

In English, quotatives are usually verbs, such as *say*, *think* and *go*, or the zero quotative which is not pronounced. Tagliamonte and Hudson (1999:148) provide the following examples of these quotatives in their data:

1. a. And she ***said***, "Would you like me to phone?"
   And I ***said***, "Don't do that 'cause Dad'll be furious."

   b. So I ***said***, "Tell him - " I ***said***, "Tell Chris, to tell him, to call me."
   It was goin' through this chain, eh! And aah, Season ***says***. "OK."

2. a. And I was ***thinking***, "Well surely they can all get on."
   I ***thought***, "Right OK." I told them what I was doing.

   b. And I ***thought***, "Well that's not very nice!"
   Yeah, tsk. Anyway, I ***thought***, "Fine."
   And I was kind of crabby after that. I ***thought***. "Nobody tells me that," you know?
   I thought that was just rude.

3. a. And then he looked at me, and smiled and ***went***, "Oh I'll lock us in."
   and ***went***, "{pouck}" and pulled them
   and they just ***went***, "{kuck}" like that.

b. And uh anyway she ***goes***, "They're not my relatives,"
She ***goes***. "I'm related to these people named Michaels."
I ***go***, "Michaels,"
I ***go***, "holy woah!"

4. This bloke trailed her in the front door.
Ø "Josephine, who's this?"
Ø "Oh, he's staying the night at our house."
Ø "When did you meet him?"
Ø "Oh, about half an hour ago in the train station."

Quotative use is constrained by linguistic and social variables. The linguistic constraints relevant to quotatives are *content of quote*, *grammatical person*, *tense and temporal reference*, and *mimesis*.

Content of quote refers to the pragmatic content of the quoted material. That is, whether a quote contains reported speech, thoughts, attitudes, or non-lexicalised sounds (D'Arcy, 2010). For the traditional quotatives, *say* and *think*, this is intuitive. A quote following the verb *say* will contain words that were spoken, whereas a quote following the verb *think* will contain words that were thought. Thus, one is used almost solely for reporting direct speech and the other for reporting internal dialogue. They have little pragmatic contribution beyond this, resulting in a versatility which likely contributes to the high frequency of their use in contrast with more descriptive quotative verbs like *exclaim* or *declare* (Tagliamonte & Hudson, 1999). The *grammatical person* constraint concerns the grammatical person (and sometimes the number) of the subject of a quotative clause. This is related to the quotative verb, as certain quotatives, such as *think*, are associated with conveying speakers' internal experiences, pairing frequently with first-person subjects, whilst others, like *say* and *go*, are used more often for the recalling of others' speech and will likely be accompanied by third-person subjects (Tagliamonte & Hudson, 1999). The *tense and temporal reference* of a phrase may also constrain quotative use. Sometimes split into two separate factors (e.g. Kohn & Franz, 2009), this constraint encompasses both the morphological tense and the temporal reference of a quotative phrase. *Mimetic re-enactment* or *mimesis* is a constraint that gained attention later in the grammaticalisation of *be like*. The term refers to "the use of voicing effects" (D'Arcy, 2010:65) in dialogue, rather than the normal speaking voice, which can include "changes in pitch, prosodic structure, intonation and accent" (D'Arcy, 2010:66).

Occasionally considered among newer quotatives, *go* has in fact been used as a quotative verb since at least the 1940s and 1950s (Singler, 2001). It is more restricted than *be like* and has been observed only to be used for direct speech, rather than internal dialogue (Singler, 2001), or for non-lexicalised sounds (Romaine & Lange, 1991). The quotative *be all* arose at a similar time to *be like* and also introduced direct speech or internal dialogue. It seems to have peaked in the USA in the 1990s and subsequently declined in use, so that it is little used today (Singler, 2001; Barbieri, 2005; Kohn & Franz, 2009). Innovative quotatives such as these seem more likely to be constrained by social variables such as speaker gender, age, ethnicity, race and socioeconomic status, as will demonstrated in the following section.

2.2 History of *Like*

The word *like* can be traced to the Old English adjective, *gelic*, which meant "having the form of" (Meehan, 1991:39). Numerous meanings and functions of *like* have subsequently been developed, requiring new dictionary definitions. In the 2008 edition of the *Oxford English Dictionary*, the entries for *like* alone span four pages.

Meehan (1991) discusses three of the different meanings, or 'senses' as she calls them, of *like*, as we use them today. In the 14th century, *like* was used to indicate that something was similar to, or approximately resembled, something else. She dubbed this the 'similar to' sense of *like*. This meaning developed, and *like* was soon used in what Meehan (1991:41) calls the 'as if' sense, which allowed speakers to pose hypothetical scenarios in which *like* scoped over a clause, rather than a noun or number, as in the first definition. In the 1800s, *like* began to be used to introduce examples, in order to demonstrate a speaker's point. This 'for example' sense allowed *like* to scope over prepositional phrases, noun phrases or whole clauses.

Meehan (1991) focuses primarily on the meanings of *like* that she believes led to the use of *like* as a FOCUS marker. Underhill (1988:238) defines focus as "the most significant new information in a sentence". He notes that focus is commonly found at the end of a sentence or in a question, which also happen to be areas in which *like* is frequently found. He provides many examples of *like* as a focus marker, marking new, significant, unusual or stereotyped information. Underhill (1988) also notes the use of *like* as an approximator, and its resultant hedging function. When *like* is used as a hedge or focus marker, it can be removed from a sentence and leave the meaning of the sentence intact. As a result, the function of *like* in a sentence is often ambiguous.

Meehan (1991:38) responds to Underhill's (1988) claims that the use of *like* as a focus marker is already decreasing, saying this function "is still very much alive rather than becoming archaic." She further suggests that the function of *like* as a focus marker developed from the 'similar to', 'as if',' and 'for example' senses of *like*. When used in these senses, the part of the sentence scoped over by *like* usually contains the focus – the new and significant information. Crucially, she shows that, unlike Underhill's (1988) recent examples of *like,* in the sentences she uses to demonstrate the three senses above, *like* cannot be removed without altering the meaning of the sentence. Thus, she suggests that the meaning of *like* has generalised, in a process not unlike the process of GRAMMATICALISATION, so that it has gained the additional function as a focus marker.

Grammaticalisation is the process by which a lexical item gains an increasingly grammatical status, or changes to a grammatical item entirely (Hopper & Traugott, 2003). Meehan (1991) poses the possibility that quotative *be like* developed from the same process of grammaticalisation as focus *like*, because they both function to further emphasise, or introduce, new information in a sentence. She also suggests that both of these functions are related to the 'similar to' sense, as they afford the speaker a degree of vagueness and an ability to retract their assertion that they really were emphasising something or quoting someone verbatim.

The use of *be like* as a quotative was first noted in Butters's (1982) editor's note in *American Speech*. It was thought at first to be restricted to American English, but *be like* was soon found in Canadian and British English too (Blyth et al., 1990; Romaine & Lange, 1991; Tagliamonte & Hudson, 1999). Later research revealed that speakers born in the 1960s were the likely innovators of quotative *be like,* but it was speakers born in the 1970s that popularised its use (Tagliamonte et al., 2016). Tagliamonte and D'Arcy (2007) speculate that the particular social group (and/or style) responsible for accelerating this linguistic change was the Californian 'Valley Girls'. Tagliamonte and D'Arcy (2007:212) elaborate:

> As part of the "preppie"[1] movement of the 1980s, *be like* gained prestige as a trendy and socially desirable way to voice a speaker's inner experience. In

[1] The term 'preppie' comes from 'preparatory school'.

> other words, linguistic change begins with a hospitable grammar environment, but requires a social force to drive it forward.

Be like was also found to be used primarily by students and teenagers, and was soon identified as a potentially global phenomenon due to the increasing globalisation of English (Tagliamonte & Hudson, 1999). This was facilitated by a widespread change in English narrative style. Tagliamonte et al. (2016) explain that the appropriation of *be like* as quotative was likely part of a larger change that has been occurring within the direct quotation systems of spoken vernacular English over the last 150 years. This change entails the shift from only reporting speech to the increasing inclusion of "quoted self-revelation" (Tagliamonte et al., 2016:826).

The adoption of *be like* into English quotative systems has been studied across the world. Research has been published on data from Canada (Tagliamonte & Hudson, 1999; Tagliamonte & D'Arcy, 2004; Tagliamonte et al., 2016), the UK (Tagliamonte & Hudson, 1999; D'Arcy, 2004), the USA (Barbieri, 2009; Barbieri, 2007; Singler, 2001; Kohn & Franz, 2009; Cukor-Avila, 2012; Cukor-Avila, 2002), Jamaica (Bogetić, 2014), Australia (Louro, 2013; Tagliamonte et al., 2016), New Zealand (D'Arcy, 2010; Tagliamonte et al., 2016), India (Davydova, 2019), Germany (Davydova, 2019) and Nigeria, (Titilola, 2017).

It is a linguistic change that may be spreading further, into the quotative systems of other languages. A calqued form of *be like*, quotative *comme* or *être comme* has been adopted in Canadian French (Sankoff et al., 1997; Levey et al., 2013), possibly due to language contact with English. *Like* is also not the only discourse marker to have been used as a quotative. Kohn and Franz (2009:262) note that the process by which discourse markers become quotatives has been documented in several other languages (Schourup 1982 in Kohn & Franz, 2009) including French (Fleischman & Yaguello, 2004 in Kohn & Franz, 2009) and German (Golato, 2000 in Kohn & Franz, 2009).

2.3 Initial Grammaticalisation of the *Be Like* Quotative

The grammaticalisation of *be like* for use as a quotative was a key focus in the early literature on *be like* (Blyth et al., 1990; Meehan, 1991; Romaine & Lange, 1991; Ferrara & Bell, 1995; Tagliamonte & Hudson, 1999). It was initially thought to be an opportunity to study

grammaticalisation in progress. This section focuses on research published in the 1990s and early 2000s.

The progress of *be like* has been studied alongside both internal linguistic variables and external social variables, which have constrained the use of the quotative. Initially, the primary constraints were found to be the linguistic constraints, *content of quote*, *grammatical person*, and *tense and temporal reference*. Early stages of the grammaticalisation of *be like* were marked by its use with first-person subjects, its use for dramatic effect, its use to quote internal dialogue and non-lexicalised sounds, as well as the greater use of the quotative by female speakers than male speakers (Ferrara & Bell, 1995; Blyth et al., 1990; Tagliamonte & Hudson, 1999). The next stages of grammaticalisation were predicted to contain semantic bleaching as well as an expansion of distribution. Semantic bleaching would entail *be like* being used to quote direct speech, not just thoughts and sounds, and being used with third-person subjects. An expansion of distribution would entail *be like* coming to be used by male speakers as often as female speakers (Ferrara & Bell, 1995; Tagliamonte & Hudson, 1999).

Tagliamonte and Hudson (1999) make use of Hopper's (1991) three principles of grammaticalisation to track the adaptation of *be like* into a quotative complementiser. These principles are LAYERING, a period of coexistence of the multiple functions of *like* without clear "form/function correspondence"; SPECIALISATION, the process *like* undergoes in order to be used as quotative complementiser; and PERSISTENCE, which refers to the tendency of a new form to retain traces of its "original referential meaning" (Tagliamonte & Hudson, 1990: 149-150). There was, at first, a tendency for *be like* to be used with the first-person and for quoting internal dialogue (Ferrara & Bell 1995). However, its ability to "introduce both kinds of speech… an attitude reaction, or thought, as well as something actually said" (Blyth et al., 1990:215) was also reported from the beginning. This distinguishes *be like* from more traditional quotatives like *say* and *think*, which have a semantic relation to the *content of quote*. The use of *be like* to quote a wider variety of constructed dialogue may be related to its origin. Meehan (1991:47) posits that this origin was from the 'as if' meaning of *like*, as "it functions as a quasi-complementiser in that its scope already encompasses an entire clause." When used according to either the 'as if' or 'for example' sense, *like* introduces a clause that is an example or demonstration of the point that is being made. The use of quotative *like* could be seen as an example or demonstration of what a person said, thought, or felt, at a certain time. This also demonstrates Hopper's (1991 in Tagliamonte & Hudson, 1999) *persistence* of older meanings of *like* in its innovative uses.

From its early use as a quotative, *be like* has occurred most frequently with first-person subjects. Then, as there was increase in the use of *be like*, it began to be used with third-person subjects increasingly often. However, the predicted semantic bleaching and expansion (Tagliamonte & Hudson, 1999; Ferrara & Bell, 1995) did not occur as expected. Whilst the percentage of third-person subjects increased, *be like* did not come to be used predominantly with third-person subjects. Rather, *be like* continued to be favoured for use with first-person subjects (Buchstaller and D'Arcy, 2009:296). As will be discussed later, Tagliamonte and D'Arcy speculate that this was due to a shift in narrative style that emphasised the recounting of personal inner dialogue which created a niche for *be like* to fill so that it was retained in quotative systems.

There has been an expansion of distribution of *be like*, most obviously demonstrated by the many countries it has been studied in. However, when it comes to speaker gender in particular, the evidence does not depict a simple expansion from predominantly female to male speakers. As noted by Tagliamonte and Hudson (1999) there is no consensus when it comes to gender and *be like*. Blyth et al. (1990) and Daily-O'Cain (2000) find *be like* is used more often by men, whereas Romaine and Lange (1991) and Singler (2001) find it is used more often by women. Ferrara and Bell (1995) found it was used more often by women initially, but men soon caught up and neutralised the gender effect. Tagliamonte and Hudson (1999:167) review the variation in their data and published research, and propose the following:

> [T]he more diffused *be like* is, the ***more likely*** it is to differentiate male and female, at least at an early phase of development of which these corpora appear to representative.

Others proposed different reasons for speaker gender variation. Barbieri (2007) found that *be like* was the preferred quotative of use by women between 16 and 26 and men between 27 and 40 and postulated this may be because older men who aspired to connect with younger women began to imitate their usage of *be like*. Singler (2001), on the other hand, took into consideration interviewer gender. He found that when a male interviewer was paired with a male interviewee *be like* would be strongly disfavoured, when a female interviewer and female interviewee were paired with one another, *be like* was strongly preferred, and when the interviewer and interviewee were of different genders, the quotative would be weakly disfavoured. This suggests there could be a host of factors influencing speaker gender, which

would account for the wide variety of results when *be like* is assessed according to this constraint.

As with sociolinguistic variants generally, speakers themselves appear to be poor judges of the distribution of *be like* according to gender. Blyth et al.'s (1990:223-224) attitudinal survey shows *be like* and *go* as stigmatised and ungrammatical with *go* associated with "uneducated, lower-class males" and *be like* with "middle-class teenage girls". The attitudinal survey found that *be like* was perceived as female but used more often by males.

Later research, such as Tagliamonte and D'Arcy (2007) and D'Arcy (2010), emphasised the significance of the *tense and temporal reference constraint*, but it was already a topic of interest for some of the early researchers. Romaine and Lange (1991) noticed *be like* tended to occur in present tense and Blythe et al. also found tense to be a significant linguistic factor constraining its use. As quotatives such as *be like* and *go* tend to be used for dramatic effect, they often occur in the present tense, which is another strategy employed to make a narrative more captivating to a listener.

This strategy is the use of HISTORICAL PRESENT (HP) tense. HP refers to a specific combination of morphological tense and temporal reference. Whereas simple present tense has present morphology and present temporal reference, and simple past tense has past morphology and temporal reference, HP is the use of "present morphology with past temporal reference" (D'Arcy, 2010:67). As mentioned above, it is commonly employed for dramatic effect.

D'Arcy (2010:68) provides the following examples of HP and *be like*:

5. a. She got me home and it***'s like*** 'So, what'd you think?'
   It***'s like*** 'Oh yeah.'

   b. I just seen her eh and then she looks at me and starts pointing at me and I***'m like***
   'Yeah? What?'
   And she***'s like*** 'I know you, eh?'

Blyth et al. (1990) also remarked that *be like* tended to occur in phrases that were in the continuous aspect, despite the fact that a quote contains words that were spoken in the past and therefore represent a completed action. They speculate that this was perhaps due to the ability of *be like* to introduce either something that was said, in a speech act, or some form of internal dialogue that expresses a speaker's emotional state.

Like the continuous aspect, there have been other speculations about potential linguistic constraints on the use of *be like* as a quotative that have not been further explored in the literature. For example, Tagliamonte and D'Arcy (2004) recount Romaine and Lange's (1991) suggestion that *be like* was potentially used to introduce hypothetical dialogue, and Ferrara and Bell's (1995) suggestion that it introduces "the collective thoughts of a group and habitual thoughts and speech" (Tagliamonte & D'Arcy 2004:499).

Age was identified as a significant social constraint from the beginning, as *be like* was a linguistic change led by students and adolescents. During the 1990s, the use of *be like* was reported to decline suddenly after the age of 25 (Blythe et al., 1990), but it was soon found to be used by older speakers too. By the 2000s, in the USA, *be like* was reported to be used frequently by speakers in their early 30s (Singler, 2001).

There are more social variables than age and gender, such as race, ethnicity and class, but the gender and age variables are the primary forces within the literature. As Buchstaller and D'Arcy (2009:295) note, the socioeconomic status (SES) of speakers has received considerably less attention than speakers' gender in studies of *be like*. Blythe et al. (1990:216) offer what was likely a common motivation for a common practice, saying, "we did not control for class or education because our hypothesis did not involve observations based on social class and the amount of time available to us was restricted." In one of the few studies that consider speaker socioeconomic status, Buchstaller (2006a) found that *be like* was used more by American speakers with a lower socioeconomic status, than by those with a higher socioeconomic status, whereas there was not a noteworthy difference in the use of *be like* between speakers with varying socioeconomic status.

There are also few studies that include ethnicity or race as a social variable. Kohn and Franz (2009:265) note that although Ferrara and Bell (1995) and Singler (2001) account for ethnicity, they do not "address the specific constraints that might distinguish African Americans' use of *be like* from White Americans' use." Those who further consider and focus on ethnicity or race include Kohn and Franz (2009) who look at the Latino and African American speaker communities, Cukor-Avila (2002, 2012) and Sanchez and Charity (1999) who look at African American users of *be like*, and D'Arcy (2010) who looks at Maori and Pakeha users of *be like*.

Kohn and Franz (2009:282) find that the Latino and African American communities they studied differed from "previous corpus studies", as *be like* was not significantly

constrained by tense. There was variation in how gender and grammatical person constrained *be like*; these constraints were prone to variation across international, and even general American, speaker populations. Cukor-Avila (2012) discusses the linguistic diffusion of *be like* into Springville. She believes it was a change begun by White teenagers that went on to spread into young urban and rural AAVE speaker communities. She further comments that the data implies that as *be like* was adopted by Springville speakers, they "simultaneously adapted it to the structural norms of AAVE that were already operative in the speech of younger generations in the community." (Cukor-Avila, 2012:635) D'Arcy (2010) also reveals striking differences between the quotative systems of Maori and Pakeha men (discussed in section 2.5).

2.4 Real-time Comparisons

After the initial research on *be like* was published, it became possible to conduct REAL-TIME studies – in which the same community is sampled twice or more over a period of time – to contrast with APPARENT-TIME studies – in which a community is sampled only once, and the behaviour of different generations is compared – and provide an improved understanding of the process of grammaticalisation that the *be like* quotative was undergoing among certain speakers. Two such studies are Tagliamonte and D'Arcy's (2004) comparison to Tagliamonte and Hudson (1999), and Barbieri's (2009) comparison to Barbieri (2007). Other real-time studies are included in the work of Ferrara and Bell (1995), Cukor-Avila (2002) and D'Arcy (2010).

Tagliamonte and D'Arcy's (2004) study shows that there had been a dramatic rise in the use of *be like* quotatives in Canada in the time after Tagliamonte and Hudson's (1999) research. This was most apparent amongst their young participants, for whom *be like* made up 58% of quotatives used, and seemed to take the place of *say*, which had shrunk to make up only 11% of quotatives. *Go* and *think* had also decreased in usage, but the percentage of zero quotatives and 'other' quotatives remained consistent.

Tagliamonte and D'Arcy's (2004) results showed both that *be like* is used most commonly by women and appeared more often alongside first-person subjects than second- or third-person subjects. Tagliamonte and Hudson's (1999) original predictions that *be like* would remain highly correlated with internal dialogue proved statistically significant for all age groups, except amidst the 17-19-year-old age group, who used *be like* most commonly for direct speech, rather than internal monologue and non-lexicalised sounds. The real-time

analysis showed that there was an increased usage of *be like* and female participants appeared to be leading the change. There was also a shift in the *content of quote* constraint and *be like* was beginning to introduce direct speech and non-lexicalised sounds more often, showing that its association with internal dialogue seemed to be lessening.

Barbieri (2009) is contrasted with Barbieri (2007) in a real-time analysis which investigates the social variables, speaker age and speaker gender, and assesses whether *be like* is an example of AGE-GRADING or a change in progress. Barbieri (2009) explains that there were expectations that *be like* was a case of age-grading, rather than fully-fledged linguistic change, encouraged by the negative perception of *be like* in attitudinal surveys which gave the impression that speakers would stop using the feature if they disliked it. Research such as Singler's (2001) study, which records the use of *be like* by speakers in their 30s, brings into question whether age-grading is taking place. Barbieri (2009) finds that by 2005 *be like* is the preferred quotative for all speakers below forty. Usage has increased for women and even more so for men, showing that speakers who were young adult and teen users of *be like* in the 1990s increased their usage of the form over the last decade.

Whilst age-grading can be assessed using apparent-time studies, real-time studies are considered more reliable (Tagliamonte & D'Arcy 2007). Real-time studies can be conducted using existing evidence or by using re-surveys. Re-surveys can be either TREND SURVEYS, which randomly sample a previously studied community, or PANEL STUDIES, in which the same participants are re-interviewed over a set time period (Labov, 1994 in Barbieri, 2009). Barbieri (2009:72) performed a "surrogate trend study" by comparing two corpora, the American Conversation corpus of the *Longman Spoken and Written English*, gathered by Biber et al. (1999) in 1995-1996 and used in Barbieri (2007), and a sample from the *Cambridge Corpus of Spoken and North American English*, gathered in 2004-2005, that shared "similar design characteristics."

Barbieri (2009) found that, although *be like* is the preferred quotative for teenage speakers of both genders, it is used most often by girls. This is consistent with Tagliamonte and D'Arcy's (2004) suggestion that the gender effect becomes more exaggerated amongst those who use the quotative most frequently. Barbieri (2009) cites Chambers (1992) and Labov (1990) and suggests that it is also consistent with the understanding that women often advance linguistic change. The real-time comparison showed that women over 40 maintained their higher use of *be like* (as found in the earlier study), whilst men older than 40 had decreased

their use of the new quotative, which Barbieri argued shows older women play a role in maintaining and advancing changes to the quotative system.

2.5 Interaction with Local Quotative Systems

Like other features from 'western' and American culture, such as kindergarten, Christianity and fast-food chains, the rapid spread of *be like* has been enabled by globalisation. But as Buchstaller and D'Arcy (2009:283) write, the process of globalisation is not simply a process of Americanisation and does not yield uniform results. Each community that adopts such practices and commodities reinterprets them and assigns them new functions and meanings according to their local worldviews and needs. The same is true for linguistic changes that originated in the USA. When *t*-flapping, or North American lexical items have spread cross-varietally in the past, they have not replaced local alternatives, rather the local systems have shifted to accommodate them as additional variants (Meyerhoff & Niedzielski, 2003 in Buchstaller & D'Arcy, 2009). This also seems to be the case for the adoption of *be like*.

Buchstaller and D'Arcy (2009:317) suggest that as *be like* spreads cross-varietally from American English, its linguistic constraints were also transferred, but their relative strengths and directions of effect differed across communities. They found that constraints on *content of quote*, *grammatical person*, and *mimetic re-enactment* were not dramatically reinterpreted, whereas *tense and temporal reference* was more likely to be lost when *be like* was transferred from one variety to another. Social constraints, on the other hand, were not consistently transmitted across varieties. Buchstaller and D'Arcy (2009) ascribe the cross-varietal variation of constraints to two causes: the specific pressure of local language systems; and the degree of contact with, and knowledge transfer from, the donor variety.

D'Arcy (2010:62-63) also remarks on the effect of system-specific pressures on *be like*, saying "[*B*]*e like* does not impose itself on an existent system but rather is integrated in ways which cohere with the system's prevailing, and pre-existing, overall configuration". This is evidenced by her study of the use of *be like* by Maori and Pakeha men, in their early 20s, in Aotearoa[2]. The difference in the ways that Maori and Pakeha men use *be like* and zero

[2] Aotearoa is the Maori name for New Zealand

quotatives demonstrates the way in which pre-existing quotation patterns influence the adoption of innovations such as *be like*.

D'Arcy (2010) found that Pakeha English and Maori English speakers shared *be like, say, think, go,* and zero quotatives as their main quotatives (all occurred over ten times in each corpus). *Be like* was the most popular quotative in both groups, and *say, think,* and *go* appeared in similar quantities in both groups. However, Pakeha speakers were found to use *be like* significantly more than Maori speakers, whereas Maori speakers used the zero quotative significantly more. Maori and Pakeha speakers use the zero quotative in different ways, as well as in different quantities. Maori men tend to use the zero quotative in consecutive clusters of dialogue that they are re-enacting (implying the use of *mimetic effect*). Pakeha men, on the other hand, tend to use zero quotatives after other lexical verbs that imply quotation. D'Arcy (2010:78) provides examples of Maori speakers' use of the zero quotative in 6, and Pakeha speakers' less frequent use of the zero quotative in 7, below:

6. a. She ***goes*** 'Oh well you get a bit of flour you put it in you put a bit of you put a bit of sugar in and then you put a bit of water in.'
   Ø 'How much?'
   Ø '<drawls> Oh <drawls> as much as you need.'
   Ø 'Sorry nana I sort of I don't understand those sort of instructions.'

   b. But fresh as, **Ø** 'No kids?'
   They ***were like*** 'You smoke?'
   Ø 'Nah.'
   Ø 'You drink?'
   Ø 'Yeah, socially.'
   Ø 'Any kids?'
   Ø 'Nah.'
   Ø 'That's good that's good, don't mind drinker, I drink too.'

   c. I fucken started off rambling all these names like, **Ø** 'Tania?'
   Ø 'Nah.'
   Ø 'Oh fuck, I don't know.'
   She***'s like*** 'Nicole.'

7. a. I was ***going*** ‘Mate let’s go. I’ve got no time for pain.’
He was ***going*** ‘Yeah mate, I ain’t got time for pain.’
Ø ‘Yeah, me too bro.’

b. I did ask what that person thought I spoke. She ***said*** ‘New Zealandish.’
Ø ‘Okay [inc] that’s English.’
She ***said*** ‘Okayish.’

c. She’ll ***go*** ‘Oh I’m Matthew’s mum.’
And they’ll **Ø** ‘Oh okay.’

The use of HP also varies among Maori and Pakeha English speakers. In D’Arcy (2010), Maori men disfavour the use of HP with *be like* and rather favour *go* with HP, but Pakeha men favour the use of HP with *be like*.

D’Arcy’s (2010) incorporates into her study a real-time comparison between quotative systems in Maori and Pakeha English which shows that their quotation patterns pre-date the introduction of *be like*. The graphs showing the real-time comparisons of Maori and Pakeha English reveal that whilst both varieties have experienced a decrease in the use of *say* and *think*, as well as a dramatic increase in the use of *be like*, there has been a decrease in the use of *go* and an increase in the use of zero quotatives in Pakeha English, whereas the zero quotative has decreased and *go* has increased in Maori English. As D’Arcy (2010) suggests, this indicates that the differences between quotative systems pre-date the arrival of *be like*, and whilst *be like* has altered the local Aotearoa quotative systems, it has also been adopted differently according to the local varieties of English.

D’Arcy (2010:69) suggests that these differences could be considered evidence of forms taking on “social significance as markers of ethnic identity in the New Zealand context.” However, she emphasises that in doing so, she is not trying to establish the two varieties as separate ethnolects. She opts rather to use Eckert’s (2008) term, ETHNOLINGUISTIC REPERTOIRE, which acknowledges the fluidity and variation the term ethnolect does not. D’Arcy (2010) makes it clear that no forms are exclusive to either Maori or Pakeha English, rather the two varieties share a repertoire of features. As both groups of speakers are in daily contact with each other, they construct their individual ethnolinguistic repertoires both in conjunction, as well as in opposition, with one another. Therefore, there are bound to be overlapping, as well as distinctive, features in both repertoires. Speakers can make use of certain forms, either consciously or subconsciously, which may serve to distinguish their ethnic identities.

Like D'Arcy (2010), Kohn and Franz (2009) conducted a community-based study that focused on the ethnicity or race of participants in order to shed light on the way a global form interacts with the systems of local identities and language varieties. They emphasise the importance of community-based studies in challenging and disrupting "the illusion of homogenous national norms" (D'Arcy, 2004 in Kohn & Franz, 2009: 260). Kohn and Franz (2009) compared a Latino community in Durham, North Carolina and an African American community in Hickory, North Carolina and found their results did not vary greatly from global findings. The differences in speaker gender, as well as the relatively low frequency of zero quotatives are both factors that have varied across communities. Their finding that *be like* continued to be more likely to introduce internal dialogue than direct speech, although it was still a popular choice for quoting direct speech, was also common at the time.

Kohn and Franz (2009) also provide an example of how quotatives form a part of individuals' repertoires and can be used to index different identities. They discuss a speaker of Latino heritage, who takes great pride in this heritage and is immersed in a Latino community. This speaker code-switches often, despite speaking limited Spanish, and they speculate that he may use the quotative *say* more frequently "to index his ethnic identity by sharing a speech pattern associated with those still transitioning from Spanish to English." (Kohn & Franz, 2009:283)

Whilst both Kohn and Franz (2009) and D'Arcy (2010) include multilingual participants in their studies, Davydova (2019) looks explicitly at quotatives in second language varieties of English. She uses the Concentric Circles Model from World Englishes studies in her comparison of quotatives in German and Indian Englishes. Davydova (2019) finds that there is more innovation in Indian English (the Outer Circle) than in German English (the Expanding Circle) because, due to colonisation, India has a greater history of the use of English in informal and sometimes domestic interactions, and thus, for some Indian speakers it is not a second language variety. Whereas German English has had limited local use because it is regarded as prestigious resource for communicating with outsiders, rather than amidst the population.

The Indian English innovations of particular interest to Davydova (2019:11) are "verb + *that/ki* and *okay(fine)*" they are quotatives that are popular enough to compete with global quotative innovations such as *be like*. *Ki* is a complementiser in Hindi, equivalent to English clausal complement *that*. Davydova (2019:6) provides the following examples of verb + *okay* (8), verb + *that* (9) and verb + *ki* (10):

8. a. I was ***okay***, 'I saved Pee!'

   b. So after some time it ***becomes okay***, 'He is like that, so even if you don't find it it doesn't matter.'

9. a. I was ***feeling*** that, 'No! Light feeling in my heart that still I have told my feeling to her.'

   b. And we were ***laughing*** on the point ***that***, 'Why she is laughing? Why he is laughing?'

10. a. But my parents ***made*** it very clearly ***ki***, 'First you're your studies!'

    b. My father ***asked*** me ***ki***, 'Now, you have to go out.'

These local innovations form part of "a [quotative] system of structured variation" (Davydova, 2019:588) which include traditional quotatives, like *think* and *say*, as well as global and local innovations. The quotatives within Indian English – which, unlike German English, do not strive to mimic North American English – can also be seen as one of the linguistic resources that make up ethnolinguistic repertoires. Davydova (2019:588) explains that "Indian speakers view the varieties of British, American and Indian English as distinct, yet self-contained, linguistic resources upon which they draw to index specific social meanings."

A community's or individual's narrative style will also influence the quotatives they tend to use. Tagliamonte and Hudson (1999) compare British and Canadian narratives and find that British English speakers have a more introspective narrative style and include the speaker's thoughts more often than Canadian English speakers, who prefer an action- and dialogue-oriented narrative style which makes use of more HP and zero quotatives. D'Arcy (2010) also understands the differences in Maori and Pakeha quotative systems to be part of a larger difference in narrative styles. She explains that Pakeha men rely on an "overt lexical scaffolding" (D'Arcy, 2010:62) to communicate 'speakerhood', whilst Maori men use voicing effects, or *mimetic re-enactment*, to convey the same information.

2.6 Further Grammaticalisation of the *Be Like* Quotative with HP and Mimetic Re-enactment

It is these two variables, HP and mimetic re-enactment, also associated with narrative styles, that came to be significant constraints later in the development of quotative *be like*.

Both HP and mimetic re-enactment were noticed in the early research on *be like*, although they only gained greater significance further into the 2000s. Buchstaller and D'Arcy (2009) note that both Singler (2001) and Romaine and Lange (1999) commented on the tendency of *be like* to introduce expressive quotes. Furthermore, as noted earlier, Blyth et al. (1990) and Romaine and Lange (1991) observed a correlation between the use of *be like* and quotative clauses in present tense. Within a decade, Singler (2001) noted that most of the present tense quotes in his data had a past time reference (D'Arcy, 2010).

Tagliamonte and D'Arcy's (2007) apparent-time study of the Toronto speech community suggests a potential explanation for the association of *be like* with HP. Looking at how different generations use *be like*, they proposed that it was initially associated with first-person subjects, present tense, internal dialogue, as the expression of speaker thought processing offered *be like* a niche it could occupy within the grammar. However, with time, the effects of these constraints weakened. Both the quotative *be like* and the expression of internal dialogue in narratives gained popularity. Tagliamonte and D'Arcy (2007:211-212) speculate that the use of HP may have risen as it bypasses the use of the potentially "complicating action clauses" of past morphology. As a result, within time, quotative *be like* became associated with HP.

However, the relationship between *be like* and HP (as well as present tense more generally) is not consistent across speech communities. Whilst it is favoured with HP in many communities, like in Canadian, American and New Zealand Englishes (Tagliamonte & D'Arcy 2007; Buchstaller & D'Arcy 2009), it is disfavoured by some, such as Buchstaller and D'Arcy's (2009) study of British English. Similarly, D'Arcy's (2010) study of the quotatives used by Maori and Pakeha young men in Aotearoa shows that although *be like* is associated with HP in Pakeha English, in Maori English *go* is associated with HP and *be like* is "strongly disfavoured in HP contexts" (D'Arcy, 2010:71).

2.6.1 Specialisation with dummy it.

The use of existential or dummy *it* in quotative phrases has been noted to occur solely with quotative *be like*. It has further been proposed that this construction is particularly likely to occur with mimetic re-enactment. Fox and Robles (2010:716) provide the following example in which a speaker, Felicia, adopts a specific tone as she enacts her own self-talk to her friends.

11. And that's what keeps me going. Y'know every time I, I get together, it ***'s like***, "Okay, Felicia, you can do it."

Singler (2001) observes that dummy *it*, when used as the subject of a quotative phrase, requires *be like* as its quotative. He speculates that "*it's like* evolved from *it's as if X said*, where X ordinarily has an indefinite referent […] but can have a specific referent" (Singler, 2001:260-261). He further suggests that this construction may have been important in the grammaticalisation of *be like* into a quotative.

Fox and Robles (2010) cite Streeck's (2002) observation that it was becoming increasingly popular for speakers to use mimetic re-enactment rather than descriptions of what they were thinking or feeling, especially in the United States. They also note Carbaugh's (1988) observation that 'self-revelation' was gaining popularity with United States speakers. Fox and Robles (2010) remark that *be like* is perfectly suited to this more performance-oriented narrative style. They further propose that mimetic re-enactment has become associated specifically the combination of dummy *it* and quotative *be like*.

2.7 The Nature of Linguistic Change

Multiple theories of linguistic change have been used in attempts to understand the widespread adoption of *be like*. Models, such as the wave, gravity and cascade models, have been used in attempts to elucidate how features spread across communities. Theories such as Labov's (1994, 2001) generational change and communal change, and Sankoff's (2005) lifestyle change attempt to explain how linguistic change takes place within a community.

Labov (2007) defines TRANSMISSION as the transfer of language within a community – or a variety that makes up one branch on language family tree – from one generation to another. Change from within the language system, or 'change from below', occurs in situations of transmission. This kind of change occurs gradually and incrementally over multiple generations.

DIFFUSION, on the other hand, occurs when there is linguistic transfer between two (or more) communities with different varieties. This is a 'change from above' which imports variables from other language systems or communities (Labov, 2007).

The wave model attempts to describe how language change diffuses outward from places of dense communications (Bloomfield, 1933 in Nerbonne, 2010). The gravity (Trudgill,

1974 in Mesthrie et al. 2009) or cascade (Labov, 2003 in Mesthrie et al. 2009) model takes into consideration social geography and claims it is not only communication density that is important. Rather, population size also affects the diffusion of change, as larger centres will experience a change first, before it moves outward to intermediate-sized and, finally, smaller centres.

Whilst these models explain how a linguistic innovation diffuses across communities, Barbieri (2009) describes the different mechanisms of change by which a linguistic innovation can diffuse within a community. That is, Labov's (1994) generational or communal changes and Sankoff's (2005) lifespan change: When a community experiences a GENERATIONAL CHANGE, the frequency of the introduced variable increases for each new generation. Whereas, when a community experiences a COMMUNAL CHANGE, the introduced variable is adopted by the whole community simultaneously. Generational change was considered the usual kind of change, as speakers' grammars change little after stabilising, whereas communal change is associated with lexical change. However, later research (Tagliamonte & D'Arcy, 2007; Barbieri, 2009) shows that the two types of change can co-occur and older adults may continue to be involved in ongoing linguistic change, as has been found to be the case for some studies of quotatives. The same is true for Sankoff's (2005) lifespan change which occurs when speakers continue to advance a community-wide change in progress over their lifetimes.

Tagliamonte and D'Arcy (2007) conduct an apparent-time study that focuses on lifelong members of the Toronto speech community. They find that speakers appear to continue to increase the frequency of their use of *be like* as they age. They support this finding, claiming that there is growing evidence that "individuals do not cease to participate in linguistic change post-adolescence but may indeed contribute to its advancement into adulthood" (Tagliamonte & D'Arcy, 2007:213 and the references within).

Tagliamonte and D'Arcy's (2007) results suggest that speakers continue to participate in change into adulthood as *be like* continues to be used by speakers in their late 30s and by a handful of speakers in their 40s and 50s. In support of their findings, they cite an observation made by Labov (2001:442 in Tagliamonte & D'Arcy, 2007:213) "that adults may not be as stable as previously thought."

Additionally, Tagliamonte and D'Arcy's (2007) results imply that forms of communal and lifespan change may be taking place in Toronto, and the peaks in the usage of *be like* that appear similar to age-grading rather display the typical 'S-curve' (Chen, 1972; 1976) of

linguistic change (Tagliamonte & D'Arcy, 2007:212) when viewed in the context of the whole community. However, the degree to which adults continue to participate in change is brought into question by the finding that, although speakers in their 30s continued to manage the frequency of their use of *be like*, the constraints on their use of *be like* do not develop as younger speakers do. Older speakers continue to advance the frequency of quotative *be like* in their speech but don't appear to advance its underlying grammar.

Cukor-Avila (2012) cites Labov's (2007) claim that diffusion does not lead to permanent structural change among adults and speculates that it is among teenagers and children that *be like* diffuses across varieties and results in structural change. This provides a potential explanation for why younger speakers participate in developments of the grammatical constraints of *be like*, but older speakers do not.

As noted above, Barbieri's (2009) real-time study, to assess whether *be like* was an age-graded change, yielded similar results to Tagliamonte and D'Arcy's (2007) apparent-time study of the Toronto speech community. Barbieri's (2009) results also suggest that speakers continue to participate in linguistic change into their adult years and that generational, communal and lifespan changes may be occurring simultaneously.

By sampling a similar corpus almost a decade after the corpus used in her previous study, Barbieri (2009) found that teenage and young adult speakers in the mid-1990s continued to use *be like* as one of the primary quotatives of choice as they entered their late-20s and early-30s, which suggested that generational change, rather than age-grading was in progress. This increased use of *be like* by speakers below 40 showed that speakers maintained the use of this quotative for longer than Singler (2001) and Buchstaller (2006a) had shown they were doing in the 1990s, when speakers decreased their use of *be like* in their late-20s to mid-30s. The fact that speakers had increased their use of *be like* during this period was also presented as evidence of lifespan and communal change.

Buchstaller and D'Arcy (2009) make use of Milroy's (2007) concepts of *off the shelf* and *under the counter* changes and Meyerhoff and Niedzielski's (2002, 2003) distinction between information transmission and tacit knowledge transmission to understand the diffusion of some of the constraints on *be like* and not others. *Off the shelf* changes do not require much interpersonal contact to occur and are the medium through which information transmission can take place, whereas *under the counter* changes require frequent social interaction and can allow for the transmission of tacit knowledge. Buchstaller and D'Arcy

(2009) propose that the form *be like*, along with its 'universal constraints' – namely, *grammatical person*, *mimetic re-enactment* and *content of quote* – make up the information transmission that occurs with minimal interaction, whilst the social constraints on *be like*, as well as *tense and temporal reference*, make up the tacit knowledge that is only diffused with regular face-to-face interaction. This would explain how teenagers, without being internationally mobile, led the change by adopting *be like* and its main constraints. They also note that the constraints making up the tacit knowledge of *be like* are likely to be functions of narrative style, which requires frequent interpersonal interaction to be transmitted.

With the additional use of Meyerhoff's (2009) taxonomy of transfer (originally intended for language contact scenarios), Buchstaller and D'Arcy (2009) compare the use of *be like* in American, English and New Zealand English, and find *be like* to be an example of weak transfer. That is to say, the same constraints – the linguistic constraints on *be like* – are found to be significant across varieties, but their strength and direction of effect varies. They did, however, identify some trends within the linguistic constraints across varieties. *Be like* was consistently used more often to quote internal dialogue more often than direct speech; it was used more often with a first-person subject than a third-person subject; and was more likely to occur with mimetic re-enactment than not, but tense and temporal reference differed across locales.

2.7.1 A change in narrative style.

Global adoption of the *be like* quotative does not appear to be simply the adoption of an innovative lexical item that replaces traditional quotatives. Rather, *be like* seems to be part of a larger change to English narrative styles. Singler (2001) made the observation in his examination of innovative quotatives "that the change in progress is not merely a replacement of some forms by other forms" (Singler, 2001:264). Whilst *be all* did not gain the popularity *be like* did, he notes that these innovations often function to communicate a speaker's emotional state and attitudes, rather than a speaker's quoted utterance. They do not imply the conveyance of literal speech but possess a degree of ambiguity that allows the listener to interpret whether or not a quotative expresses an utterance or a thought.

Tagliamonte and D'Arcy (2007) pick up on this shift to a performative, dialogic narrative style, with more frequent utilisation of quotation, that has accompanied the diffusion of *be like*. They ask when it was that speakers began "regaling their audience to a running

stream of their own inner thought processes" (Tagliamonte & D'Arcy, 2007:211) and find indications within their data that this trend emerged and drastically increased in the last 65 years. Their apparent-time data finds that octogenarians quote internal dialogue less than 10% of the time, but the percentage of quoted internal dialogue increases with each younger generation until it reaches speakers in their 30s, where it peaks. Tagliamonte and D'Arcy (2007:211) note that this "immediately precedes the acceleration of *be like* among the 30-year-olds" and suggest that, prior to the rise of *be like*, the "stylistic option of inner monologue in narratives of personal experience" gained popularity and created a niche that *be like* shortly filled. From there, its functions could specialise among younger generations.

2.7.2 The role of mass media in change.

Davydova (2019) studies the adoption of *be like* by second language speakers of English in Germany and India. She proposes that the variable grammar of *be like* has a smaller cognitive load than phonological variables do, which allows speakers with "low levels of naturalistic but high levels of mass media exposure" (Davydova, 2019:12) to acquire quotative *be like.* They do so without acquiring all of the constraints, particularly the *content of quote* and *tense and temporal reference* constraints which possess more semantically and pragmatically complex distinctions than *grammatical person* and *mimetic re-enactment*.

Whilst personal communication has been lauded much responsibility for the diffusion of change, mass media has recently emerged as an alternative, or additional, medium for change. The degree to which mass media contributes to change is a contested topic, as is apparent from the *Media and Language Change Debate* (Bell & Sharma, 2014:213-286) Some, such as Trudgill (2014), regard it as highly unlikely that the media could result in the diffusion of grammatical and phonological features. However, he contests the classification of *be like* as a grammatical feature and considers it rather a superficial lexical change that is unlikely to result in any lasting change. Perhaps it is only these surface-level changes that can diffuse through the media.

One of Trudgill's (2014) arguments for the irrelevance of mass media is the fact that speakers do not speak to, and interact with, their television. Sayer (2014), however, proffers two concepts which could provide an explanation for why speakers may acquire innovative features via media. The first is Coupland's (2009) MEDIATION, which refers to the inclusion of non-standard vernaculars within media, and the second is Rubin et al.'s (1985) PARASOCIAL

INTERACTION, which describes the psychological, emotional and creative ways in which audiences experience media content.

Tagliamonte (2014) offers a pertinent reminder that it is not only television, but the internet, that can be considered mass media with the potential to diffuse changes across varieties, and the internet, unlike television, does allow for interaction. However, she reminds us that social media did not exist until the 1990s to act as a possible catalyst to diffuse and accelerate the initial adoption of *be like*. Moreover, the use of *be like* in the media in the 1970s and 1980s was negligible enough to rule it out as a medium for that particular change.

2.7.3 The diffusion of be like *as a black swan event.*

As there were only estimations about the point at which *be like* arose, Tagliamonte et al. (2016) set out to determine, in a large apparent-time study, the moments at which *be like* arose in Canada, Australia and Aotearoa. They aimed to provide the "empirical validation" (Tagliamonte et al., 2016:825) that former speculations about the dates *be like* entered different varieties of English had lacked, by comparing the presence of *be like* in corpora collected in four different cities across the globe: Perth, Christchurch, Victoria and Toronto.

Tagliamonte et al. (2016) found the same results across cities. Speakers born from the 1960s onward showed an increase in the use of *be like*, and an accompanying decrease in the use of the traditional quotative *say*. Speakers born in the 1970s and 1980s showed a dramatic increase in the use of *be like*. The corpora that included speakers born in the 1990s and 2000s (Toronto and Perth) showed that this younger generation use *be like* even more than the others.

Tagliamonte et al. (2016:832) note that, because the data in each of the corpora were collected at different points in time, at first glance it seemed like *be like* was more advanced in the Perth and Victoria corpora, which were created in the early 2010s, whereas the Christchurch and Toronto corpora were created in the early 2000s. With collection times taken into consideration, the results not only showed that the use of *be like* had surpassed that of traditional quotatives within three decades, but that it had done so simultaneously in all four cities.

Tagliamonte et al. (2016:832) used conditional inference trees (ctrees) to try to identify the "points of deviation (either in space or time) in the diffusion of *be like*" which would represent the stages that characterise incremental language change. Whilst they could identify which generation initiated and accelerated the adoption of *be like* (the ctrees returned similar

dates of birth as the first significant split for all four communities: 1970, 1971, 1973 and 1975, for Victoria, Perth, Toronto and Christchurch respectively), they found "no evidence for incremental geographic spread of *be like*." (Tagliamonte et al., 2016:834)

Such a globally simultaneous language change had not been documented before and Tagliamonte et al. (2016) declared the unprecedented change to be an example of a linguistic black swan event.[3] That is, an event that could not have been predicted by any pre-existing models like the wave and gravity models, but was made possible by global developments that have spawned black swan events in other fields within the humanities, like "the internet, social media, September 11, and the market crash of 2008" (Tagliamonte et al., 2016:842). They motivated their classification of the simultaneous emergence of *be like* as a black swan event as follows:

> Language is systematic, constrained, regular, and probabilistic, but is not entirely predictable or deterministic. This makes language, like other complex adaptive systems, susceptible to black swan events. (Tagliamonte et al., 2016:842)

Tagliamonte et al. (2016) suggest the particular global development allowing for this black swan event to occur was the increase in the demand and ease of independent international travel that began to occur in the 1960s and 1970s. They claim this resulted in formation of international and 'supralocal' social networks as travellers from "the United Kingdom, Western Europe, North America, Australia, New Zealand, South Africa and Japan" (Tagliamonte et al. 2016:839) became acquainted with one another in gathering places for tourists, like coffee shops and hostels. Such interaction allowed for Milroy's (2007) *off the shelf* theory of linguistic diffusion, which does not require continuous contact between individuals, to occur. Linguistic items diffusing this way would possess a widespread currency due to their socio-symbolic meaning within supralocal networks.

[3] Coined by statistician Nassim Nicholas Taleb and inspired by a Latin metaphor about birds so rare they were presumed not to exist, a 'black swan event' refers to a difficult to predict event that has a large impact.

Tagliamonte et al. (2016) also found that within their data the same main constraints on *be like* were statistically significant: *tense and temporal reference*, *grammatical person* and *content of quote*. They found *tense and temporal reference* to have the strongest effect and *content of quote* the weakest. This was in contrast with the results of initial inquiries into the grammaticalisation of *be like*, implying that grammaticalisation has continued. They suggest that this potentially provides further support for the link between HP and *be like*, as well as for the notion that there exists an "underlying grammar [which] operates uniformly across corpora, regardless of the demographic makeup of the city or its relative proximity/distance to the putative epicentre of *be like*." (Tagliamonte et al., 2016:837)

They speculate that if they are correct, and *be like* is indeed a linguistic black swan event, there will soon be another to follow. They note that there are already other "increasingly broad-scale changes underway in language, such as the use of *like* as a pragmatic particle (D'Arcy 2007), fronting of the GOOSE vowel (Cox 1999, Labov et al. 2006, Mesthrie 2010, Harrington et al. 2011), high rising terminals (Guy et al. 1986, Britain & Newman 1992, Grabe et al. 2000, Fletcher et al. 2002, Ritchart & Arvanti 2014), /l/ vocalisation (Horvath & Horvath 2001, 2002), and perhaps others" (Tagliamonte et al., 2016:841). However, these changes would require new studies with similarly large, local samples and statistical analyses, as used in Tagliamonte et al. (2016) to establish whether they are also diffusing simultaneously on a global scale with consistent global constraints.

2.8 Conclusion

The quotative *be like* can be used to express either direct speech, internal dialogue or non-lexicalised sounds, but it was originally used most often for the expression of first-person inner dialogue. A degree of semantic bleaching occurred, but in many varieties, it continued to be used most frequently alongside first-person subjects and quotes that express the speaker's thoughts. *Be like* was sustained because it found a niche in changing English narrative styles which preferred a more dramatic and speaker-centric story-telling style, and thus it became correlated with the use of HP and mimetic re-enactment. Whilst these trends are identifiable in multiple communities, constraints on *be like* usage, especially the social constraints, vary across varieties. The linguistic constraints on *be like* appear fairly consistent across varieties, but their strengths and directions of effect vary. The use of *be like* is not age-graded, rather its use appears to be maintained and increased into adulthood. The exact nature of this linguistic

change is unclear, as *be like* appears to have emerged simultaneously in multiple locations across the globe with no evidence of incremental geographic diffusion. This occurrence defies previous understandings of language change and has been suggested as the first example of a linguistic black swan event.

3. Methodology

3.1 Ethics

The data used was collected prior to the commencement of this Masters dissertation and ethics clearance was granted for each interviewer. Each interviewer possesses the consent forms used for their participants, informing them that they could opt out of the study at any point. As the author of this dissertation, I remain in contact with all primary researchers, and they have given me permission to use the data they collected.

3.2 Sample

With permission, Professor Rajend Mesthrie's Database of South African Englishes was used for the purpose of this dissertation. The database consists of sociolinguistic interviews recorded by Professor Mesthrie and other researchers from the University of Cape Town's linguistics department, in different cities across South Africa, including Kimberley, Durban, Johannesburg, Port Elizabeth and Cape Town. This dissertation focuses on interviews recorded in Cape Town with speakers who spent the majority of their lives in Cape Town.

Table 3. 1

Composition of the sample by sex and race and age groups

	Younger (16-35)				Older (55-82)
	Black	White	Coloured	Indian	White
Male	4	10	10	3	2
Female	6	10	10	5	4
Subtotal	10	20	20	8	6

A judgement sample of speakers from Cape Town that would contain participants from the 'Black', 'Coloured', 'White' and 'Indian' racial groups, was selected so that the results of this dissertation may contribute to Rajend Mesthrie's (2010) research on the deracialisation of South African Englishes. Some of the original researchers also collected their data using judgement sampling, making this process easier (Chevalier, 2016). The sample (table 3.1) consists of interviews with 20 young White speakers, 20 young Coloured speakers, 8 young Indian speakers and 10 young Black speakers, recorded over a decade, between 2006 and 2016.

The researchers and main interviewers were Professor Rajend Mesthrie, Dr Tracey Toefy and Dr Alida Chevalier, although a handful of interviews were conducted by students as part of the completion of their undergraduate studies. The younger speakers were between 16 (only one participant this young) and 35 years old, and older speakers were all over 50 years old. The older White speakers served as a proxy control group, to provide a sample of older English quotative systems that were formed prior to the introduction of *be like*.

A more even distribution of speakers across racial groups is desirable, but this dissertation was limited by the interviews available in the database. Interviews with missing data, for example when recordings had been cut short, were excluded. It was necessary to have an audio file to accompany each transcripts, as it wasn't always clear from the transcript alone whether verbs were being used quotatively, and the audio recordings were consulted in these instances, as the speaker's intonation could often clarify uncertainty. When there was uncertainty about whether a verb was being used as a quotative, it was excluded.

All of the quotatives that could be identified in each interview were entered into a database. This meant that some speakers used many quotatives, with one participant producing as many as 119 quotatives during an interview, whilst others used no quotatives at all. This could be due to a variety of different social factors, which may be interesting to look at in future research with a larger database, such as how well the interviewer and interviewee knew one another and the age, gender and race of the interviewer. These interviewer factors were coded and included in the database, but they did not appear significant in initial statistical runs, and thus were excluded in the final results.

Other researchers have also used pre-existing corpora as their data. Barbieri (2009:76-77) compared samples from the *Cambridge Corpus of Spoken North American English* (CAMSNAE) and the American conversation corpus of the *Longman Spoken and Written English* (LSWE). Both corpuses contain a variety of spoken interactions recorded in different settings. CAMSNAE specifically took recordings from multiple states – Oregon, Arizona, Colorado, Minnesota, Michigan, Virginia and Maryland – "so as to represent four major US regional dialects: the Northwest, the Southwest, the Midwest, and the East Coast," and the LSWE contained "set of informants […] selected to represent the range of speakers of the United States across age, sex, ethnicity, level of education, social class, and regional distribution" (Barbieri 2009:76-77). D'Arcy (2010:64) made use of data from the *Canterbury Corpus*, which is made up of "non-structured sociolinguistic interviews that include the reading

of a word list" and is most similar to the database of sociolinguistic interviews used in this project. Tagliamonte et al. (2016) also relied on multiple corpora for their data.

Most data were collected over a shorter time frame than the data used in this dissertation. Similar to D'Arcy (2010:64), for whom "it was not possible to rely on interviews conducted only in 2006" to have sufficient Maori speakers to match the sample of Pakeha speakers, the data analysed in this project consists of interviews conducted over a period of years. Whereas D'Arcy (2010) used interviews carried out between 2005 and 2008, the interviews in the Database of South African Englishes were conducted over the period of a decade. Davydova (2019:5-6) also collected sociolinguistic interviews over a longer period of 8 years. She collected a sample of 80 Hindi speakers between 2007 and 2014, and 97 German speakers between 2013 and 2015.

3.3 The sociolinguistic interview

The interviews conducted by the primary researchers took the form of sociolinguistic interviews. Most interviews were one-on-one interviews with participants the interviewers were already familiar with, or with whom they shared a mutual a friend (using Milroy's (1980) 'friend of a friend method'), whilst other participants were found through university-internal communication. There were also some interviews, including those conducted by third year linguistics students at UCT, in which multiple participants were interviewed at a time.

The interview lengths varied, but most were approximately 45 minutes long and consisted of a word list testing minimal pairs and a narrative section, with some also including a passage to be read aloud. Each researcher made efforts to reduce the formality of the interviews, interviewing participants in their homes when possible and opting to sit on couches, rather than desk chairs, when interviewing them in an office (Mesthrie, 2010; Chevalier, 2016; Toefy, 2014). During the narrative section, which was intended to elicit 'natural speech' and encourage the interviewees to recount stories from their life, the interviewer asked about their memories from school, their working life, their families, childhood games they had played and whether they had any encounters with crime, such as being mugged or burgled. Often the interviewer and interviewee were already acquainted, and they could easily ask about the person's life, as they were already familiar with each other them.

3.4 Data Entry Process and Coding

The interviews were recorded on site, and later transcribed by researchers and student assistants. This section details the author's approach to processing the transcripts and audio files. Additionally, in sections 3.4.2 – 3.4.4 the author shares her notes and decision-making process regarding unexpected aspects of the data, including the ways in which speakers were using *just, like* and 'you know', the influence of style and the possibility of potential outliers.

3.4.1 Approach

Each transcript was manually read, and all quotatives used by the participants identified. The audio files were consulted when the transcribed speech was unclear, or if it was unclear whether a verb was being used as a quotative. In this case, hearing the speaker's tone and potential use of voicing effects known to sometimes accompany quotatives, often helped clarify whether the speaker was recounting constructed dialogue or not (see also Buchstaller & D'Arcy, 2007).

Every instance of constructed dialogue identified in the interviews was entered into the database (in Excel) and coded for both social and linguistic variables. If multiple quotatives were used, the final quotative used before the quoted material was included, rather than the ones before it – as in examples (13) below. If a quotation was interrupted by a second quotative or words such as "like", "just" and "you know", but with no change in speaker or grammatical person, it was counted as one quotative as in examples (12) below. Whereas a change in speaker would signal that another instance of constructed dialogue had begun.

12. a. I **was like**, "Hi," you know, "What's up?"

    c. Like, people **were like**, "Aw! It's so lame," like, "Who would wanna read that?"

    d. People would **say**, "Ooh, blonde!" you know, "Beautiful hair, blonde."

    e. The whole point of your thesis, you not gonna go and **say**, "This is my thesis" to a company and say "Employ me."

    f. I **was like**, "This is not gonna be good," you know, like "This is going to be worse than anything I've ever experienced."

    g. You've gotta go and **say**, "These are my skill sets that I've, I've developed in my thesis in order to do this."

13. a. I took it more personally when he told me, he **said**, "Are you a foreigner? What, what country do you come from?"

b. My doctor told me, he**'s like**, "Unless you change your lifestyle entirely, you won't be able to walk, then."

The variables each quotation was coded for were the response variable (the quotative used) and the explanatory variables made up of social variables of age, school and gender, as well as linguistic variables of *content of quote, tense and temporal reference* and *grammatical person*. Other factors that seemed potentially relevant, such as grammatical number, aspect, 'double quotatives' and code-switching, were also noted during the initial data entry phase, but were later excluded after the first preliminary statistics runs in R, when they either occurred too infrequently or did not seem statistically relevant.

Other exclusions included instances of uncertain or unclear *grammatical person* or incomprehensible speech (following the example of Tagliamonte & Hudson, 1999; Tagliamonte & D'Arcy, 2004), as well tokens of *like* that were used for purposes other than quoting. Tagliamonte & Hudson (1999:153) explain that they "excluded tokens of *like* with the meaning 'for example/such as/in other words/as if to say.' Also excluded were iterations of *like* used for purposes other than quoting, such as hedging, comparing or approximating functions, use as a focus marker or introducing examples, as laid out by Meehan (1991). However, *like* was sometimes used prior to an instance of constructed dialogue that also happens to function to demonstrate a point or provide an example (see 14, below). Whilst most studies included only English, quotations that demonstrated code-switching were not excluded for this dissertation, and whilst some studies (Tagliamonte et al., 2016; Tagliamonte & D'Arcy, 2004) chose to exclude instances of second-person subjects or of non-referential subjects, such as the use of dummy *it* with *be like,* were included. There were a few instances when speakers deliberately created constructed dialogue in order to demonstrate a local accent or the way an acquaintance speaks and after some consideration a decision was reached not to exclude these either.

Tagliamonte and Hudson (1997:253) provide the following examples of "tokens of *like* with the meaning 'for example/such as/in other words/as if to say'", which they excluded:

14. a. He just ***went***, "Oh yeah." Like, "Yeah, really."

b. So my Mends said like, to go away, but not very politely, like, “Fuck off, go away.”

According to the approach used for this dissertation, the second quote in (14a) would similarly not have been a new entry in the database. However, rather than being excluded entirely, the text in both quotes would have been included as a single instance of constructed dialogue introduced by quotative *go*. The example in (14b) would be considered a quotative phrase, despite the *like* that appears to be operating according to Meehan’s (1991) ‘such as’ sense as it precedes an instance of constructive dialogue demonstrating the impoliteness of what Mends said, because it would be presumed that a zero quotative was used to introduce the quote.

3.4.2 Instances of ‘like’, ‘just’ and ‘you know’ in participants’ speech

Some studies, like Kohn & Franz (2009:268), include “bare like” and “Verb + *like*” [verb plus like] as additional, separate quotatives. This dissertation followed the examples set by Tagliamonte & Hudson (1999), Tagliamonte & D’Arcy (2004) and D’Arcy (2010) and did not include ‘bare *like*’ as a separate quotative when it was found after - or before - another quotative (including zero quotative), and chose rather to assume that *like* was being used according to Meehan’s (1998:42) “for example” sense of the word. The following examples have been drawn from Tagliamonte & D’Arcy (2004:500) (15 and 16) and D’Arcy (2010:78) (17).

15. I have a bumper sticker that ***says***, like, ‘I am Canadian.’

16. Like, for example like, ***Ø*** ‘(nonsensical mumbles).’

17. I fucken started off rambling all these names like, ***Ø*** ‘Tania?’
Ø ‘Nah.’
Ø ***‘***Oh fuck, I don’t know.’
She***’s like*** ‘Nicole.’

It is interesting to note the difference between the non-quotative *like* with the zero quotative in the first part of (15) (D’Arcy, 2010:78) and the *be like* quotative at the end of (17). If non-quotative like can combine with a zero quotative, it seems possible that non-quotative *like* could sometimes combine with quotative *be*, and there would be no way of distinguishing between that and quotative *be like*.

Singler (2001:261) observes that *say* can perform a similar complementiser function in AAVE, “but only when the verb of the higher sentence is itself a *verbum dicendi*, most often *tell*”. He provides the following example from Martin and Wolfram (1998:15 in Singler, 2001:261).

18. They tell him say, "You better not go there."

However, he goes on to note that *like* doesn’t have the same limitations as *say* in the above example. He provides the following examples to demonstrate the way *like* can combine with a wider variety of “higher verbs” and occurs in more dialects than just AAVE (Singler, 2001:261):

19. a. And then I went through all this guilt ***like***, "Oh, I must be such a bad person."

b. I was pretending I was a reporter ***like***, "Oh, we are here to ... "

c. I was running around the office ***like***, "Oh my God, there's a cop on the line!"

d. I was like the quasi-supportive friend ***like***, "Oh, it really doesn't look that bad. No one'll notice. Don't worry."

e. He was always the one ***like***, "I'm upset."

Singler (2001:267) also notes that it is also possible for quotative *like* to appear without a higher sentence and provides the following example:

20. Remember Jane? Fucking no-style, like smelled-weird Jane? Like, "Oh my god! You cut your bangs! They look great, Jane! Great, Jane, great!"

This project takes a different approach to Singler (2001) above, and would not have considered the above instances of the use of *like* as quotative. In fact, Singler (2001:262) himself vacillates, saying:

> Given the range of sites where *like* can appear and its range of functions (cf. Underhill 1988), particularly in the speech of the young, it might be argued that the like in examples (15) through (20) is not always a quotative marker.[4]

Singler (2001:274-275) also observes the phenomenon of quotatives verbs, such as go, tell and say, combining with *like* to form what he calls 'double quotatives' which he suggests are representative of an earlier stage in the development of *like* into a quotative. Singler (2001) also seemed to regard *like* as a quotative complementiser, whereas in later literature it is the specific combination of *be + like* that is considered a quotative verb.

During data entry, it was observed that *like, you know* and *just* occurred commonly throughout quotative phrases. These words were found after quotative verbs, between 'be' and 'like', and in the middles of quotes, disrupting the constructed dialogue. As these words and phrases seem to occur in similar places, they may be serving similar or interchangeable functions. This may prove interesting for future research, but it has provided me with additional motivation not to include 'bare *like*' as a quotative. Example 21 demonstrates the approach taken in our data:

21. I think it was one day I just looked around and, like, you know **Ø**, "I'm the only one that's not wearing spectacles here."

Singler (2001:276) also notes that 'just' occurs alongside *like* several times in his database, saying, "Of all the *like* tokens in the corpora, 6% of them are *just like*". However he has more questions than answers about this, asking:

> Is *just like* best understood at face value as the combination of *just* 'only' plus quotative *like*, or is it most often simply a variant of *like*? What, if any, is the difference in distribution between *like* and *just like*? Did *just like* have a distinctive role to play in the emergence of *like* as a quotative, and/or is it now serving some transitional function in *like*'s ongoing evolution?

[4] Singler's examples 15-20 have been relabeled 19 a-e and 20 for the purposes of this dissertation.

3.4.3 The impact of style on quotative usage

One of the variables that the data used for quotative studies is not coded for is style. Style is not included as a variable in studies, nor is it explicitly commented on, other than to note that interviewers adopt techniques to make participants comfortable, so that they can achieve a casual, conversational style that encourages them to recount narratives. Other interviewers (Tagliamonte & Hudson, 1999) work specifically with recorded personal narratives, rather than full conversations or interviews, which could be considered a type of style. As discussed in the literature review, the adoption of the *be like* quotative is part of a larger change that could be interpreted as a change in conversation style, as speakers make greater use of mimetic re-enactment and a more dialogic way of speaking and recounting narratives.

Researchers employed various techniques to ensure that their participants would be comfortable during the data-gathering stage of their research. Tagliamonte and D'Arcy (2004) used data collected by members of the speech community they were studying. Tagliamonte and Hudson (1999) and Blythe et al. (1990) ensured that their interviewers and interviewees were already acquainted with one another. Kohn & Franz (2009) had their child participants interviewed in pairs to reduce formality. And Barbieri (2009) used solely conversations recorded in social and family contexts.

This dissertation followed the example set by previous researchers and has not included style as a variable to be coded for. Nor was it recorded at which stage quotatives occurred during each interview, as the vast majority of quotatives occur during the conversational part of each interview, when the interviewer questions the participant about their life. It is, however, interesting to consider, and a comparative investigation of the quotatives used in different setting and conversation styles may make for intriguing future research.

3.4.3 Various approaches to potential outliers

There was great variation in the number of quotatives produced by each speaker, which could be due to a number of factors. Four of our 64 participants produced no quotatives at all during their interviews, whereas one participant produced 119 in one interview. Individuals who use *be like* in particularly unusual quantities could potentially skew the data, so some researchers implemented strategies to deal with outliers in their databases.

The strategies for dealing with potential outliers vary. Kohn and Franz (2009) simply removed the two participants in their database with a much higher token count of quotatives than the rest of their participants. Tagliamonte and Hudson (1997:154) "included approximately the same amount of speech from each speaker" to minimise the possibility of skewing effects. Other studies placed greater emphasis on the number of tokens for each social group, rather than looking at the number of tokens used by individual participants. D'Arcy (2010:68) ensured that there was an approximately equal number of tokens in each of the two varieties she studies, so that there were "265 for Maori English, 241 for Pakeha English."

Others implemented norming procedures. Barbieri (2009:78) implemented "standard corpus-based methodology" in which "the raw frequency of quotatives in the age groups was normed to 10 000 words". In other words, they worked out the number of quotatives produced by each gender in each age group for every 10 000 words that group spoke. Acquiring a rate of occurrence for quotatives, allowed them to compare the number of quotatives produced by each age group, despite the fact that the amount of speech produced by each group varied greatly. Barbieri (2009:79) clarifies:

> Normalized counts are rates of occurrence, and as such are in principle not influenced by sample size. Thus, in general, the uneven size of the samples of talk in the present corpus does not influence the variability found here in the use of quotatives by men and women in some of the age groups […] However, samples that are very small […] may result in inflated counts. Findings for these age groups should thus be considered with this caveat.

Discrepancies may also occur as a result of having a long data collection period, as the adoption and use of *be like* could have progressed considerably in a decade. Tagliamonte et al. (2016) dealt with a similar issue as they used data collected between 2000 and 2013. They note their concern about these discrepancies and offer a heuristic solution:

> This has ramifications for synchronic analysis, particularly when dealing with an innovative feature that has diffused extremely quickly, both across and within populations. In order to normalize across generationally constituted materials collected at different times, we focus on speaker date of birth (see also Sankoff & Blondeau 2007, Sankoff 2013). Such accountability is a necessary analytical heuristic for tracking the diffusion of

> be like and the subsequent constraints on its use as operative in the various speaker cohorts considered here. (Tagliamonte et al., 2016:828)

For the purposes of this dissertation, the interviews have been included as is, as they were conducted over similar periods of time with similar structures. None of the participants have been excluded and statistical approaches such as mixed-effects models capable of taking into consideration participants as a random variable, are used to counteract the potential skewing effects of any one speaker.

3.5 Dependent Variable

The dependent variable is also known as the response variable or the output variable. For the purpose of this research, the response variable is the quotative used by the speaker. The main quotatives are those that have the highest count in the data. In the Cape Town sample, all quotatives that occurred more than 50 times were included as main quotatives. *Be like, say, think, go, tell* and zero quotative made up the main quotatives. The remaining quotatives were grouped together and labelled 'other quotatives'.

Table 3. 2

Main quotatives and examples

Quotative	Example
be like	He ***was like,*** "What are you doing reading that?"
say	So I ***said***, "No, you can't leave without him."
think	So, I was so worried, I was ***thinking***, "Ooh, where is this child?"
go	And she ***goes***, "Ooh, are you ready for lots of planning?"
tell	I ***told*** him, "Go sit down Tyler or I am going to hit you."
zero	You know, so if my back hurts ***Ø***, "Oh my gosh you must go to the doctor!"

The 'other quotatives' consisted of various verbs that seemed to be performing a quotative function. For example, *mumble* has a quotative function in the following example from our database:

22. She quickly ***mumbled*** quickly “Oh no sh- she’s my ma- maid of honour, cause uh, uh, she’s like a sister to me.”

Be like, say, think, go and the zero quotatives most commonly make up the main quotatives in studies (see Tagliamonte & Hudson, 1999; Tagliamonte & D’Arcy, 2004), although some studies focused solely on *be like*. Quotative *tell* is fairly uncommon in the literature and is primarily mentioned as one of the traditional, but less frequently used quotatives (Barbieri, 2005, 2007; Buchstaller, 2006; Cukor-Avila, 2002; D’Arcy, 2010). It does not feature in the main analyses of these papers, which tend to focus on zero, *go, be like, think* and *say*, or only on be like and other innovative forms. Franz and Kohn (2009), however, do include *tell* as a main quotative in their study looking specifically at quotative use in African American and Latino communities. Franz and Kohn (2009:268) also listed “verb + *like*” and “bare *like*” as part of their frequently used quotatives. Singler (2001:174-275) too observes instances of ‘double quotatives’, when *like* occurs with another quotative, and suggests that they may represent an earlier, experimental stage in the grammaticalisation of *be like*. Within the data used for this dissertation, it was unclear whether instances of *like* were performing a quotative function or one of the many other functions of *like*. As such, quotative verbs preceding *like* were coded alone or, in the absence of a verb, zero quotatives were coded. Sometimes it is even unclear whether *like* is intended to be part of the quotative phrase or the constructed dialogue itself.

Zero quotatives that occurred alongside *like* in our database:

23. a. They'll look at you like ***Ø,*** "What're you doing?"

b. And how she called me ‘dear’, I'll never forget it, like ***Ø***, “Listen here, Dear.”

c. He, you know, knocked and like ***Ø***, "Hey, you forgot your lunch."

d. Cape Malay Muslims have a reputation for smokkeling[5]**,** like ***Ø***, "Ok I can get you this for half the price instead of paying for laminated flooring. My auntie's uncle's brother can come put it in for you."

e. They automatically make fun of the person, like ***Ø***, "You're trying to be White."

[5] A slang word meaning ‘to conduct illicit business’, originating from the Afrikaans word for ‘smuggle’. (Dictionary Unit for South African English, 2020)

This is not the only method that could have been used. Rather than manually reading through interviews to identify quotatives, researchers such as Barbieri (2007, 2009) have made use of concordancing software to identify certain quotative verbs in their simple present and past tense forms before manually excluding non-quotative forms.

Singler (2001:262) expressed concern that the quotative verbs are not always equivalent in meaning. E.g. the morphosyntax of *be like* does not change when it paraphrases something in the form of indirect speech, rather than quoting it directly. Other verbs, like *say*, display a change in morphosyntax or, in the case of *go*, cannot be used to paraphrase speech and are only used to introduce direct quotes.

3.6 Explanatory Variables

These are variables that potentially explain or predict our categorical response variable. The explanatory variables were split into linguistic factors and social factors. Linguistic variables were distinguished from social variables, as they represent the language-internal factors and grammatical environments which may influence quotative use, whereas the social variables may reveal patterns of distribution, as well as which quotatives may be indexical of community identities.

3.6.1 Linguistic variables

The linguistic variables focused on for the purpose of this dissertation are *content of quote*, *grammatical person* and *tense and temporal reference*. *Mimetic re-enactment* has also been shown to affect the use of *be like* (D'Arcy, 2010). However, time constraints meant that, whilst audio files were occasionally consulted and intonation and voicing effects used to determine when a verb was being used as a quotative, there was not time to listen to each interview and record every instance of *mimetic re-enactment*, especially once Covid-19 lockdowns rendered campus (and the computers on which the data were stored) inaccessible. There are a few linguistic variables included in early research on *be like*, but excluded from later research, such as aspect, sentence type and narrative sequence (Blyth et al., 1990). These variables were not included the final analysis of this dissertation.

When it comes to *be like*, it can be difficult to differentiate between when it is used to quote thought and when it is used to quote speech (Ferrara & Bell, 1995; Tagliamonte &

Hudson, 1999; Singler, 2001). In these instances, the procedure described by Tagliamonte and Hudson (1999:156) was employed:

> "We made use of the broader narrative context in which the quotative occurred as well as intonational cues. If the constructed dialogue reported an attitude or a general feeling of the narrator or group of people, then it was considered internal dialogue. If it was contained in a sequence of reported dialogue (i.e. complicating action) which advanced the story-line, or was part of an utterance to which the protagonists responded, it was coded as direct speech."

However, admittedly this is an imperfect process and there is room for error and inaccuracy. Speakers may even use *be like* in order to be intentionally ambiguous about whether their constructed dialogue reflects inner or spoken narrative.

For the *content of quote* variable, each instance of constructed dialogue was coded either as "im" for internal monologue and thoughts, "s" for speech and "e" for the expression of non-lexicalised sounds and exclamations. Following the lead of Fox and Robles (2010:727), exclamations such as "*m:::*, *oh no*, *wow*, *man*, *hey*, and *o:::h*" were treated as response cries (Goffman, 1978:800 in Fox & Robles, 2010:727) and when they occurred alone, rather than as part of a sentence, they were coded the same way as non-lexicalised sounds. Provided below are examples from our database of *be like* being used to express inner dialogue or an attitude (24), to express exclamations and non-lexicalised sounds (25) and to express speech (26):

24. You***'re like***, "What? The lecturer didn't tell me about this!"

25. a. In the middle of their songs this like hard core heavy moment and I***'m like***, "Wooah!"

    b. Then the next time you look at it, it**'s like**, {squirting sound}

26. My doctor told me. He ***was like***, "Unless you change your lifestyle entirely you won't be able to walk, then."

The *grammatical person* of each quote was originally also coded by number, so that each quotative phrase (that phrase which expressed the speaker and preceded the constructed dialogue) was coded as "1sg", "1pl", "3sg", "3pl" or "2sg" (2pl was not included, because second person does not inflect for number in English), but later collapsed into "1st" (27), "2nd" (28) and "3rd" (29), using R, when grammatical number was excluded from the analysis.

27. I ***was like***, “How do you play guitar?”

28. We have quiet nights sometimes. Then you stand there and you***'re like***, “Please can I go home!?”

29. The one guy, this, um, security guard, stops me and he***'s like***, "Excuse me, can we like search you?"

The *tense and temporal reference* of the quotative/quotative phrase was coded either as ‘past’, ‘present’, ‘future’, ‘HHT’ or ‘HP’. ‘Past’ and ‘present’ refer to quotative phrases in past and present tense respectively. ‘HP’ refers to historical present tense, in which the present tense is used in sentences that have a past time reference. Some researchers (Tagliamonte et al., 2016; Cukor-Avila, 2012; D’Arcy, 2010; Buchstaller & D’Arcy, 2009; Kohn & Franz, 2009) coded only for past, present, HP and ‘other’. However, rather than including an ‘other’ category, phrases that did not fit into the usual tense categories were divided between ‘HHT’ and ‘future’ categories. “HHT” stands for hypothetical, habitual and timeless and includes phrases which serve these, as well iterative and generic functions. Many of these phrases were in irrealis moods, such as hypothetical, conditional, potential or eventive moods, often making use of the verb ‘would’. Most instances of hypothetical speech were expressed in the eventive (both conditional and potential) mood and made use of the verb ‘would’. The ‘future’ tense was used only for quotative phrases which contained the auxiliary ‘will’, although there were also a few instances in which ‘gonna’ or ‘going to’ were used to indicate future time reference. In an effort to further simplify the tense and temporal reference predictor, Tagliamonte et al. (2016:829-830) excluded “less frequent tense and aspectual constructions such as future temporal reference, past perfect, and habituals”, from their data and focus instead on the contextual effects already recognised and discussed in the literature. The following examples demonstrate the use of *be like* in past (30), present (31), HP (32), HHT (33), and future tense (34) within out database:

30. I ***was like***, "No, I'm not from the scouts."

31. But she's too scared to ***be like***, "I don't want to join your club anymore"

32. She ***says***, "You should really come and like chill and hang out with me.” So I***'m like***, "Okay fine I can do that."

33. You know, I would ***be like***, "Ja bru”

34. I'll ***be like***, " Come on, give me a hint for the test."

Franz and Kohn (2009:270) included quotative phrases without a finite verb, coding the tense as “other” in their analysis. In contrast, other researchers would have excluded cases such as the following example they provide, because it contains no tense information:

35. Then, “No, I was doing my homework,”

Instances such as these (35), without a verb to provide tense information, were all included as zero quotatives in our database.

Tagliamonte and D’Arcy (2004) excluded quotatives accompanied by incomprehensible quotes or uncertain *grammatical person* from their database, and excluded second person, hypothetical and written quotes from some of their tests. Tagliamonte et al. (2016) excluded any data that was not fully expressed for all the constraints, as well as instances of the second person, dummy *it* and zero subjects, as excluding them had become established as common variationist practice in quotative analysis. However, second person, dummy *it* and hypothetical quotations were included in this dissertation as they seemed to perform interesting functions. Dummy *it* occurred only with *be* (36) and *be like* (37), and the second person seemed to be performing a hypothetical/rhetorical function (38).

36. a. And then it ***was***, "Wat gaan aan met jou hare?"[6]

    b. Just through being w- good at academics, it ***was*** sort of, "Ah, he's the teacher's pet,” or “He's the nerd,” or whatever.

37. a. It***’s like***, “You're all going to amount to something. You're, like, a Bishops boy. You're gonna, you're gonna do something with yourself at the end of this.”

    b. It***’s like***, “Who does this? Who chops these up?”

    c. And it***’s like***, “Ja, I want to be like them.”

    d. You know, it***’s like***, just, “I'm putting up with your accent; I'm putting up with the fact that you can't speak English; we can't understand each other; I'm drawing pictures with my hands; I'm drawing pictures on a pad and paper to understand what you're saying. Please tip!”

38. a. When you come and ***say***, "I want food"

[6] Afrikaans meaning “What’s going on with your hair?” in English.

b. And you***'re*** just ***like***, "Please just give me my order."

c. When people walk past you, and they're speaking Afrikaans, and you***'re like***, "What?"

d. Initially, you ***thought***, “I can't understand this guy!” And, and, all of a sudden after, six months or so, you, you get used to it.

3.6.2 Social variables

The social variables under consideration relate to the age, schooling, race, and gender of the speakers. These are factors which may influence speakers’ use of quotatives.

For the variable *age*, the exact age in years of the participant at the time of interview was recorded. When this was missing from the interview, the speaker’s age was calculated by subtracting their date of birth from the date of interview. Unfortunately, the exact dates were not always available, so the decision was made to code *age* according to two discrete categories, ‘young’ and ‘old’. Based on the range of ages within the database, ‘young’ speakers were below 35 years old and ‘old’ speakers were over 50 years of age. Buchstaller and D’Arcy (2009) used a similar approach, dividing their American English, British English and New Zealand English speakers into either a ‘Younger’ or ‘Older’ age category. Their exact age ranges also varied based on the ages of the participants in each area. For example, in England their older speakers were between 38 and 69 years old, whereas in New Zealand they were between 40 and 63 years old.

Speakers’ schooling was coded as “modelc” for participants who attended an FMC school and “not” if they attended an FDET school. Speakers’ race was coded as “b” for Black, “c” for Coloured, “w” for White and “i” for Indian. Participants’ self-identification and parental history were used in order to assign race (Mesthrie, 2012). Speakers’ *g*ender was coded as “m” for men and “f” for women.

Multilingualism could have been included as another social variable. Most researchers focused on monolingual English speakers or did not account for either mono- or multilingualism (Blyth et al., 1990; Tagliamonte & Hudson, 1999; D’Arcy, 2004; Buchstaller, 2006; Tagliamonte & D’Arcy, 2009; Barbieri, 2009; Buchstaller, 2015), with only a few including multilingual speakers as we have (D’Arcy, 2010; Davydova, 2019). Franz and Kohn (2009:283) look explicitly at second language speakers of English within the Latino

community and note that they "quickly acquire quotative *be like*, along with recognition of its pragmatic function to frame both thought and speech." However, they also suggest that second language speakers with either brief or much longer exposure to English would use quotative *say* more. This implies the possibility of a period of hyper-correction.

Other studies, such as that conducted by Buchstaller and D'Arcy (2009) include socioeconomic status, calculated by assessing speakers' education and occupation, as a social variable. Singler's (2001) results showing the impact of different combinations of interviewer and interviewee genders revealed the potential significance of also including interviewer social variables. Interviewer variables, including interviewee, interviewer and interviewer gender were coded during the data entry phase. Each interviewee was assigned a number (e.g. "P01") instead of their name for anonymity and convenience. The interviewers were coded as "Alida", "Raj", "Tracey" and "AXL3" for interviews conducted by third-year linguistics students. Interviewer gender was coded as 'm', 'f', or 'm&f' when AXL3 students conducted interviews together, and on student was male and the other female. However, the first statistical runs in R showed that these variables were not significant, so they were excluded from the final analyses.

3. Statistical Methods

As Buchstaller and D'Arcy (2009:295) note, there is not an established methodology for mapping the way that "global cultural flows" enter and travel through local spaces. They acknowledge that there are certain heuristic frameworks, such as Milroy and Milroy's (1992) social network theory and Eckert's (1989; 2000) understanding of communities of practice, for understanding how small, local groups of speakers may use global linguistic variables in various ways to accomplish different aims. However, Buchstaller and D'Arcy (2009) call for more micro-linguistic studies that record in greater detail the way local and global resources combine and impact one another. This is what this dissertation, and various other publications, has attempted to provide. The data gathered was first explored using descriptive statistics, after which I performed a regression analysis and made use of conditional inference trees to better understand the factors constraining *be like* across different local varieties.

3.7.1 Descriptive statistics

After the initial entry of data into Excel spreadsheets all the database management and statistical analysis for this dissertation was performed using R (R Core Team,2019) and, more

specifically, the integrated development environment, RStudio (RStudio Team, 2019). The following packages were also used for exploratory statistical methods, including data visualisation and rudimentary analysis, as well as managing and manipulating the dataset: dplyr (Wickham, François, Henry & Müller, 2020), reshape (Wickham, 2007), ggplot2 (Wickham, 2016), ggmosaic (Jeppson, Hofmann & Cook, 2018) and ggpubr (Kassambara, 2020).

3.7.2 Regression analysis

Once the data had been summarised and potential trends identified using descriptive statistics, various multivariate logistic regression analyses were used to determine the constraints on *be like* and other quotatives. The first popular tool for performing these regression analyses was VARBRUL (Blyth et al, 1990; Singler, 2001). As this became less popular, researchers turned to tools like GoldVarb (Tagliamonte & Hudson, 1999; Cukor-Avila, 2002) Goldvarb 2.0 (D'Arcy, 2004; Davydova, 2019) and GoldVarb X (Kohn & Franz, 2009; Buchstaller & D'Arcy, 2009), before using Rbrul (Johnson, 2009) or simply making use of statistical packages in R (R Core Team, 2019). As R is free, easily accessed and frequently updated, I have opted to use lmerTest (Kuznetsova, Brockhoff & Christensen, 2017), rms (Harrell, 2020) and mlogit (Croissant, 2020) to perform logistic regression analyses on R.

Singler (2001) published an article titled, 'Why you can't do a VARBRUL study of quotatives and what such a study can show us', in which he explains that VARBRUL requires response variables to be in free variation, but this is not the case for quotatives. The use of quotative verbs is restricted according to the *content of quote* variable, and certain quotatives can be used for functions that other quotatives cannot. For example, *think* and *say* cannot be used interchangeable as they alter the meaning and interpretation of the quotative phrase. This makes quotatives unsuitable subjects for a typical VARBRUL study. He explains:

> It is a general principle in the use of VARBRUL that, for a given set of variants, only those tokens should be included for which all of the variants are permissible. If a given environment blocks full variation, then that environment is excluded from tabulation. In variationist studies of AAVE copula, for example, sentence-final tokens are routinely excluded from consideration because only full forms of the copula can occur in that environment (cf. Blake 1997). The distribution of quotatives in the corpora and subsequent confirmation from native speakers make the point that there

> are several environments where quotative choice is constrained. Crucially, when that which is "quoted" is not speech itself, the choice of quotative is restricted. (Singler, 2001:259-260)

Singler (2001:264-265) has a further objection to VARBRUL, which concerns the nature of the change in which *be like* has been adopted. He highlights what was discussed in section 2.7 of the literature review, that change taking place in quotative systems is not simply the replacement of certain quotative verbs with new forms:

> Rather, a more fundamental change has occurred, a change in the content of quote of usage. If by "variants" we mean alternative forms that are equivalent in meaning, then *say/go*, on the one hand, and *like/all*, on the other, are not congruent in content of quote, hence are not variants of each other. (Singler, 2001:264-265)

However, multivariate studies are useful for determining the different constraints on the various quotatives, giving us an indication of their functions. Tagliamonte and Hudson (1999:155) also observe that it is worth looking at all the quotatives making up a quotative system, despite the fact that not all of them can be used interchangeably, because understanding where *be like* "may or may not be expanding" help us "to track its developmental pathway." It was with this in mind, that a multivariate analysis was performed. This analysis was, however, limited by the exclusion of zero quotatives, which frequently did not contain the necessary *tense and temporal reference* and *grammatical person* information.

Tagliamonte et al. (2016:830) used both the "fixed-effects logistic regression models in GoldVarb (Sankoff et al. 2012) to test the simultaneous and multifaceted effects of the predictors on the use of *be like*", as well using the "mixed-effects logistic regressions using the lme4 package in R (R Core Team 2007, Bates et al. 2011)". The GoldVarb fixed-effects regression models were used "to examine the operation of the variable grammar constraining be like in each of the two primary cohorts isolated by the conditional inference trees: that is, speakers born in the 1970s, and speakers born in the 1980s, who fast-tracked the change" (Tagliamonte et al., 2016:834) for each of the four cities they examined. They then used the lme4 mixed-effects model, applying it to their entire dataset "in order to control for the random effects of individual in the sample, as well as to assess the interaction between predictors" (Tagliamonte et al., 2016:834). They used a binomial choice mechanism for the mixed-effect

model, "contrasting *be like* versus all other forms" (Tagliamonte et al., 2016:834), so that they could examine how linguistic variables constrained *be like* specifically across the different cities.

After applying the mixed-effects logistic regression model to each of the cities, Tagliamonte et al. (2016:836-837) ran a mixed-effects analysis for all of their speakers born in the 1970s and 1980s. For this purpose, "city is treated as a fixed effect" and unlike in their previous models, individual participants were included as a random effect.

Tagliamonte et al. (2016)'s mixed-effects model of logistic regression found that the statistically significant linguistic variables were *tense and temporal reference*, *grammatical person* and *content of quote*. *Tense and temporal reference* had the strongest effect and *content of quote* the weakest.

> Logistic regression models the relationships between a categorical response variable with two or more possible values and one or more explanatory variables, or predictors. This technique is particularly popular in probabilistic multifactorial models that explain and predict the speaker's choice between two or more near synonyms or variants on the basis of conceptual, geographic, social, pragmatic and other factors. (Levshina, 2015:253)

Various iterations of the GoldVarb multivariate, multiple regression package (Rand & Sankoff, 1990; Sankoff, Tagliamonte and Smith 2005; Sankoff et al. 2012) have been used by researchers such as Franz and Kohn (2009), D'Arcy (2010), and Davydova (2019) to compare the effect of linguistic variables on different quotatives across different groups of speakers. Similar to VARBRUL, the effect of various constraints is determined by factor weights.

> Factor weights […] indicate the probability of occurrence with each factor relative to the input value. These figures range between 0 and 1; figures closer to 1 indicate a higher probability and those closer to 0 a lower probability. The range provides a non-statistical measure of the strength of a factor group relative to other groups in the model (Poplack and Tagliamonte 2001: 93–94; Tagliamonte 2006: 242). (D'Arcy, 2010:69)

Tagliamonte et al. (2016:834-835) also offer an explanation of factor weights in GoldVarb:

> [F]actor weights ('FW') report weighted probabilities; those closer to 1 indicate contexts that favour selection of be like, whereas those closer to 0 indicate disfavouring contexts. When the values within a group hover close to .50, the predictor is typically not significant, as the contexts neither favour nor disfavour selection of the dependent variable. Three lines of evidence apply to the results: statistical significance (assessed at the 0.05 level), constraint hierarchy (the direction of effect within a predictor), and constraint ranking (the relative effect of a predictor vis-à-vis other predictors in the model, reflected via the nonstatistical measure of range, which gauges the strength of the factor group in the model) […] Factor weights are uncentred in GoldVarb, meaning that they are estimated based on the number of tokens in a level (i.e. more robustly represented levels have a larger effect on the estimates within groups). (Tagliamonte et al., 2016:835)

As factor weights are the primary measure used to report the strength and effects of constraints on quotatives within the literature, but also seem to be unique to regression software generated specifically for use in linguistics, the author wanted to ensure that there was a similar or equivalent measure in logistic regression in R, so that it would be possible to compare the results of this dissertation to those in the literature. Unlike in Tagliamonte et al. (2016), a GoldVarb analysis, which may have circumvented the problem, was not included. Logistic regression results in R provide log odds ratios as a measure for the effect of various explanatory variables on the response variable. In his Rbrul manual, Daniel Ezra Johnson (n.d., section 2.2) describes log odds ratios as "the raw coefficients of the logistic model" and explains that they relate directly to factor weights and, if the factor weights were centred, rather than uncentred as in GoldVarb, "the correspondence [between log odds and factor weights] would be exact."

It is also easy to transform the log odds ratios into odds ratios. Log odds ratios "compare the odds of [a specific] outcome for each level of a predictor with the reference level" (Levshina, 2015:260), whereas odds ratios "are a simple ratio of the probability of one event to the probability of another event" (Levshina, 2015:261). When one interprets factor weights, a value closer to 1 conveys a strong positive correlation between the response variable and the explanatory variable in question, and a value closer to 0 conveys a strong negative correlation between the two variables. Whereas, in an odds ratio, a negative correlation will be conveyed by a value between 0 and 1, and a positive correlation will be conveyed by a number greater than 1, meaning that odds are centred around 1. Log odds ratios, on the other hand, are centred

around zero, with positive correlations represented by positive values and negative correlations represented by negative values (Levshina, 2015:260).

The lrm and glm function in the rms (Harrell Jr, 2020) and base R (R Core Team, 2019) packages were used to run fixed-effect binomial logistic regression model. In order to include a multinomial logistic regression model, the mlogit function in the mlogit package (Croissant, 2020) was also employed. However, the mlogit model encountered unsolvable errors when random effects were included, so the model in chapter 5 is a fixed-effect model. In order to obtain more nuanced and possibly more accurate results, the glmer function in the lmerTest (Kuznetsova et al., 2017) package (an alternative to lme4 that was better suited to logistic regression models at the time of running statistics) was used to fit a mixed-effects binomial logistic regression model which includer speaker as a random effect.

Singler (2001) objected to the use of multivariate regression analyses for quotative studies, as quotatives are not in the appropriate free variation required for these models, which caused the author to question the suitability of quotative data for logistic regression analyses. According to Levshina (2015:271) there are three assumptions one should be able to make about data that is to be used for logistic regression: That "the observations are independent (of one another)", that "the relationships between the logit and the quantitative predictors are linear", and that "no multicollinearity is observed between predictors."

Considering that the Cape Town database is made up of quotatives that occurred in individual interviews, one cannot confidently say that all observations are independent of one another, as some of the quotatives occur in chains of constructed dialogue during narratives recounted by speakers. However, this can be accounted for in mixed-effects models that are fitted to count speaker as a random effect. The second assumption is not of concern, as the Cape Town data consists entirely of qualitative predictors. Multicollinearity, in the third assumption, is when two predictors are highly linearly related with one another, resulting in a correlation between them that is equal to 1 or -1 (Levshina, 2015). There may be instances in the database when explanatory variables are highly linearly related with each other. In fact, the two social variables of race and schooling can be considered an example of multicollinearity, as all White students in the database attended FMC schools. Whilst logistic regression is useful for comparing our results to the literature, there may be statistical methods that are better suited to the type of data used in this dissertation

3.7.3 Conditional Inference trees and random forests

Conditional inference trees (ctrees) are "non-parametric tree-structure models of regression and classification that can serve as an alternative to multiple regression" (Levshina, 2015:291), generated using the partykit package (Hothorn, Hornik & Zeileis, 2006; Hothorn & Zeileis, 2015). The branches of the trees are created when the data splits into two groups because an explanatory variable is highly correlated to different response variables, which will be represented on each of the two branches resulting from the split. The initial split is the one with the highest correlation with the response variable. This continues to occur along each branch of the tree, until there are no longer any statistically significant splits to be made.

> A split is implemented when the criterion exceeds the value given by mincriterion as specified in ctree_control. For example, when mincriterion = 0.95, the p-value must be smaller than 0.05 (sic) in order to split this node. This statistical approach ensures that the right-sized tree is grown without additional (post-)pruning or cross-validation. The level of mincriterion can either be specified to be appropriate for the size of the data set (and 0.95 is typically appropriate for small to moderately-sized data sets). (Hothorn, Seibold & Zeileis, 2019:8)

Conditional inference trees were included in this dissertation for two primary reasons. Firstly, they are well-suited to data with few observations, but many potential predictors, complex interactions between variables and "highly correlated predictors" (Levshina, 2015:292), such as the *content of quote* variable. Secondly, conditional inference trees, unlike other regression analyses, use unbiased variable selection, rather than showing bias "toward variables with many potential splits" (Levshina, 2015:292). This is because they rely on binary recursive partitioning.

It is also useful that conditional inference trees provide *p*-values at every split, providing a measure of confidence or the reliability of each split. They also do not require any 'pruning' in order to prevent overfitting the model, and are "considered to be robust in the presence of outliers" (Levshina, 2015:292). You can set the minimum *p*-value for a split to occur, by specifying the 'minicriterion' value:

> For example, when mincriterion = 0.95, the p-value must be smaller than 0.05 (sic) in order to split this node. This statistical approach ensures that

> the right-sized tree is grown without additional (post-)pruning or cross-validation. (Hothorn, Seibold & Zeileis, 2019:8)

The *p*-values at each split are calculated using permutation. Levshina (2015:293) provides the following explanation of permutation:

> Permutation means that the labels on the observed data points are rearranged many times, and for each rearrangement the relevant test statistic is computed. This is a way of obtaining the distribution of the test statistic under the null hypothesis of no difference, no association, etc. Next, one determines the proportion of the permutations that provide a test statistic greater than or equal to the one observed in the original data. If that proportion is smaller than some significance level, then the result is significant at that level. Permutation is similar to boot-strapping in that they both are non-parametric resampling methods, which use the same data for validation of results. However, they are not identical, since the former involves reshuffling of the labels, whereas the latter draws numerous random samples from the original sample.

Permutation is also used in the creation of random forests (Hothorn, Buehlmann, Dudoit, Molinaro & Van Der Laan, 2006; Strobl, Boulesteix, Zeileis & Hothorn, 2007; Strobl, Boulesteix, Kneib, Augustin & Zeileis, 2008), which are grown from a specified number (usually a large number, for example 1000) of conditional inference trees. Random forests produce an average measure of the importance of each variable in a model by reviewing the importance of each variable across the many conditional trees it creates. "This measure reflects the impact of each predictor given all other independent variables." (Levshina, 2015:293) Because of the variable nature of permutation, the exact value of each importance measure may vary every time a random forest is generated. For this reason, it is recommended that you create multiple random forests and ctrees to ensure that the results are consistent. (Levshina, 2015)

Tagliamonte et al. (2016:830) also used conditional inference trees "to reveal the relevant temporal juncture points in [their] data" and "to test the simultaneous and multifaceted effects of the predictors on the use of *be like*." In their conditional inference trees, they contrasted speaker date of birth, which represented a continuous predictor, with quotatives which they divided into two groups, one containing all the *be like* quotatives and another

containing all other quotatives. Their intention was to "isolate the points of deviation (either in space or time) in the diffusion of *be like*" that represent the "bursts of change (Lass 1997:304)" that characterise "typically discontinuous (Jand & Joseph 2003:20)" language change (Tagliamonte et al., 2016:832). This allowed them to "identify the precise generational divisions in the upsurge of *be like*" (Tagliamonte et al., 2016:832).

4. Descriptive Results of Quotatives in Cape Town

This chapter reviews the initial findings on quotative usage among the young Cape Town speakers in Professor Rajend Mesthrie's Database of South African Englishes. The proxy control group of speakers from an older generation is included as a means of gauging the degree to which *be* like has been adopted by the younger speakers. The quotatives used by participants during sociolinguistic interviews are examined, tallied and visualised using R. Once the spread of quotatives for each age group has been established, the findings for *content of quote*, speaker schooling, speaker gender, *tense and temporal reference* and *grammatical person* are presented for each age and race group.

From the 64 interviews, 1791 quotatives were identified and recorded. They were then imported into RStudio, where the base R (R Core Team, 2019), dplyr (Wickham et al., 2020), tidyr (Wickham, 2020), reshape and reshape2 (Wickham, 2007) packages were used to organise the data and perform any calculations. These descriptive statistics aim to review the quotatives found in the database and examine how the most common quotatives, *be like*, *say*, *think*, *go*, *tell* and zero quotatives are used according to *grammatical person*, *tense and temporal reference* and the content of each quote, as well as the age, race and schooling of the speaker. This allows for a broad understanding of the database and enables the identification of potential trends that may warrant further investigation via inferential statistics in sections 4.2 and 4.3.

Table 4. 1

Count of all quotatives

Quotative	N
Be like	793
Say	542
Zero	177
Think	74
Tell	59
Go	51
Ask	32
Be	16

Decide	8
Look	6
Feel	4
Realise	4
Mean	3
Remind	2
Would	2
Greet	2
Notice	2
Start	2
All like	1
Be all	1
Find	1
Go on	1
Joke	1
Know	1
Message	1
Mumble	1
Spread	1
Pronounce	1
Warn	1
Will	1
Total	1791

A total of 30 different quotatives appeared in the data (see table 4.1). *Be like* is the most frequently used quotative, uttered a total of 793 times, followed by *say*, which is used 542 times and zero quotatives, which are used 177 times. This, alone, reveals that *be* like has been enthusiastically adopted among Cape Town speakers. *Think* only occurs 74 times, *tell* 59 times, *go* 51 times and *ask* 32 times. *Be* is used 16 times, *decide* 8 times, *look* 6 times, *feel* and *realise*

are used 4 times, *mean* 3 times and *remind*, *would*, *greet*, *notice* and *start* are all used twice, respectively. For the sake of convenience, quotatives which occurred over 50 times in the data, *be like*, *say*, *go*, *tell*, *think* and zero quotatives are considered the main quotatives, which we will focus on, and the rest of the quotatives have been grouped together as 'other quotatives'. All graphs and charts in section 4.1 were generated using the ggplot2 (Wickham, 2016) package and the viridis colour-blind-friendly colour maps (Garnier, 2018) in R.

Figure 4. 1

Pie chart of all quotatives

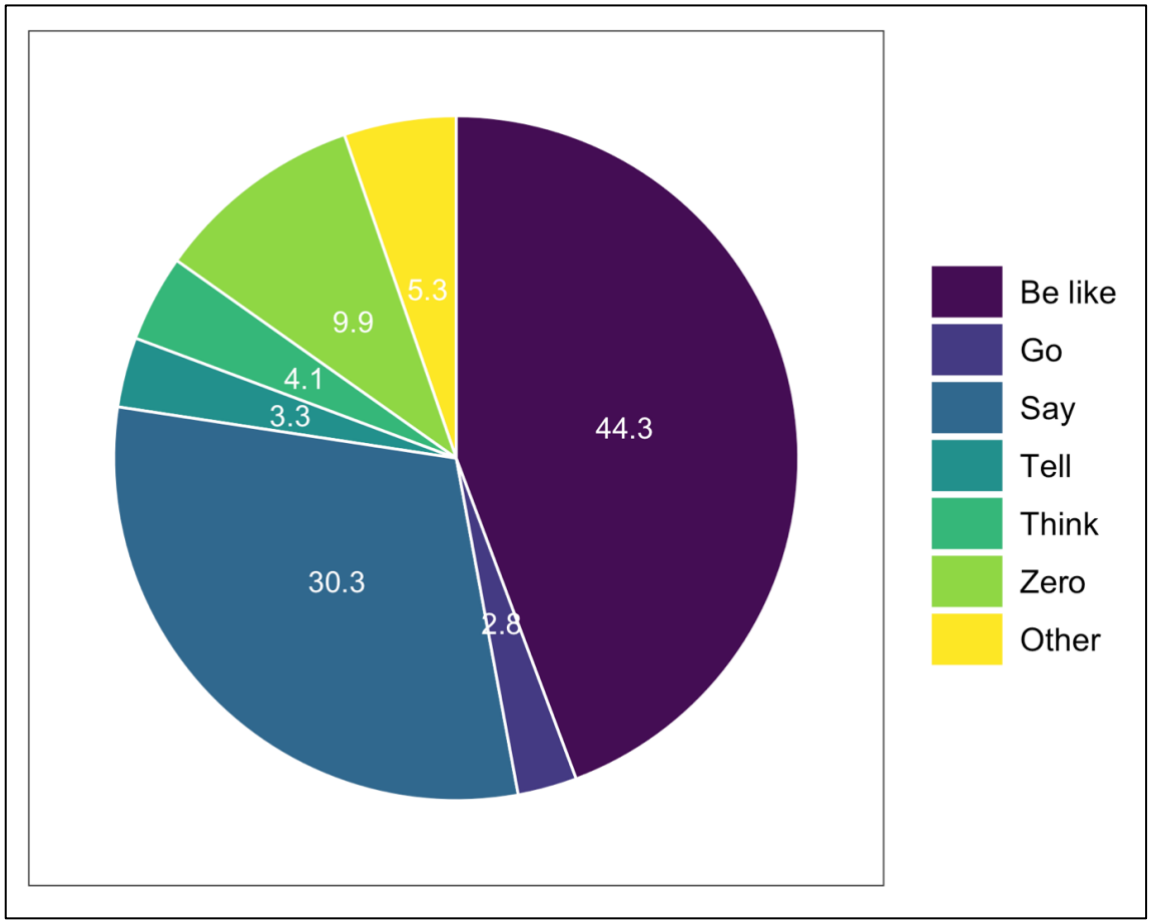


Table 4. 2

Counts and percentages of quotatives

Quotative	N	%
Be like	793	44.3
Say	542	30.3
Zero	177	9.9
Other	95	5.3
Tell	59	3.3
Think	74	4.1
Go	51	2.8
Total	1791	100

Altogether, the main quotatives form 94.75% of the database, whilst the other quotatives are only 5.3% of the database (see figure 4.1 and table 4.2). As the two most commonly used quotatives, *be like* and *say* make up 44.3% and 30.3%, respectively. However, the complete set of quotatives will not be the primary focus of this chapter. Most of the following statistics have been conducted using only the main quotatives that were produced by younger participants. Depending on whether *grammatical person* and *tense and temporal reference* are included in the visualisation, calculation or analysis, zero quotatives may be included or excluded.

Figure 4. 2

Pie chart of all quotatives used by young participants

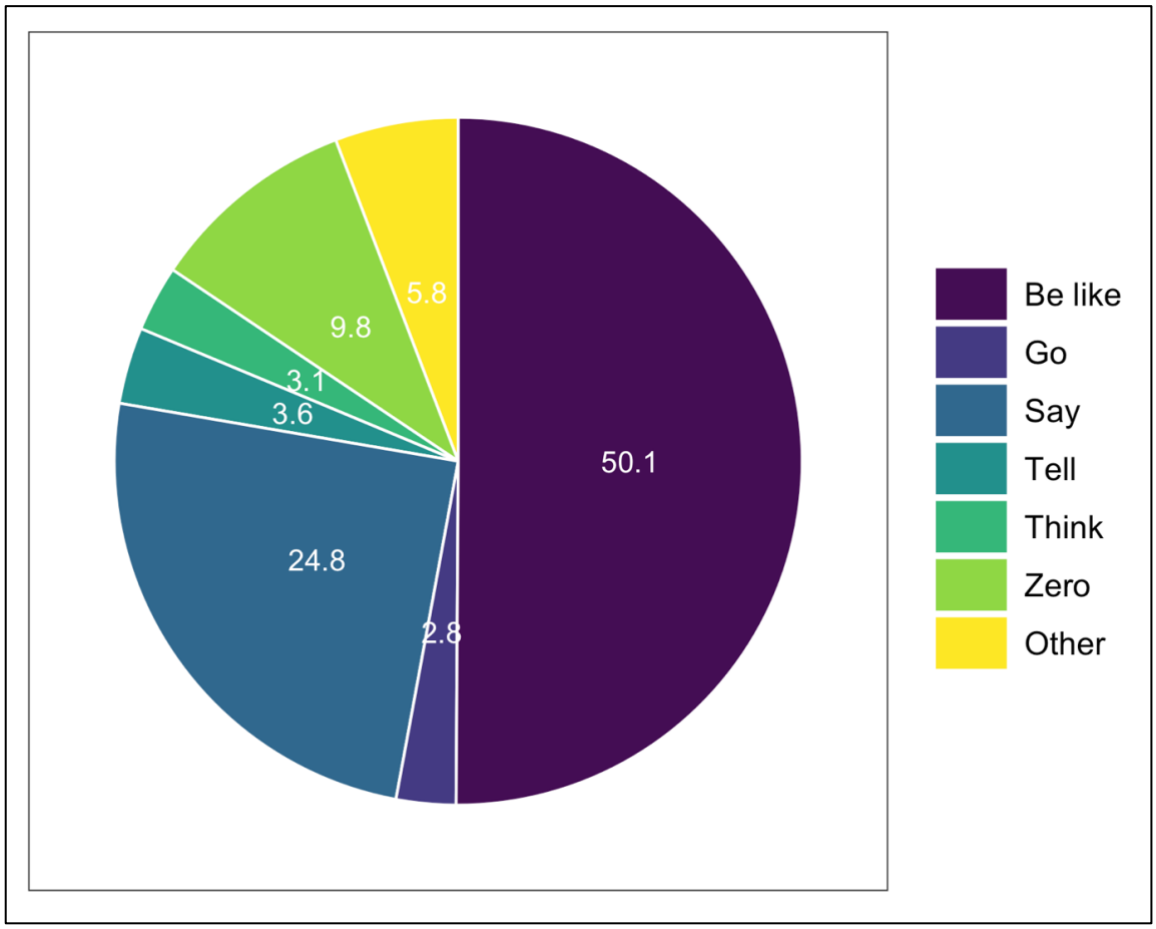


Table 4. 3

Counts and percentages of all quotatives used by young participants

Quotative	N	%
Be like	791	50.1
Say	391	24.8
Zero	154	9.8
Other	92	5.8
Tell	57	3.6
Think	49	3.1
Go	45	2.8
Total	1579	100

Among younger speakers, all of whom were between ages 16 and 35, the preference for *be like* is even stronger, as it makes up just over half of the quotatives expressed in their interviews. With the exclusion of older speakers, the percentage of *say* quotatives drops from 30.3% to 24.8% and *think* from 4.1% to 3.1%. The percentages of *go* and zero quotatives remain fairly consistent, with *go* remaining the same and zero quotatives decreasing by only 0.1%, whilst *tell* and other quotatives increase slightly from to 3.3% and 5.3% to 3.6% and 5.8% respectively (see figure 4.2 and table 4.3).

Figure 4. 3

Pie chart of main quotatives (excluding zero quotatives) used by younger speakers

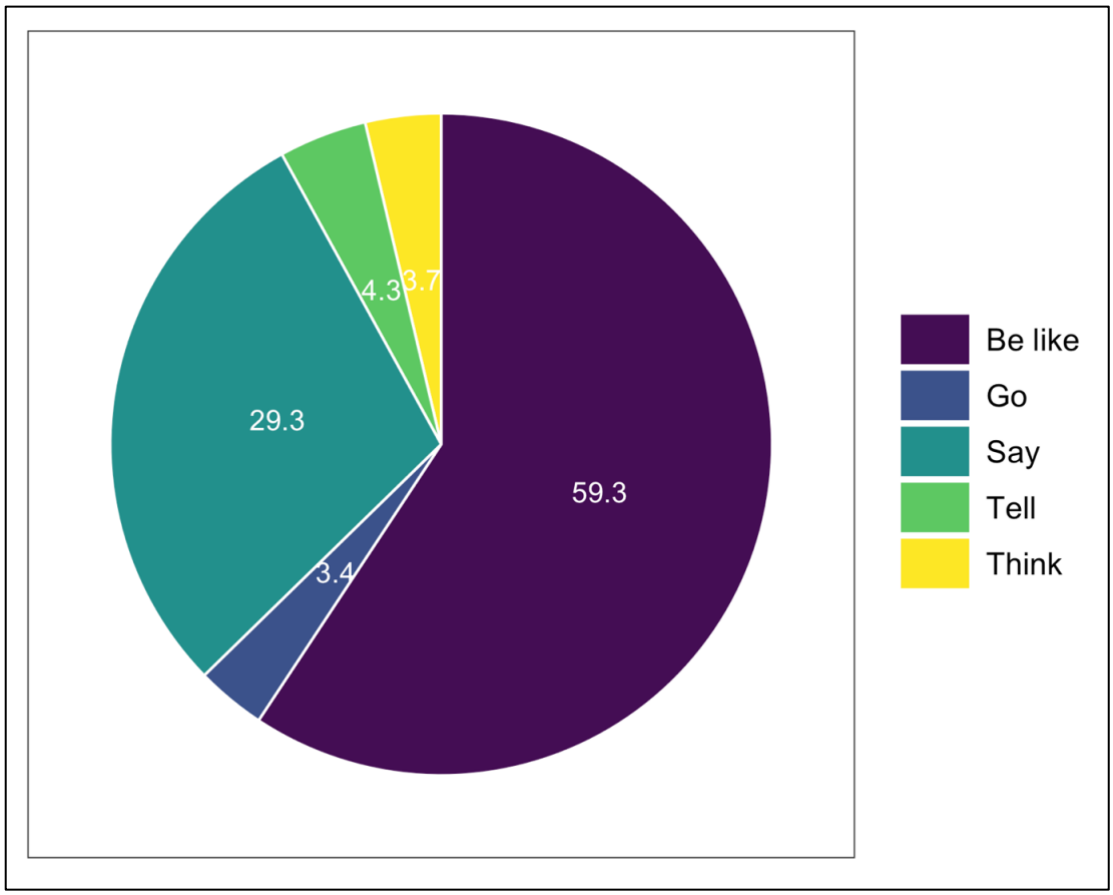


Table 4. 4

Count and percentages of main quotatives (excluding zero quotatives) used by young participants

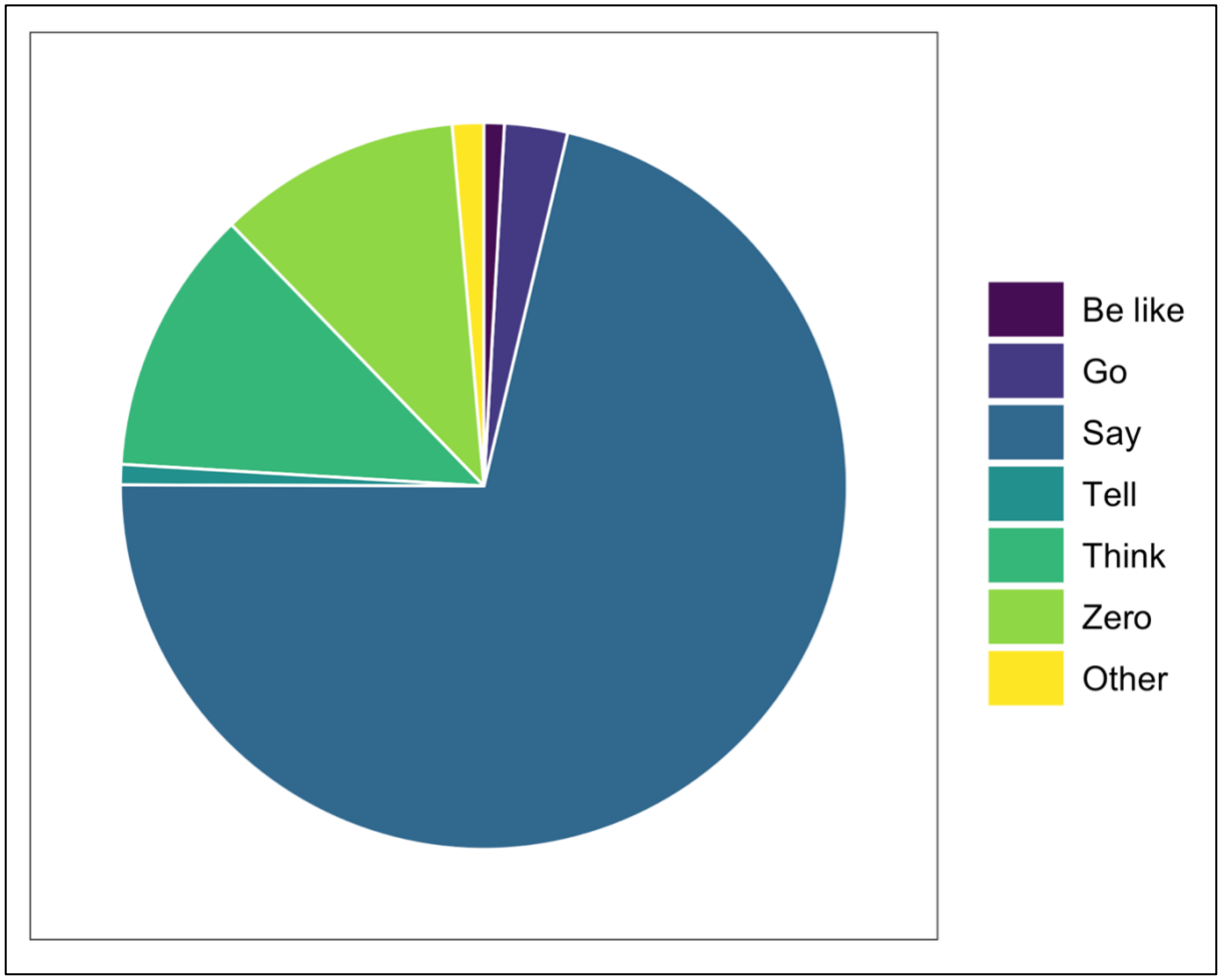
Quotative	N	%
Be like	791	59.3
Say	391	29.3

Tell	57	4.3
Think	49	3.7
Go	45	3.4
Total	1333	100%

When other and zero quotatives are excluded, in order to form the main, decreased database that is used for inferential statistics, *be like* makes up 59.3%, *say* 29.3%, *think* 3.7%, *go* 3.4% and *tell* 4.3% of the quotatives (see figure 4.3 and table 4.4).

Figure 4. 4

Pie chart of all quotatives used by older participants[7]



[7] Percentages are provided in the graph below, as they were not visible in the pie chart due to the small segments representing quotatives with low counts, such as *be like* and *tell.*

Table 4. 5

Counts and percentages of quotatives used by older participants

Quotative	N	%
Say	151	71.2
Think	25	11.8
Zero	23	10.8
Go	6	2.8
Tell	2	0.9
Be like	2	0.9
Other	3	1.4
Total	212	100

The older participants, in contrast with the 35 and younger group, were all White people over the age of 50, who had spent most of their life in Cape Town. Among the speakers from this proxy control group, *say* was used the most frequently to quote direct speech and made up 71.2% of the quotatives they used (see figure 4.4 and table 4.5). *Think* and the zero quotative were the next most common quotatives, used 11.8% and 10.8% of the time, respectively. *Go* was used 2.8% of the time and other quotatives were used 1.4% of the time (two instances of *ask* and one of *decide*). Interestingly, there were two instances of both *be like* and *tell*, each making up 0.9% of the quotatives. One use of *be like* was by the youngest of the older speakers, who also happened to teach at a university which therefore facilitated constant interaction with younger students, and it was used to quote speech. The second instance of *be like* was an example of *be like* with the impersonal or neuter pronoun, 'it's like', and was also used to quote speech.

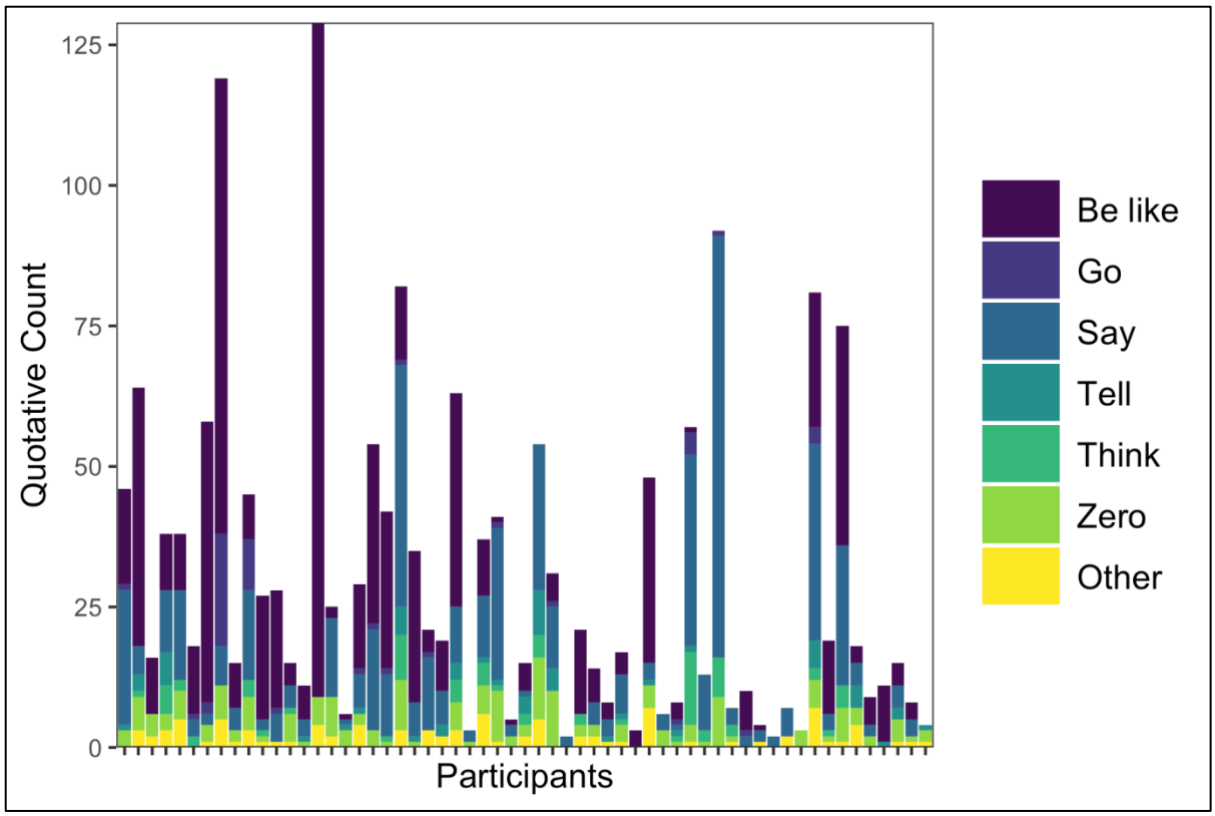
When the quotative cohort of the older participants (figure 4.4 and table 4.5) is compared with that of the younger participants (figure 4.2 and table 4.3), the differences are clear. Whilst *be like* has become the dominant quotative of choice among young speakers, at 50.1%, they still produced high numbers of *say*, using it 24.8% of the time, and did not use *be like* as in the same proportion as the older speakers used *say*. The *think* quotative was also used in smaller numbers by the younger generation, as it was used only 3.1% of the time. Intriguingly, the use of *tell* seems to have increased among the younger speakers, who used it 3.6% of the time, whereas, at 0.9%, the older speakers used it only as frequently as they used

be like and one of the other quotatives, *ask*. The younger generation also showed more frequent use of other quotatives, as they made up 5.8% of their database, but only 1.4% of the older generation's quotatives. However, both generations seemed to use *go* and zero quotatives in a similar proportion.

In order to better understand the changes in the younger generations' quotative systems, bar pie charts and bar charts are used to visualise trends within the database. Whilst this is a useful way to visualise the data, and graphs can indicate potential correlations within the data that may warrant further investigation, graphs alone cannot provide any conclusive or reliable results. Furthermore, the following bar charts reflect only the counts or percentages of quotatives in the database and should be interpreted with caution as an individual speaker's unusual use of quotatives may skew the results for an entire group of speakers under review. This is especially true for groups with lower numbers of participants.

Figure 4. 5

Bar chart of quotatives used by each participant

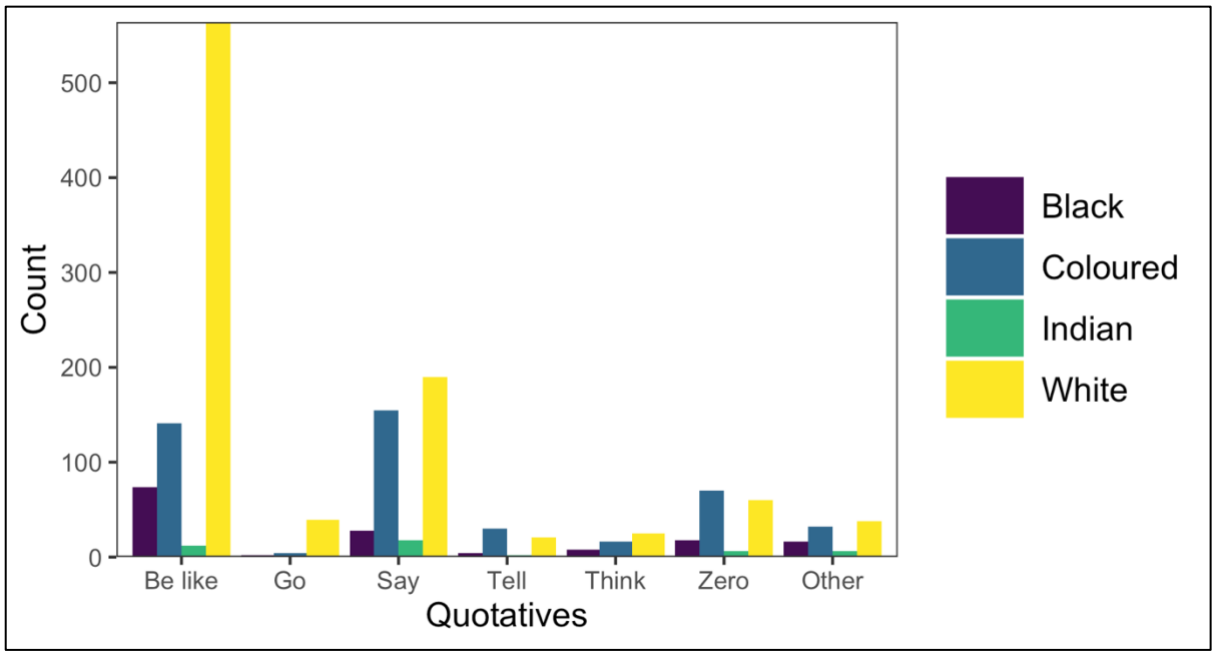


The great variety in the quantity of quotatives expressed by each participant is apparent in figure 4.5. The highest number of quotatives produced by a single participant was 129. In contrast, quotatives were absent in the speech of five interviewees from varied backgrounds:

males 4 (2 falling into the Indian and 1 each into the Coloured and Black categories; females 1 in the Coloured category). For those who did use quotatives, the mean number of quotatives produced was 30.4. However, this varied according for speakers of different races, gender and schooling. Among only young speakers, the mean number of quotatives produced was 29.8; among Black speakers it was 16.7; among Coloured speakers it was 23.6; among Indian speakers it was 6.3; among young White speakers it was 46.9 and among older White speakers it was 35.3.

Figure 4. 6

Bar chart showing the counts of main quotatives used by young speakers of different races

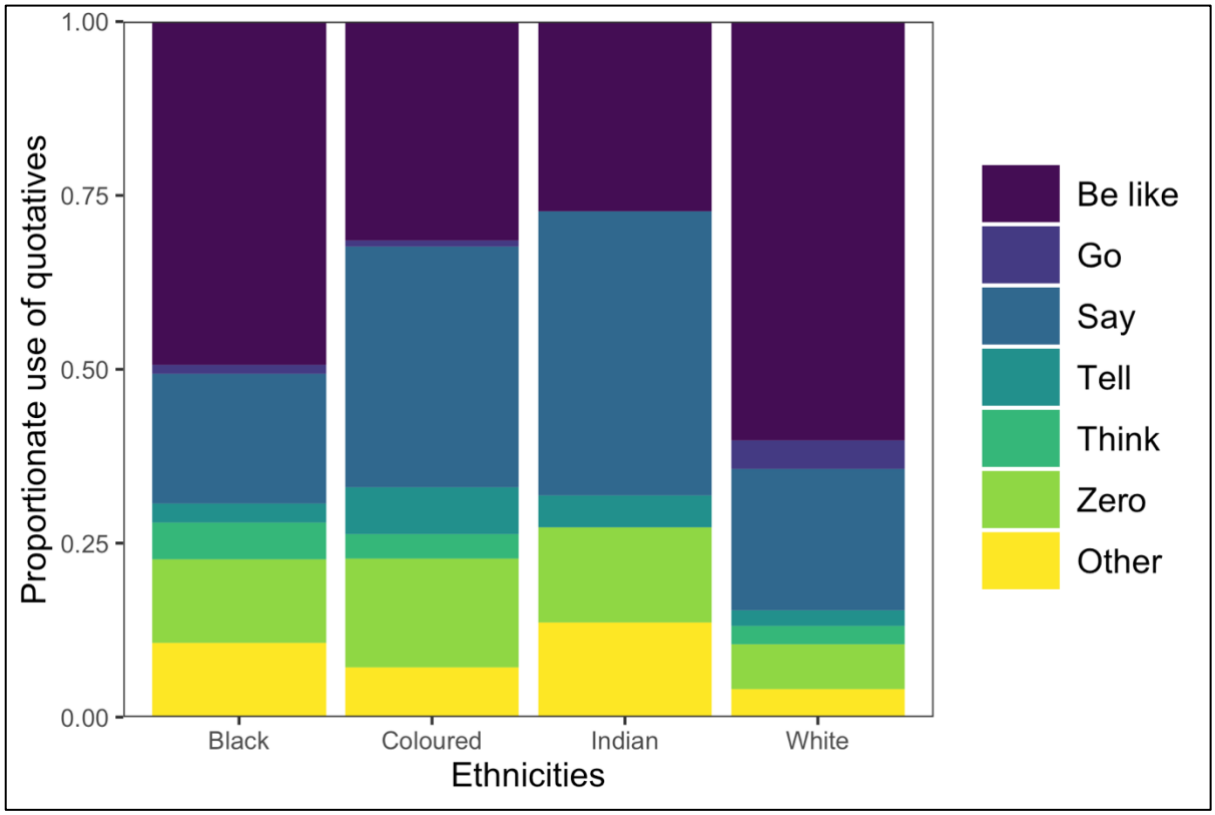


The differences in quotative count for different racial groups is also demonstrated in figure 4.6. Of the 1791 quotatives in the database, 212 were produced by the comparison group of older White speakers, 937 by the younger White speakers, 448 by Coloured speakers, 150 by Black speakers and 44 by Indian speakers. This is primarily due to the different numbers of participants in each racial group. Twenty participants were randomly selected for the young White speakers and young Coloured speakers, but there were only 8 Indian participants, 10 Black participants and 6 older White participants whose interviews were available for the purposes of this dissertation. The results for groups with fewer participants will be less statistically reliable than groups with a greater number of participants. For example, the fact that there are no columns for Indian participants for both *think* and *go*, would not be sufficient evidence to claim that speakers of Indian race in Cape Town never use the *think* or *go* quotatives, they simply were not expressed during the Indian participants' interviews.

4.1 Content of Quote

Figure 4. 7

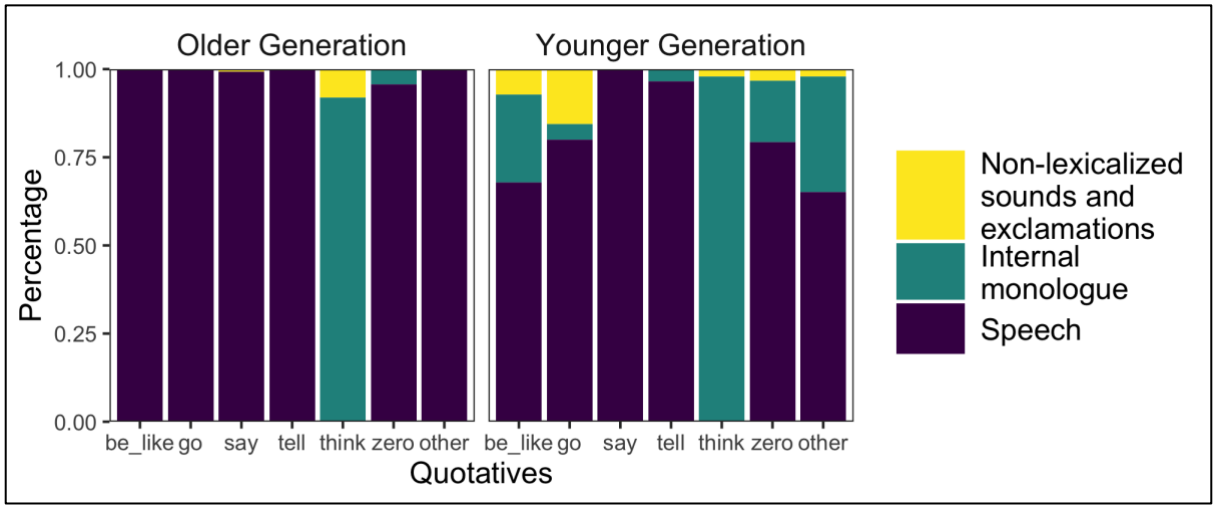
Bar chart showing the percentages of main quotatives used by young speakers of different races



In some cases, the bar charts will present the quotatives by percentage, rather than count, for easier comparison between different groups of speakers. Figure 4.7 presents the same information as figure 4.6, but more clearly displays the proportionate usage of each quotative for speakers of different races. Whilst *be like* is the most common quotative expressed by White and Black speakers, Indian and Coloured speakers seemed to favour *say* most of all quotatives. *Tell*, *think* and zero quotatives are most popular among Coloured and Indian speakers, whereas *go* is more common White speakers, followed Black then Coloured speakers and wasn't used at all by Indian speakers. Indian speakers, however, showed the highest expression of other quotatives, followed by Black, Coloured and then White speakers.

Figure 4. 8

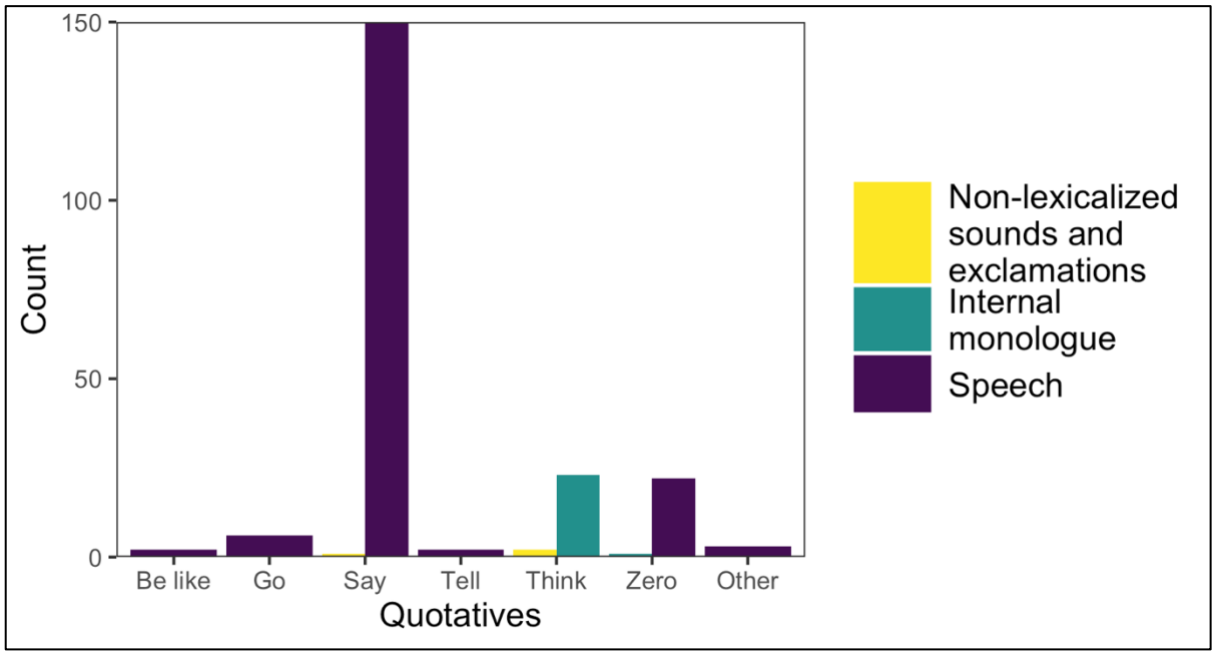
Bar charts showing contents of quote by percentage for each of the main quotatives used by older and younger participants



A comparison between the older and younger generations' quotative usage according to the *content of quote* (figure 4.8) reveals that in the time that speakers have adopted *be like*, they have also begun to use quotatives to express a greater range of *content of* quote. Where previously there had been a clear division between quotatives that expressed speech and quotatives that expressed inner monologue (thoughts may include exclamations as seen above), single quotatives are now being used to express internal monologue, speech, exclamations and non-lexicalised sounds. This indicates that a greater change is occurring in English quotative systems than simply a replacement of old quotatives with innovative ones.

Figure 4. 9

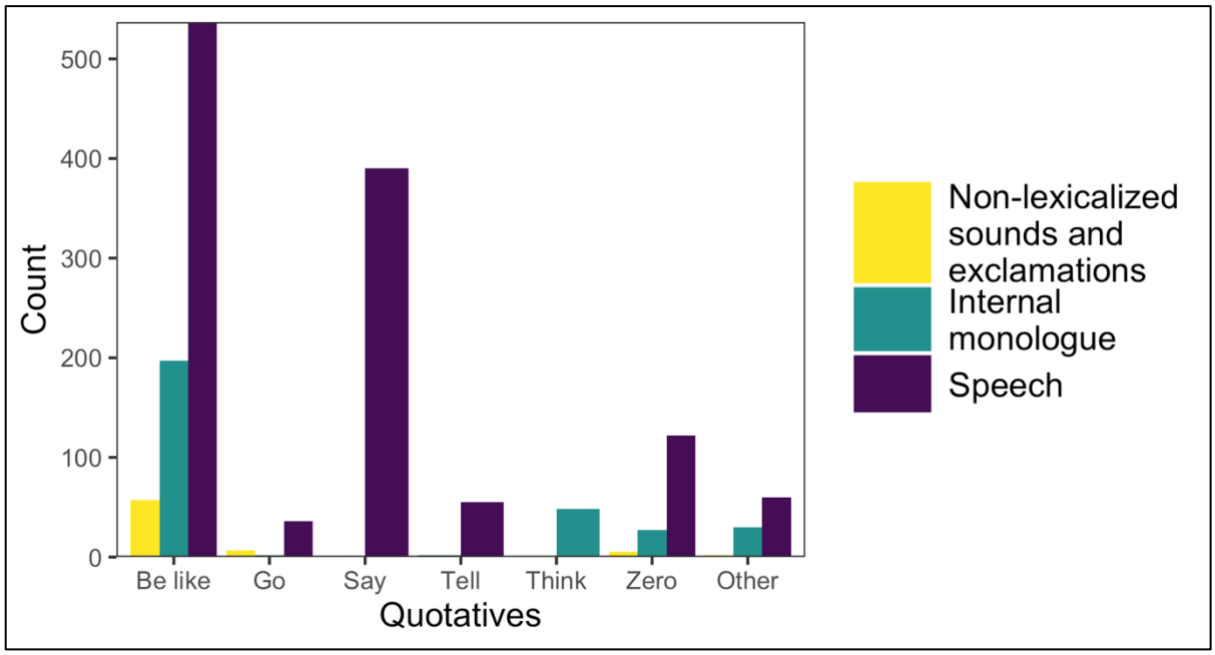
Bar chart showing the contents of quote for the main quotatives used by the older generation



The older generation expresses fewer quotes containing non-lexicalised sounds and exclamations than the younger generation. Such exclamations or non-lexicalised sounds were introduced either by the *say* or *think* quotatives. *Say* is the most commonly used quotative and primarily used to express speech. *Be like, go, tell* and other quotatives are only used to introduce quotes containing speech. *Think* is the predominant quotative used to express quotes containing internal monologue. Zero quotatives, whilst most commonly used to introduce quoted speech, were also used to quote internal monologue.

Figure 4. 10

Bar chart showing the contents of quote for the main quotatives used by the younger generation



The younger generation used *be like* in the highest numbers, followed by *say*. *Be like* was also the most common choice for all three quotative functions, with the highest number of quotes containing speech, internal monologue, and exclamations and non-lexicalised sounds. Thus, among the younger generation, *say* appears no longer to be the primary quotative for introducing quoted speech, just as *think* is no longer the primary quotative for introducing quoted internal monologue. *Go*, as well as zero and other quotatives also appear to be used to introduce all three contents of quote. Whereas *say* was used to introduce quoted speech and rarer cases of internal monologue, *think* was used to introduce quoted internal monologue and exclamations and non-lexicalised sounds, and *tell* was used to introduce quoted speech as well as a couple instances of quoted internal monologue.

When *tell* was used to quote internal monologue it occurred in the past tense with a first-person subject (39). Whereas, when *tell* was used to quote speech, the *grammatical person* of the subject and the *tense and temporal reference* both varied (40).

39. a. He is also all into soccer, and I ***told*** myself, "I'm going to make sure he gets out there where he's supposed to be."

    b. I ***told*** myself, "Don't worry."

40. a. I ***told*** him, "You're overworking yourself."

b. The people who used to work there, like, always used to ***tell*** me, "Oh, have you read this? Oh, you must read that."

As mentioned above the primary use of *think* is to quote internal monologue (41), but speakers from both the older (42) and younger (43) generations also used *think* to quote non-lexicalised sounds or exclamations.

41. I was so worried, I was ***thinking***, "Where is this child now?"

42. a. And I just ***thought*** to myself, "Sjoe!"

b. I used to be so energetic and now I just ***think***, "Aaahh…"

43. In that sense they're always ***thinking***, "Ewww."

Whilst *say* was primarily used to quote speech, the older generation also had two instances of *say* being used to quote exclamations and non-lexicalised sounds (44) and the younger generation had a couple instances of *say* being used to quote internal monologue (45).

44. They'd come along and ***say***, "Ah!"

45. So I just ***said*** to myself. "One month, only one month 'till the month of July…"

However, the older generation only used *go* to quote speech (46), whereas the younger generation used it to quote internal monologue (47) and exclamations and non-lexicalised sounds (48), in addition to speech:

46. She'd ***go***, "Oh no, that's simply ghastly!"

47. And your brain sort of ***goes***, "Ah!" [laughter].

48. I'm ***going***, "Okay, this is going to be interesting…"

Both instances of *be like* expressed by speakers of the older generation (49) were used to quote speech. However, it is interesting to note that in one of the cases, *be like* was accompanied by dummy *it* (49 b). As discussed previously, *be like* was the most common choice, among the younger cohort, for the expression of quotes containing speech (50), internal monologue (51) and exclamations and non-lexicalised sounds (52).

49. a. In the evenings when we got home my mother ***was like***, "How('re) you gonna make me laugh?"

b. If you're gonna take on this job, people are going to button you hard all the time, you know, it***'s like***, "I just thought I'd pop in and see when you're busy…"

50. a. My dad and my sister ***were like***, “Come swim with us!”

b. And my sister, when she speaks to White people, she***’s like***, “Yes.” [mimetic re-enactment].

51. a. So many people laughed and I ***was*** just ***like***, “There could be someone in the audience whose father had the same thing.”

b. It ***was like***, “What the fuck are you talking about?”

52. a. I ***was lik****e,* “Woah!”

b. You got to Joburg and you***’re like***, “Ooohhhh shit.”

c. And they ***were like***, “Mmm, ah.”

Like *be like* and *go,* zero quotatives and other quotatives were also used to introduce quotes containing internal monologue (53), speech (54) and exclamations and non-lexicalised sounds (55).

53. a. Then I ***realise***, “Oh, damn,” like, “maybe this isn’t a good idea.”

b. So she just ***decided***, “Ugh, I’m tired of being here.”

c. So, I’m just surprised about that. ***Ø*** "Oh, you guys like school né? OK, that’s good.”

54. a. So I’m ***askin****g* her, “What month is it?”

b. They’ll ***greet,*** “Hey, how are you?”

c. I was describing the smell, ***Ø*** "It's like a met- cold metally smell, you know, it must have been the detergents they use to clean the place.”

d. Obviously we're not being forced in the sense someone is putting a gun to your head, ***Ø*** "Hey you have to be friends with her."

e. My brother came running, like, ***Ø*** “Oh my gosh, Ania,” you know, “Where have you been?”

f. You know, to stand next to that person. That you’ve been there for through and through. And, um, ***Ø*** “Do you want to be my bridesmaid?”
Ø “YAYY!"

55 a. Then I ***realise***, “Oh, damn…”

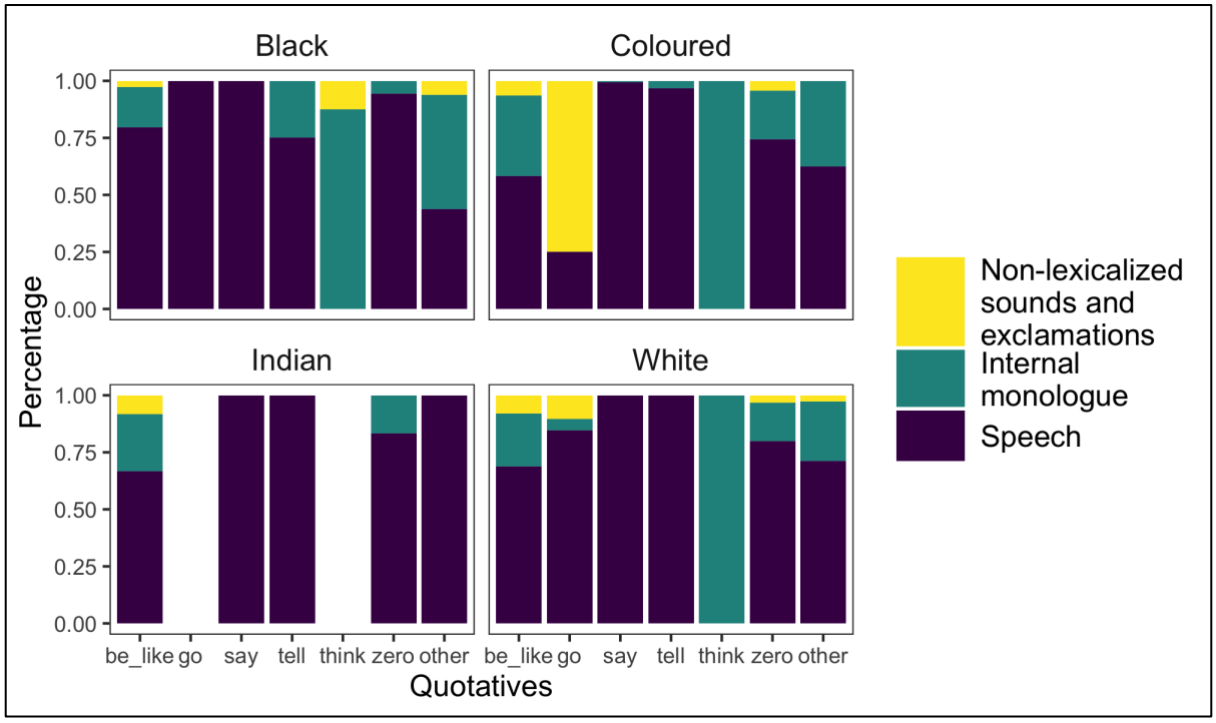
b. I ***mean***, getting to matric… I ***mean***, like, “Woah…”

c. That play we did [...] we also did a concert with dance, **Ø** {Singing melody} "Hey, hey, baby…"

d. Yes, she's making so, **Ø** **{**imitates sound}

Figure 4. 11

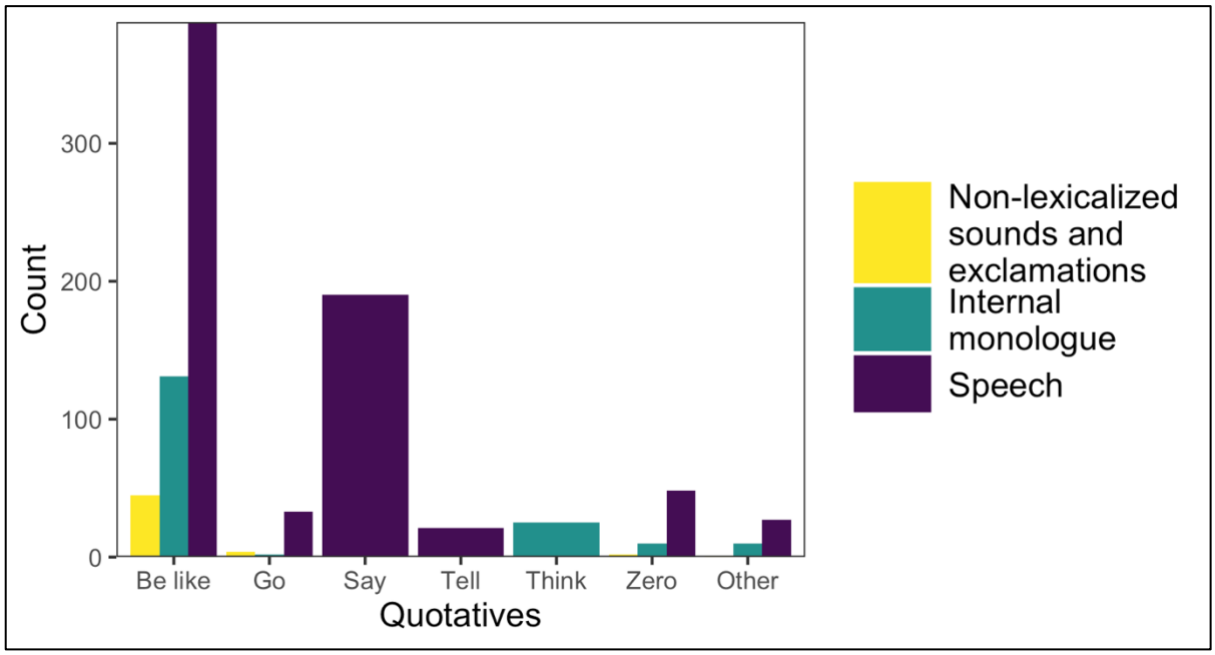
Bar charts showing contents of quote by percentage for each of the main quotatives used by younger participants of different races



Young speakers of all races used *be like* to express all three *content of quote*s and in similar proportions. *Be like* was most commonly used for speech, then for internal monologue and, finally, for a small number of non-lexicalised sounds and exclamations. In addition to *be like*, White and Coloured speakers used zero quotatives to introduce all three contents of quote in a similar proportion. The same was true for the use of other quotatives by Black and White speakers. However, Black and Indian speakers only used zero quotatives to introduce speech and internal monologue, Coloured speakers used other quotatives to quote thought and speech and Indian speakers used other quotatives only to quote speech. *Say* was used solely to introduce speech by all participants. White and Indian participants also used *tell* only to introduce speech, but Coloured and Black participants used *tell* to quote internal monologue in some cases (as in above examples). Black speakers used *think* to quote internal monologue which also contained some non-lexicalised sounds and exclamations, whereas White and Coloured speakers used it only to quote internal monologue.

Figure 4. 12

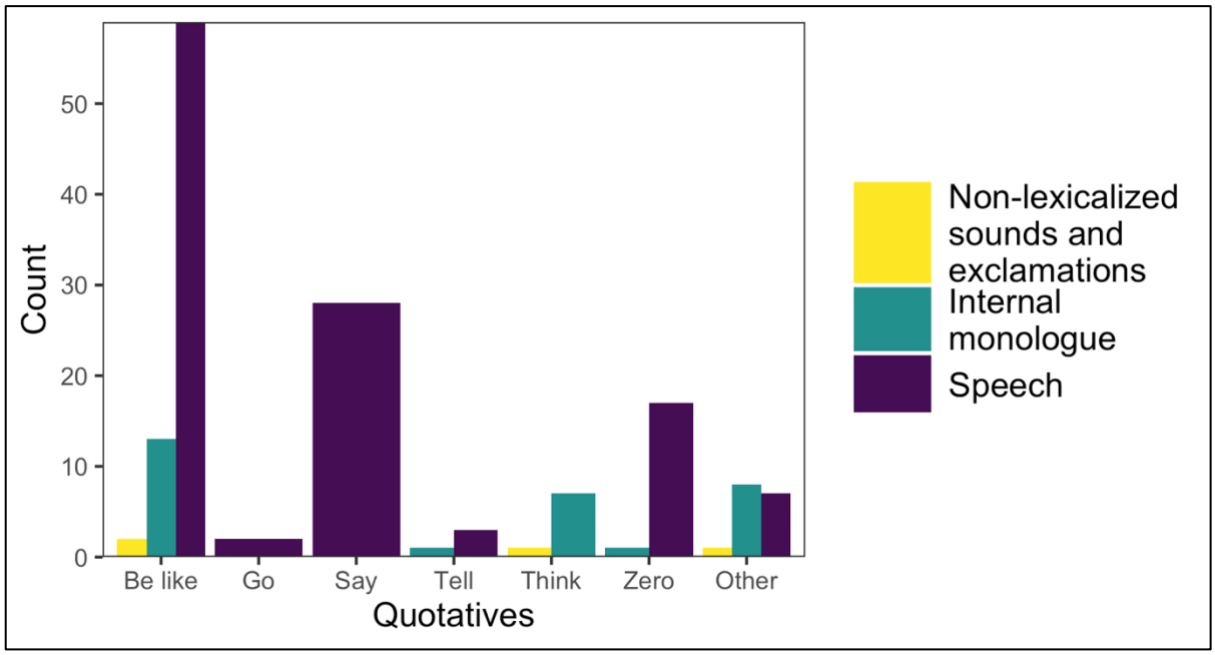
Bar chart showing the content of quote for the main quotatives used by young White speakers



White speakers, like all other young speakers most often quoted direct speech, followed by internal monologue and only quoted a few non-lexicalised sounds and exclamations (figure 4.12). Most quoted speech was introduced by *be like*, with *say* as the second most common quotative used to introduce speech, followed by zero, *go*, other quotatives and finally, *tell*. *Be like* was also the most frequently used quotative for introducing internal monologue, followed by *think*, zero and other quotatives, and a few instances of *go*. The majority of on-lexicalised sounds were quoted using *be like*, but *go*, zero and other quotatives were occasionally used.

Figure 4. 13

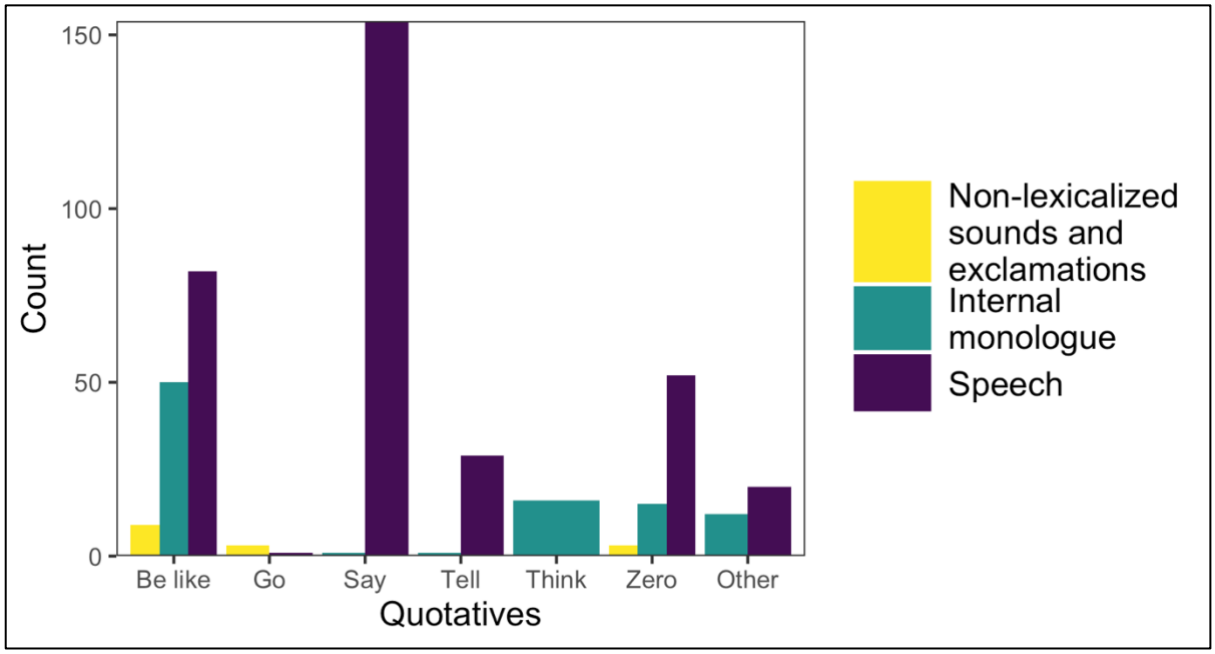
Bar chart showing the content of quote for the main quotatives used by young Black speakers



Be like was also the most commonly used quotative, regardless of *content of quote*, for young Black speakers (figure 4.13). Most quoted speech was introduced by *be like*, followed by *say*, zero quotatives, other quotatives, *tell* and, finally, *go*. Internal monologue was most frequently quoted through the use of *be like*, followed by other quotatives, *think*, *tell* and zero quotatives. *Be like*, *think* and zero quotatives were used to quote non-lexicalised sounds and exclamations.

Figure 4. 14

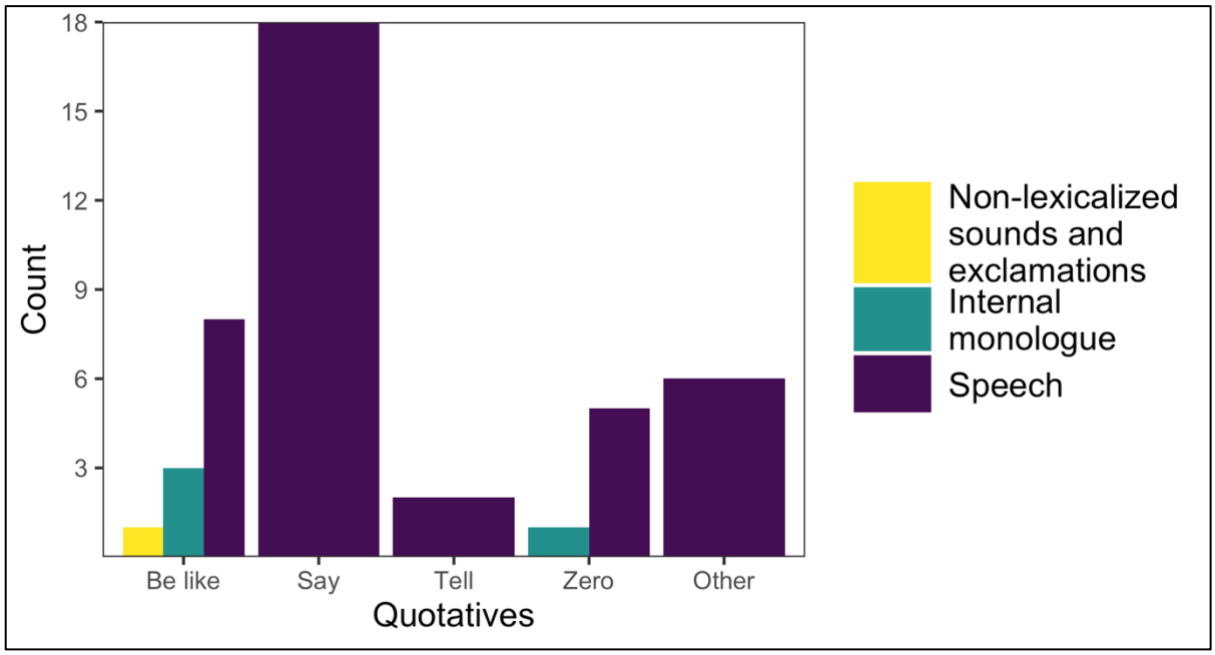
Bar chart showing the content of quote for the main quotatives used by young Coloured speakers



Whilst *be like* was the most commonly used quotative for the introduction of quotes containing internal monologue and non-lexicalised sounds, Coloured speakers primarily used *say* to quote direct speech (figure 4.14). *Be like* was the second most popular choice for quoting speech, followed by zero quotatives, *tell*, other quotatives and *go*. Internal monologue was most frequently quoted using *be like*, then *think*, zero quotatives, other quotatives and *tell*. After *be like*, non-lexicalised sounds and exclamations were introduced in equal numbers by *go* and zero quotatives.

Figure 4. 15

Bar chart showing the content of quote for the main quotatives used by young Indian speakers[8]



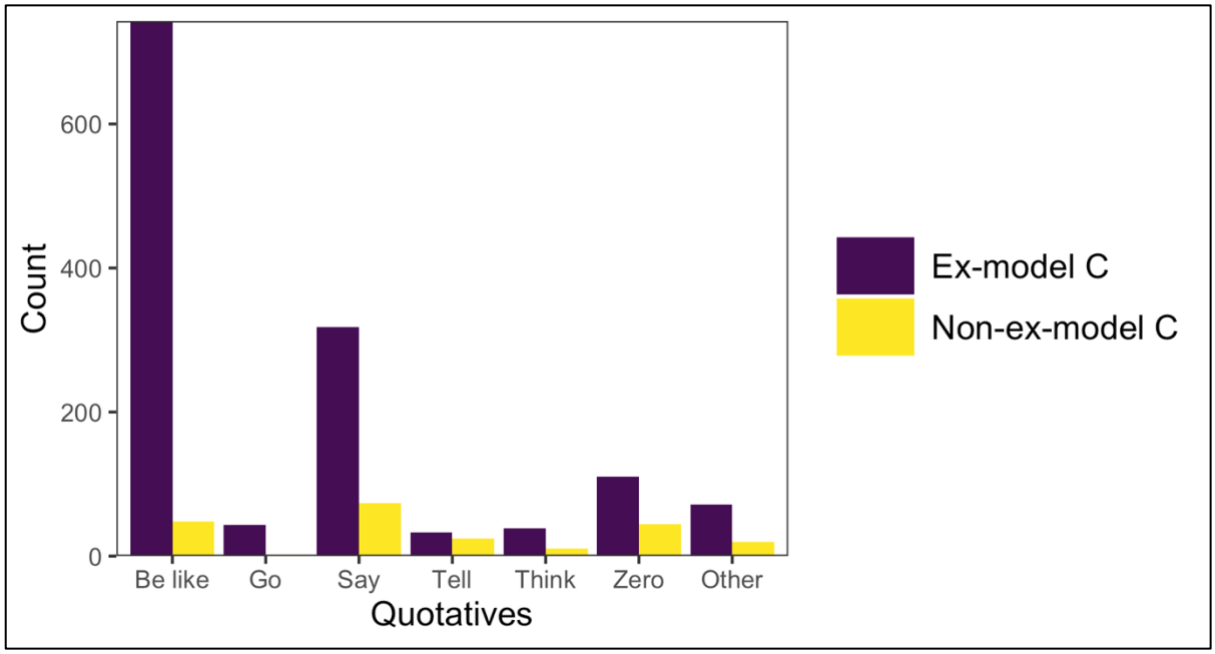
Say was also the preferred quotative for quoting speech during the interviews with Indian speakers (figure 4.15). *Be like* was the second most popular choice followed by other quotatives, zero quotatives and *tell*. Internal monologue was primarily expressed using *be like*, but zero quotatives were also used. Non-lexicalised sounds and exclamations were only quoted through the use of *be like*.

[8] Column width may vary for the bar charts, as they are divided according to the levels of each variable, and not all levels were present in the data for Indian speakers.

4.2 Schooling

Figure 4. 16

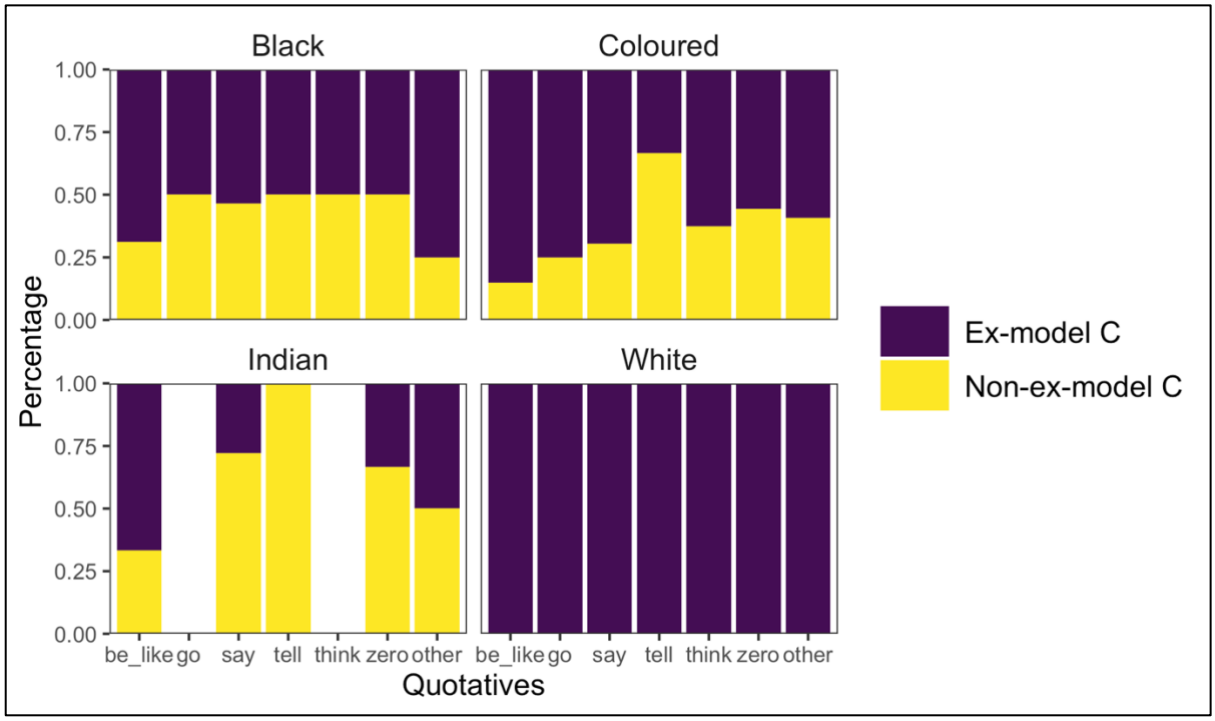
Bar chart of speaker schooling for the main quotatives used by young participants



Schooling was recorded only for the younger participants, as the older participants would have attended school during apartheid. Figure 4.16 demonstrates the higher numbers of tokens produced by participants who had attended FMC schools. Of the Coloured participants,15 of 21 attended FMC schools; of the Black participants, four of ten attended FMC schools; and of the Indian participants, two of eight attended FMC schools. The younger White speakers only attended FMC schools. Looking at the younger population as a whole, it seems speakers who attended FDET schools used *say* most frequently, followed by *be like*, then zero, *tell*, other, *think*, *go*. Whereas speakers who did attend FMC schools used *be like* in the highest numbers, followed by *say*, zero, other, *go*, *think* and *tell*.

Figure 4. 17

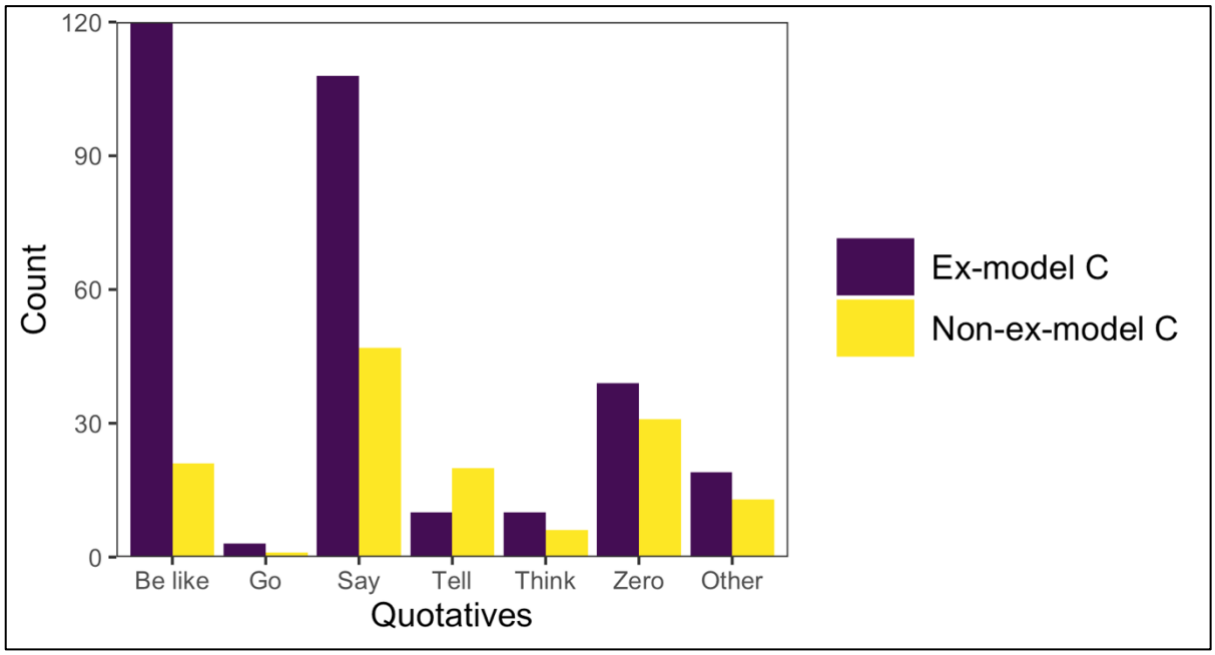
Bar charts showing schooling by percentage for each of the main quotatives used by younger participants of different races



An examination of the quotatives produced by the four groups, further divided by schooling, suggests that, for all groups, most instances of *be like* were expressed by speakers who attended FMC schools. However, for Indian and Coloured participants, most instances of *tell* were expressed by speakers who went to FDET schools. After *tell*, *say* and zero quotatives were expressed by a greater proportion of Coloured and Indian speakers who attended FDET schools, than *go*, *think* and other quotatives were expressed by the participants in those racial groups. As there is little to analyse for the White participants who attended only FMC schools, what follows is a closer look at quotatives and schooling Indian, Black and Coloured participants.

Figure 4. 18

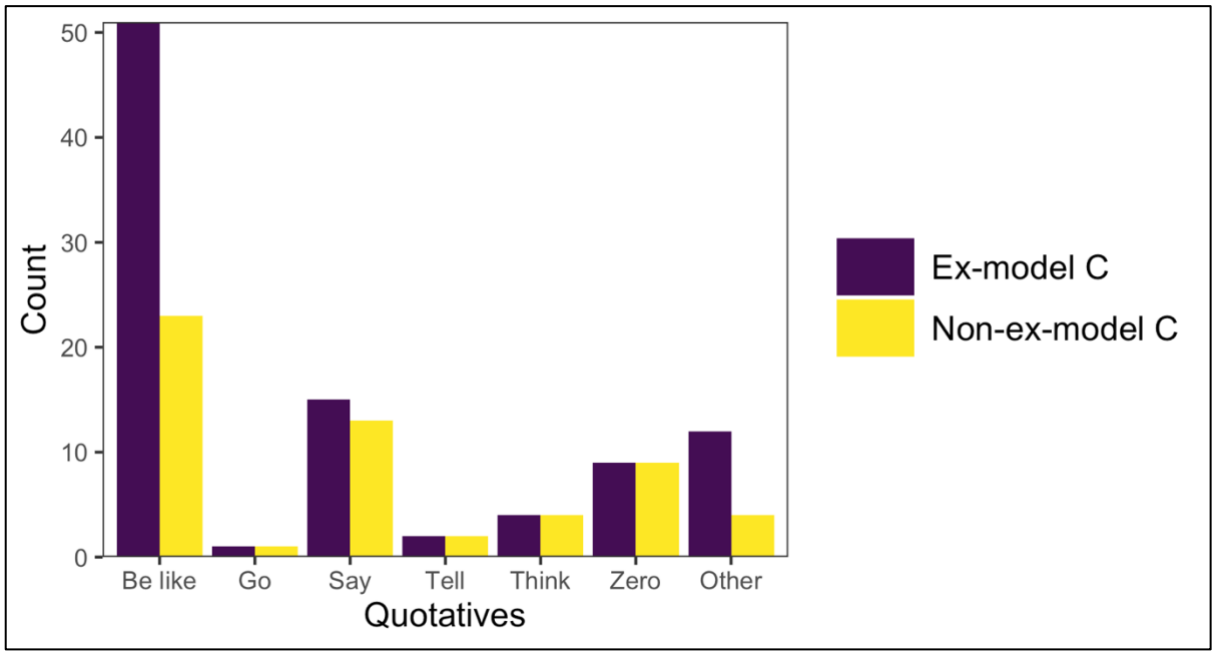
Bar chart of speaker schooling for the main quotatives used by young Coloured participants



Amongst the Coloured participants, all quotatives other than *tell* were used in greater numbers by speakers with FMC schooling than those with FDET schooling (figure 4.18). For those who attended FMC schools, *be like* was the most commonly used quotative, followed by *say*, zero quotatives, other quotative, *think* and *tell*, and, finally, *go*. Students with FDET schooling, used *say* most frequently, followed by zero quotatives, *be like*, *tell*, other quotatives, *think* and *go*.

Figure 4. 19

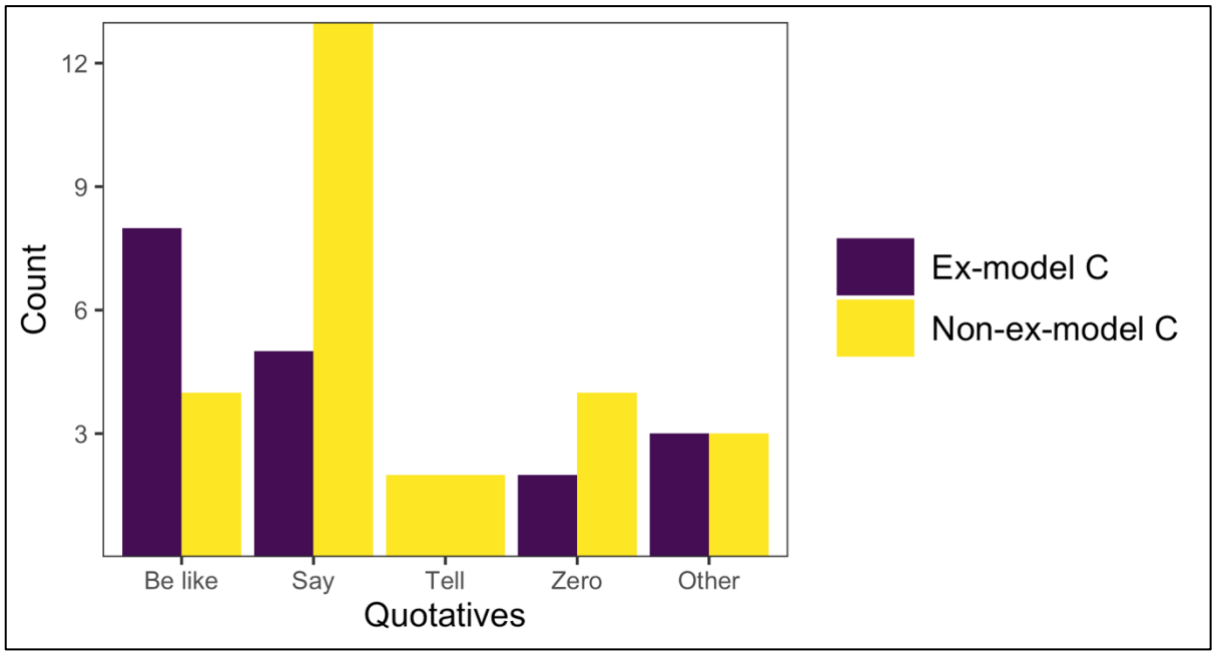
Bar chart of speaker schooling for the main quotatives used by young Black participants



Interestingly, among Black participants, most instances of *be like*, *say* and other were produced by participants who attended FMC schools, but *go*, *tell*, *think* and zero were expressed in equal numbers by both FMC and FDET students (figure 4.19). Speakers who went to FMC schools used *be like* more than other quotatives, followed by *say*, other zero, *think*, *tell* and *go*. Speakers who went to FDET schools also used *be like* in the highest proportion of all quotatives, followed by *say*, then zero, and equal number of *think* and other, then *tell* and *go*.

Figure 4. 20

Bar chart of speaker schooling for the main quotatives used by young Indian participants

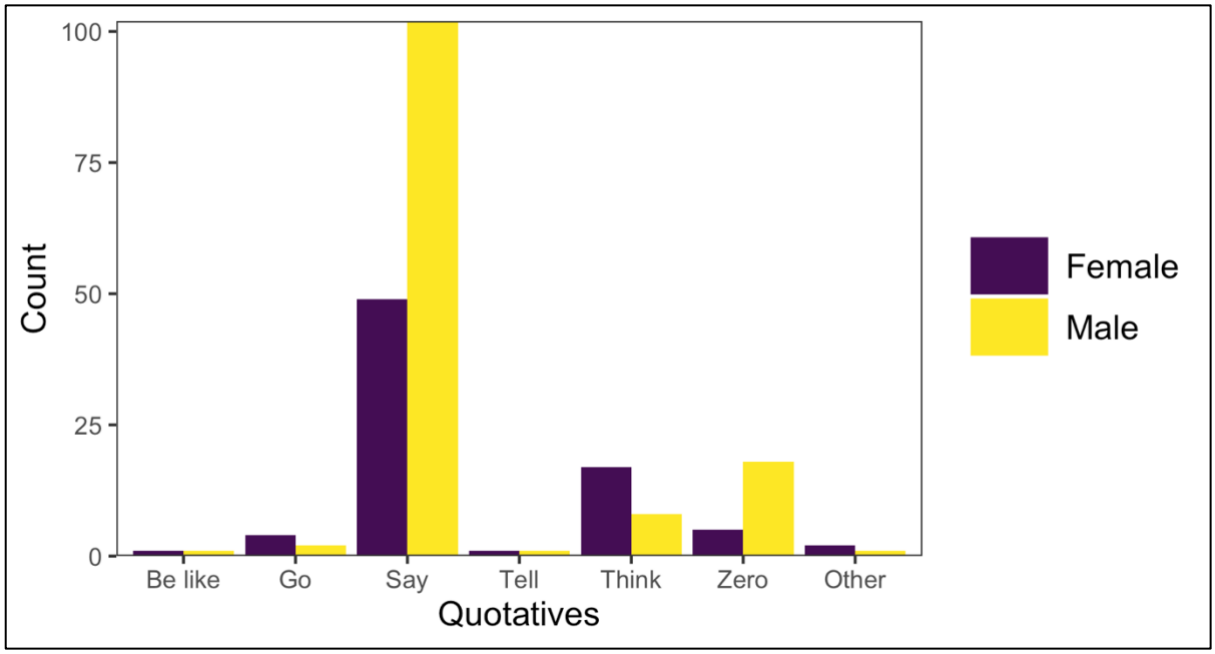


Immediately observable among the Indian speakers, is that *tell* was solely produced by speakers who attended FDET schools (figure 4.20). Students with FDET schooling also showed a strong preference for *say* and used an equal number of *be like* and zero quotatives, followed by other quotatives and then *tell*. Whereas students with an FMC schooling preferred the use of *be like*, followed by *say*, other quotatives and, finally, zero quotatives.

4.3 Gender

Figure 4. 21

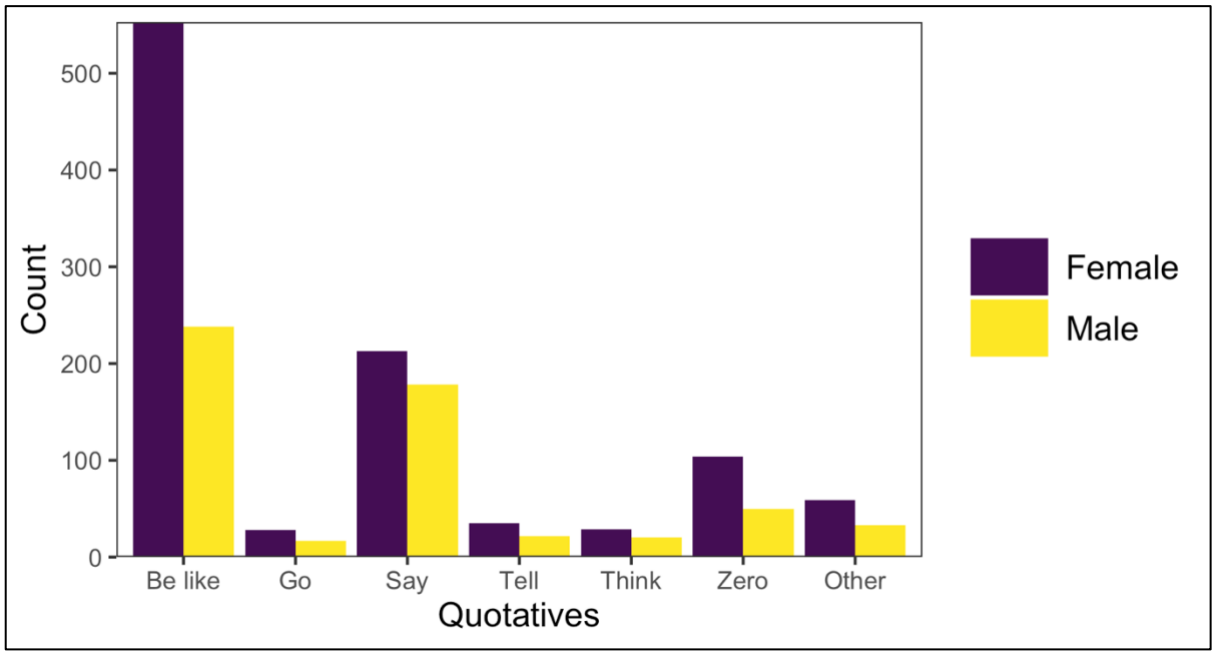
Bar chart of speaker gender for the main quotatives used by older participants



Amongst the group of older White speakers, which consisted of two men and four women, the majority of quotatives was expressed by men (figure 4.21). However, although men used *say* and zero quotatives more than women, women used *think, go* and other quotatives more than men did. There was equal usage of both *be like* and *tell* by speakers of either gender.

Figure 4. 22

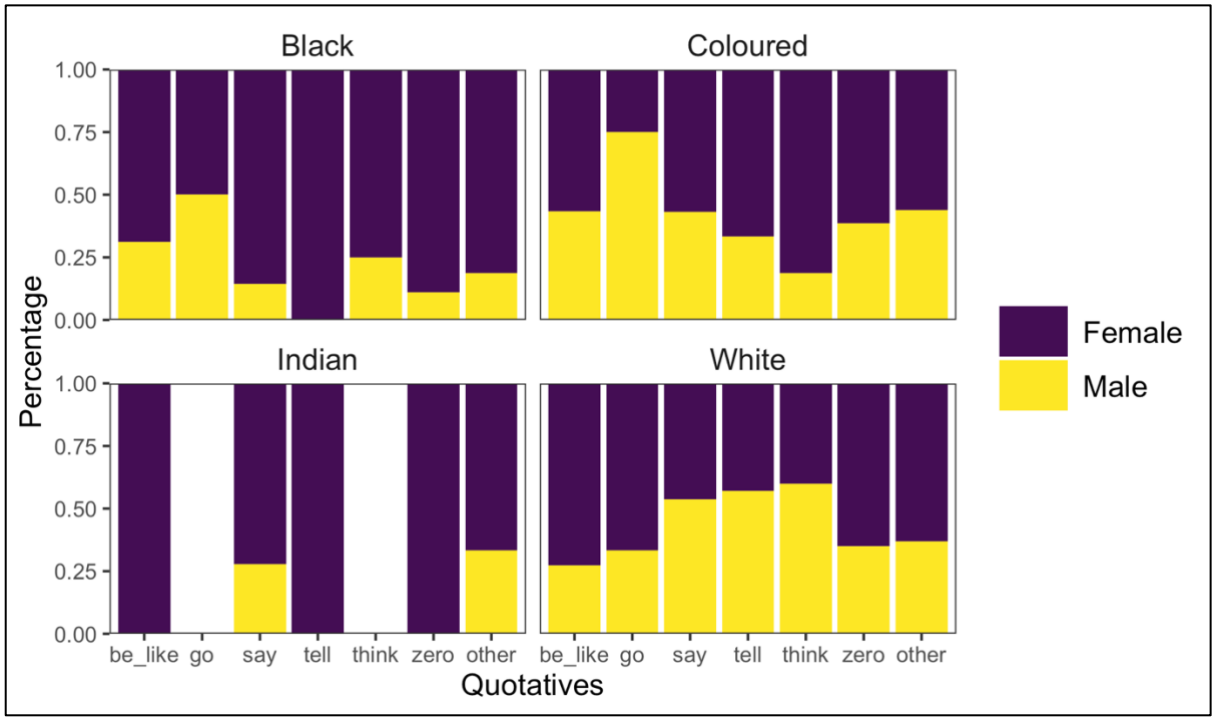
Bar chart of speaker gender for the main quotatives used by young participants



Amongst the younger speakers, all quotatives were more frequently expressed by female participants than by male participants (figure 4.22). This could be due partly to the uneven numbers of participants of each gender, as there were 31 young female participants and 28 young male participants. However, for both sexes, the most frequently used quotative was *be like*, followed by *say*, zero quotatives, other quotatives, *tell*, *think*, and, finally, *go*.

Figure 4. 23

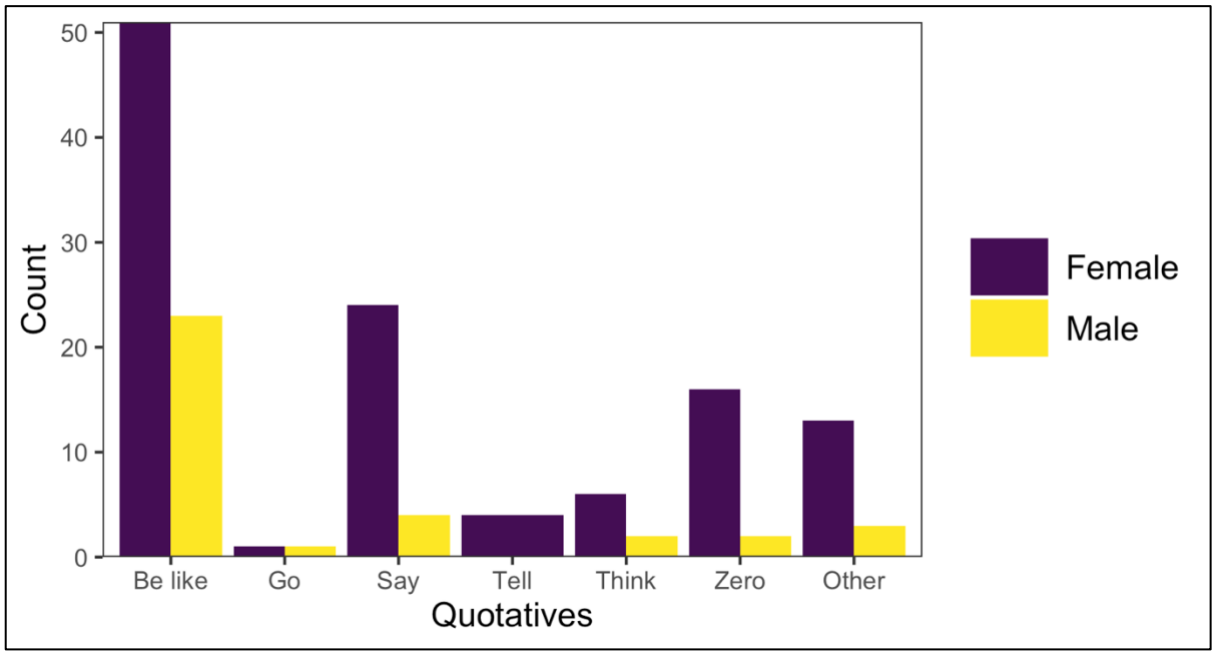
Bar charts showing speaker gender by percentage for each of the main quotatives used by younger participants of different races



In figure 4.23, the disparity in the graphs by gender for the Black and Indian groups needs to be read against the dispratiy in numbers (cited previously). The only quotatives expressed by male Indian participants were *say* and other quotatives, and even for these a greater number was produced by female Indian participants. Black male participants used all quotatives other than *tell*, but once again the majority of each quotative other than *go* - which had equal distribution between the sexes - was expressed by women. White and Coloured speakers were more equally distributed, with eleven men and ten women for Coloured participants, and ten men and ten women for White participants. Amongst White speakers, the majority of *be like*, *go*, zero and other quotatives were expressed by women, whereas the majority of *tell*, *think* and *say* were expressed by men. Amongst Coloured speakers, immediately noticeable was the way that men expressed a higher proportion of the instances of *go*, whereas all other quotatives (and particularly *think* and *tell*) were produced in greater numbers by female speakers.

Figure 4. 24

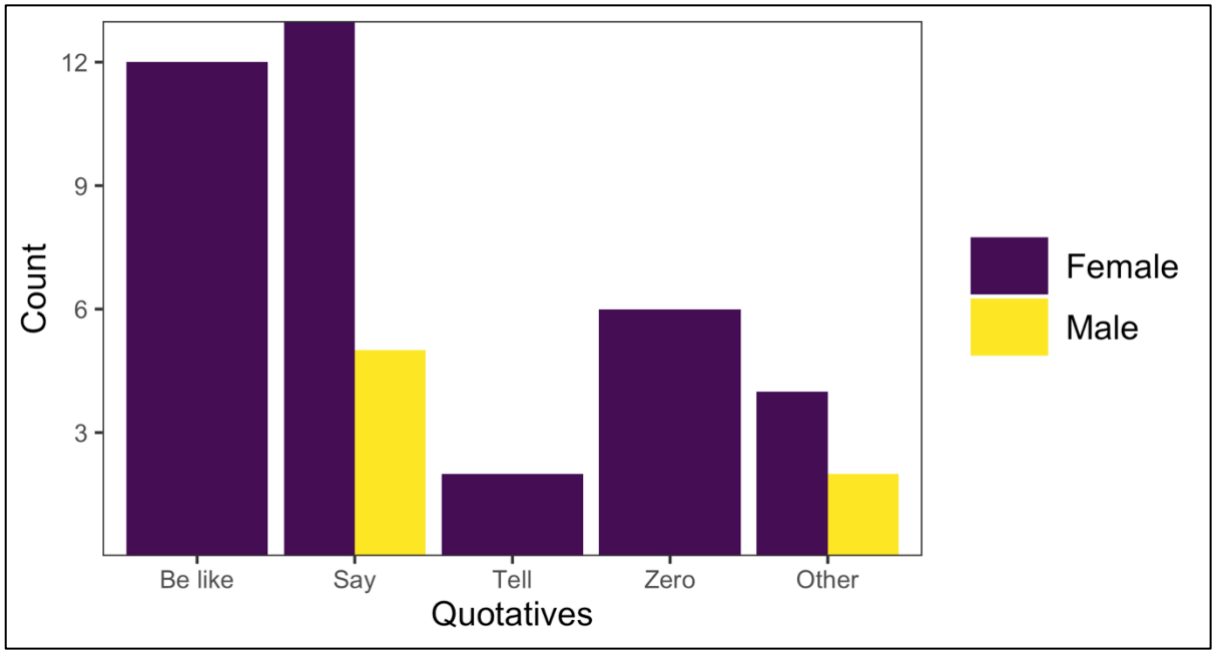
Bar chart of speaker gender for the main quotatives used by young Black participants



The greater number of female participants is apparent in figure 4.25, as all quotatives other than *go* were more frequently used by female than male participants. Women used *be like* most often, then *say*, zero quotatives, other quotatives, *think*, *tell* and *go*. Men used *be like* most often, then *say*, other quotatives, zero quotatives, *think* and *go*, but didn't use *tell*.

Figure 4. 25

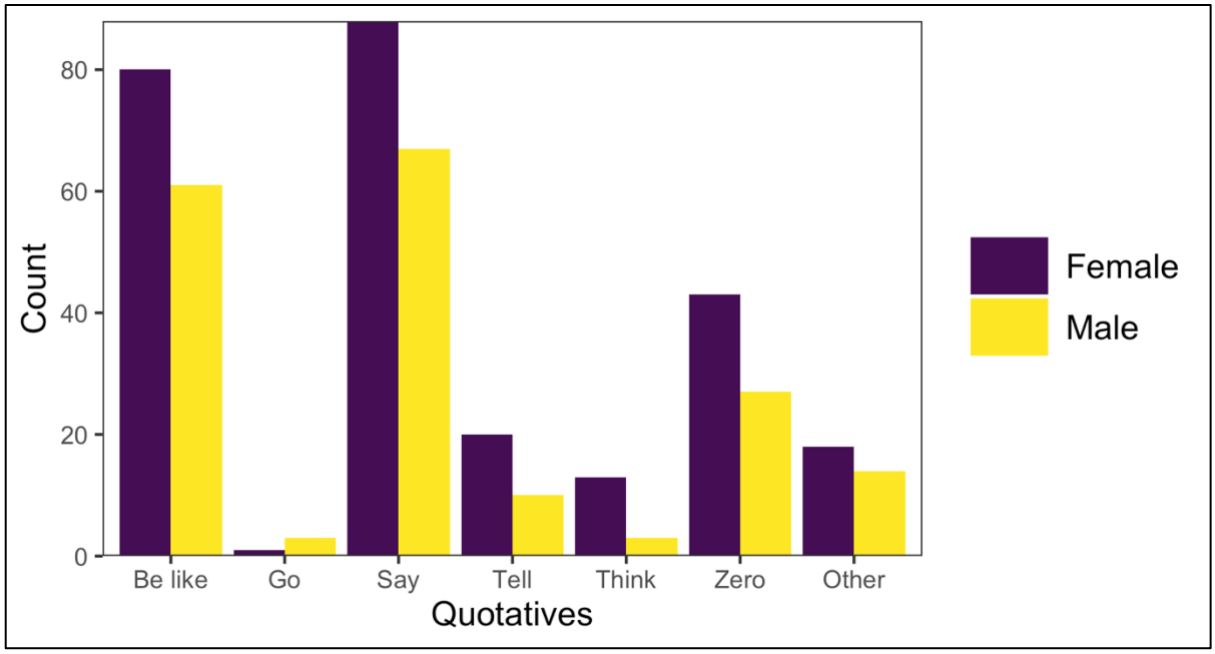
Bar chart of speaker gender for the main quotatives used by young Indian participants



Amongst the Indian participants, both male and female speakers used *say* most frequently of the quotatives. For women, this was closely followed by *be like*, then zero quotatives, other quotatives and *tell*. As there was a small number of men, only *say* and other quotatives were used by male speakers.

Figure 4. 26

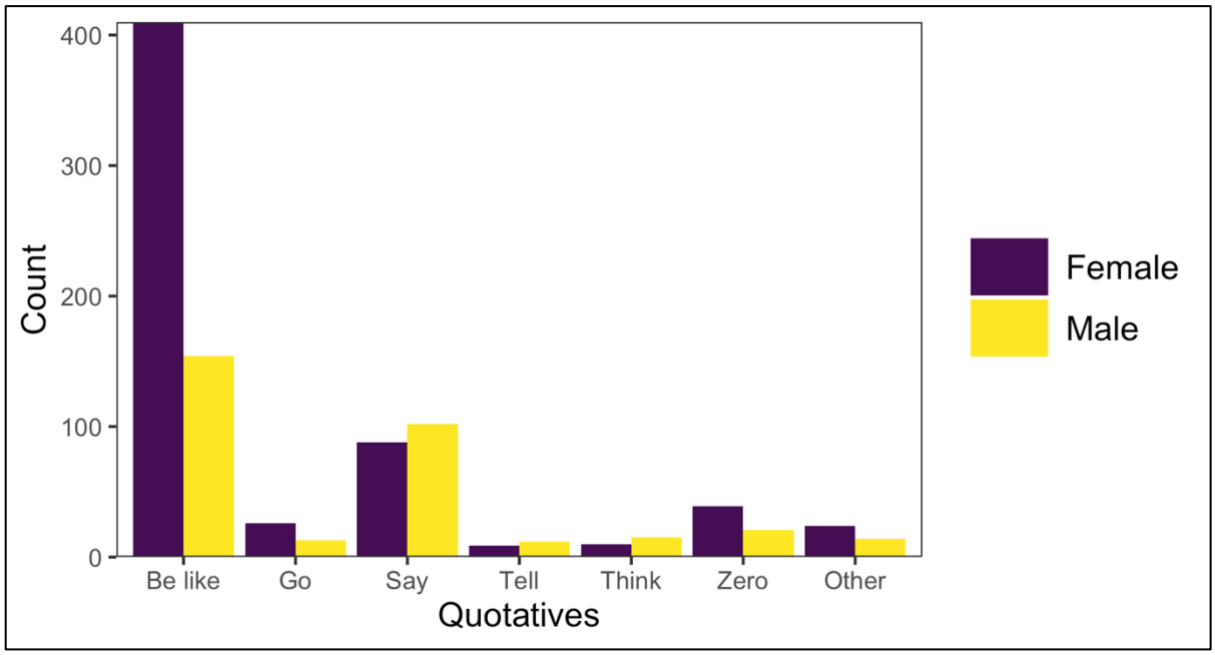
Bar chart of speaker gender for the main quotatives used by young Coloured participants



Within the group of Coloured participants, both men and women favoured *say* of all the quotatives, followed by *be like*, zero, other and *tell* (figure 4.26). For the remaining quotatives, women used *think* more than *go*, but men used both *think* and *go* equally often.

Figure 4. 27

Bar chart of speaker gender for the main quotatives used by young White participants



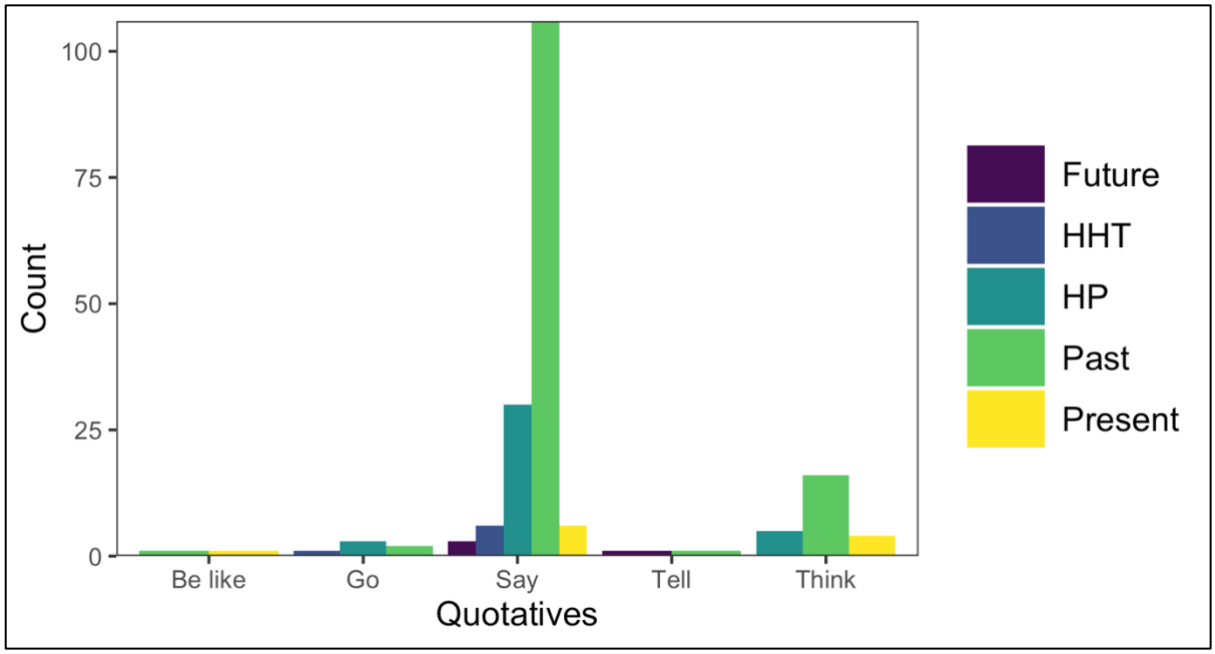
Of the young White participants, men and women used *be like* most frequently of the quotatives, followed by *say* and zero quotatives (figure 4.27). For the female speakers, these were followed by other quotatives, *go*, *think* and *tell*, but for male speakers these were followed by *think*, other quotatives, *go* and *tell*, in decreasing order.

The remaining bar charts do not include zero and other quotatives, choosing only to focus on the 5 most commonly used quotatives, other than zero quotatives, as zero quotatives did not always include the necessary *tense and temporal reference* and *grammatical person* information to be accurately coded.

4.4 Tense

Figure 4. 28

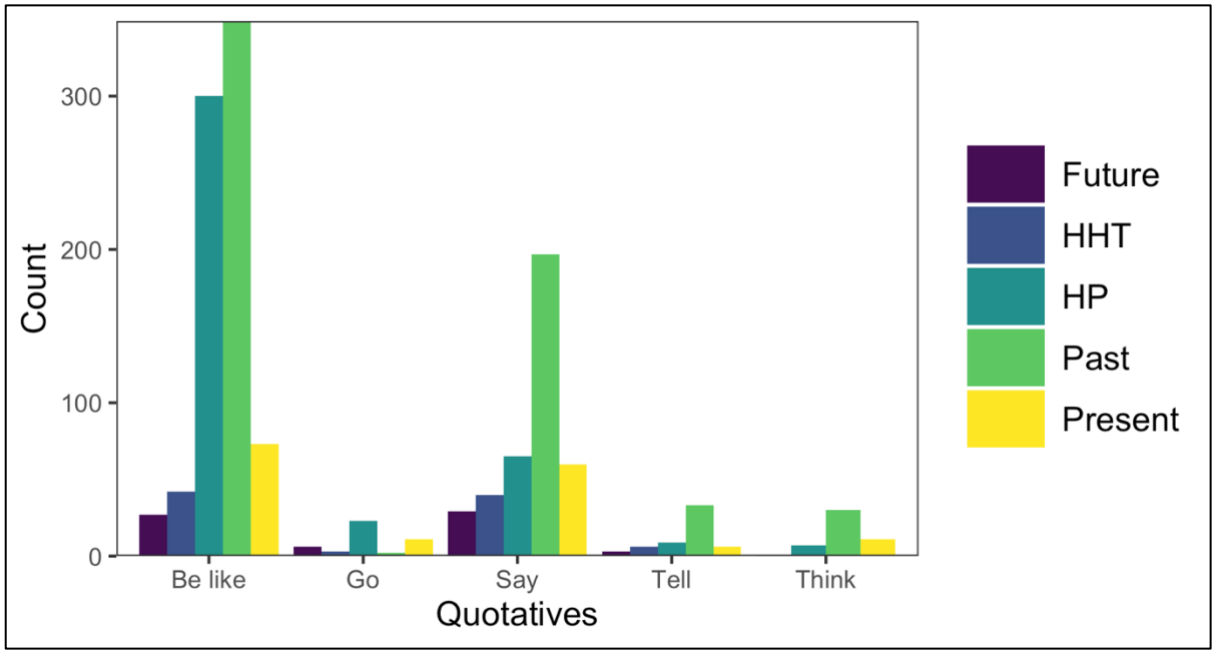
Bar chart of tense and temporal reference for the main quotatives used by older participants



The older generation most frequently used the past tense in their quotative phrases (figure 4.28). For *say*, their most commonly used quotative, this was followed by HP. Present tense and HHT were less frequently used than HP and future tense was least frequently used.

Figure 4. 29

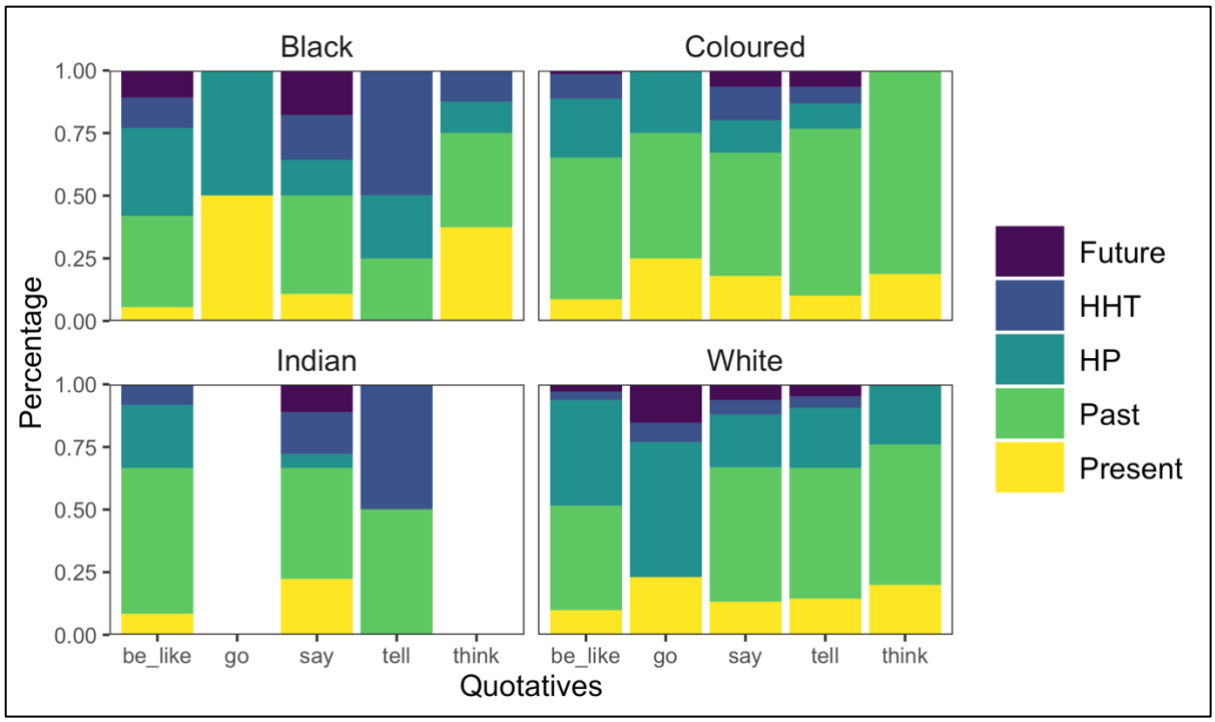
Bar chart of tense and temporal reference for the main quotatives used by younger participants



The younger participants also showed a strong preference for the use of past tense in their quotative phrases (figure 4.29). However, for *be like*, after past tense there is a very high count of HP, especially when compared to the *tense and temporal references* in which both the younger and older generations expressed quotative phrases containing *say* (figures 4.28 and 4.29). The younger generation also shows a higher usage of future tense than the older generation for *say*, *go* and *be like*.

Figure 4. 30

Bar charts showing tense and temporal reference by percentage for each of the main quotatives used by younger participants of different races

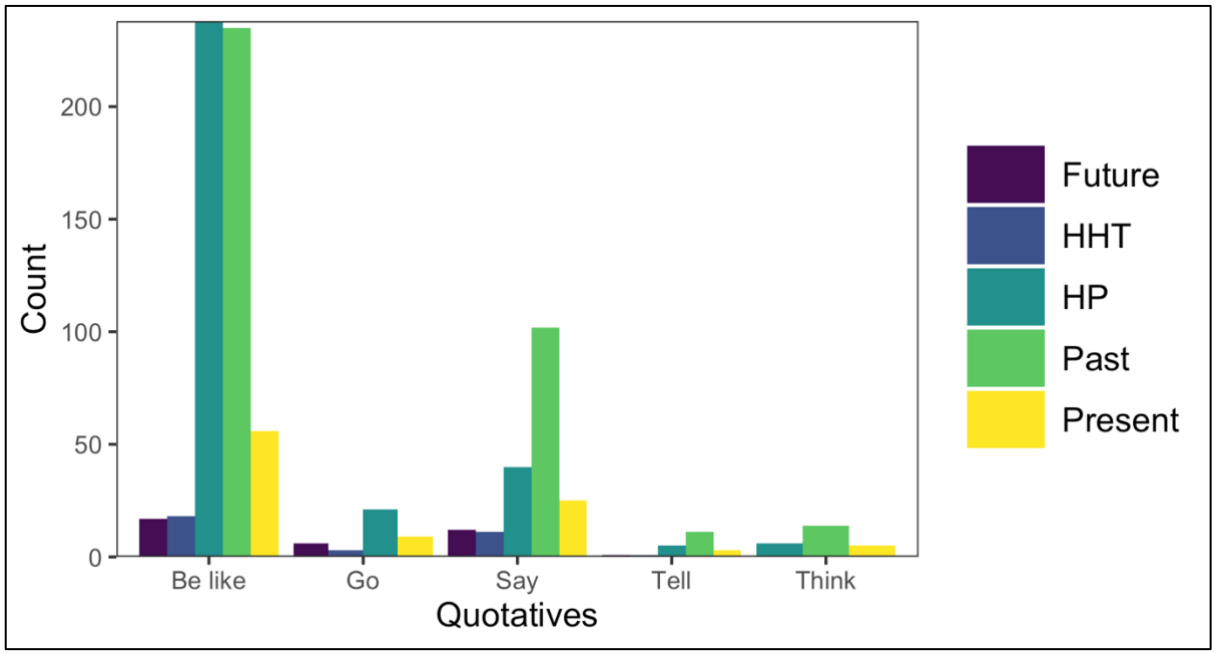


As was established above, it is clear that past tense is most commonly used for quotative phrases expressed by both younger and older participants. The highest instances of *be like* occurring in an HP quotative phrase were for White and Black students. The proportion of *be like* quotatives that occurred in HP phrases, was smaller for both Indian and Coloured students, for whom there was a greater quantity of *be like* in past tense phrases. *Go* also showed a high rate of occurrence in HP phrases, again especially for White and Black speakers, who did not use *go* in past tense quotative phrases at all. However, the incidence of *go* was lower than all other main quotatives, which may be why the proportion of different *tense and temporal references* may seem more exaggerated or to vary greatly from those of other quotatives. There was a higher count of HHT phrases, but not future or present tense phrases, for *tell* for both

Black and Indian students. This too, maybe due to a lower number of quotative entries as they were the two groups with the fewest participants. HHT was less common than present tense and HP. However, all three of the main quotatives expressed by Indian speakers occurred in HHT phrases, all main quotatives expressed by Black speakers, other than *go*, also occurred in HHT phrases and, amongst the Coloured speakers, *be like*, *go* and *tell* were used in HHT phrases. In general, the percentages of HHT phrases containing each of the main quotatives seemed to be lower for White speakers than speakers of other race groups, despite the fact that HHT occurred with *be like*, *go*, *say* and *tell* for White speakers. Future tense was least commonly used, with White speakers expressing *be like*, *go*, *say* and *tell* in future tense phrases, Black speakers expressing *be like* and *say* in future tense phrases, Coloured speakers expressing *say* and *tell* in future tense phrases and Indian speakers only using the future tense for phrases containing *say*.

Figure 4. 31

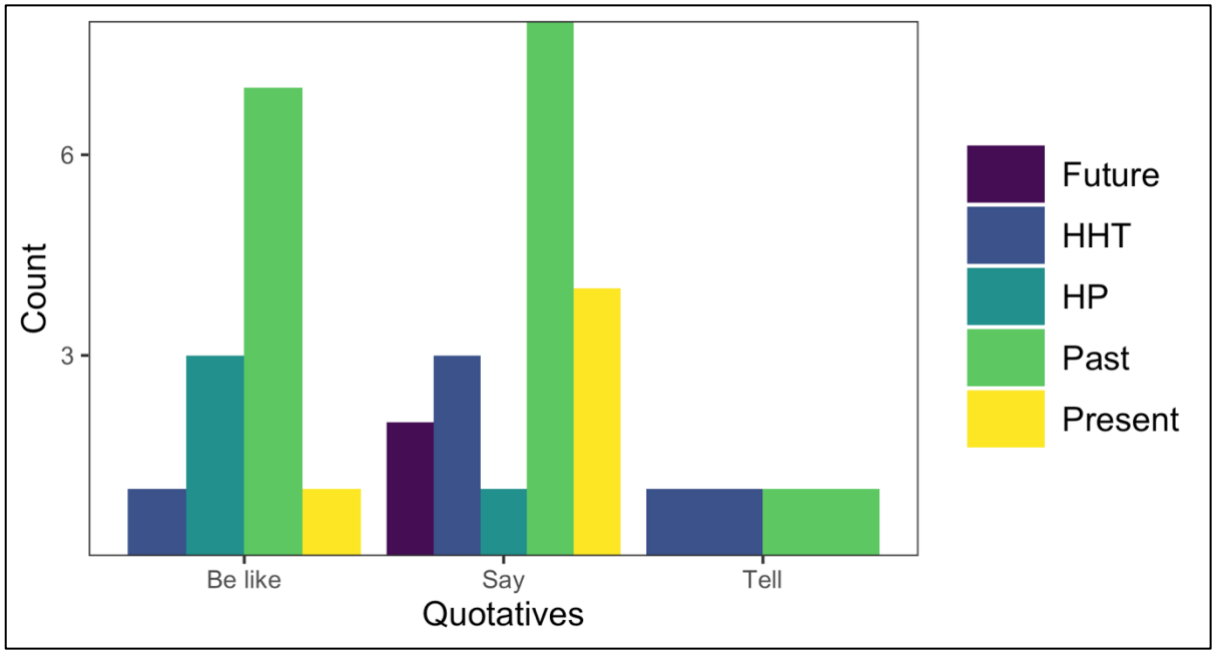
Bar chart of tense and temporal reference for the main quotatives used by young White participants



Young White speakers used *be like* more frequently in HP quotative phrases than they did in past tense quotative phrases (figure 4.31). HP was also the most common choice for expressing quotative phrases containing *go*.

Figure 4. 32

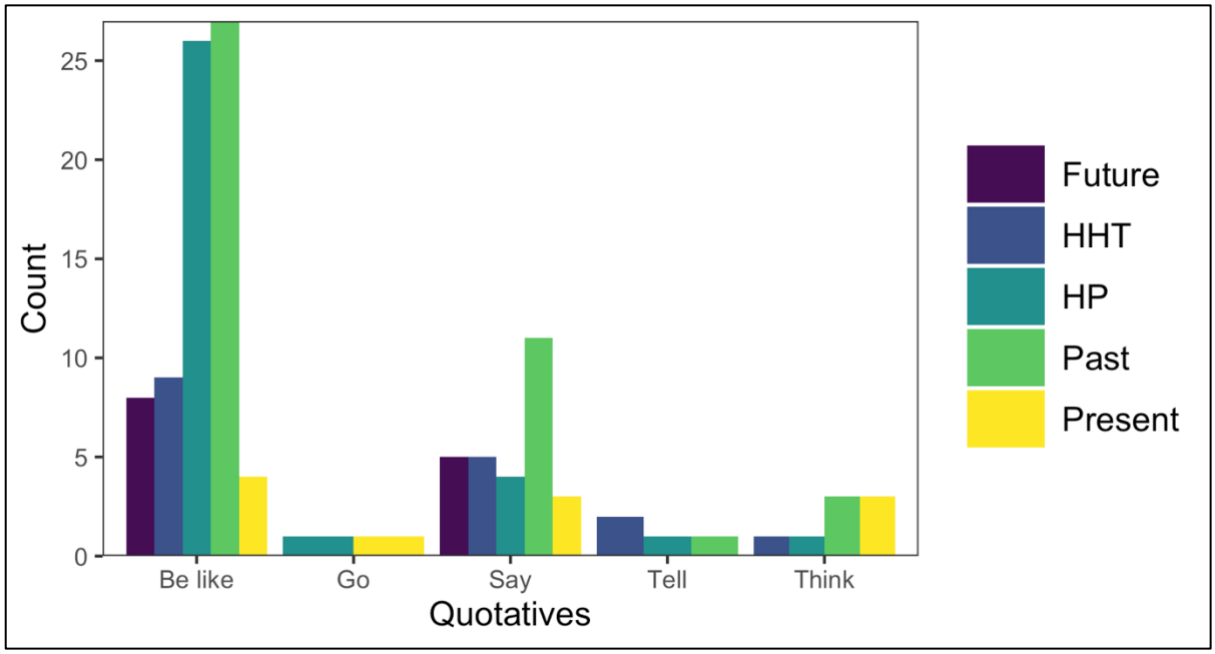
Bar chart of tense and temporal reference for the main quotatives used by young Indian participants



Indian speakers did not use HP as often whilst expressing quotative phrases containing *be like* or *say* (figure 4.32). However, there was a high occurrence of present tense quotative phrases expressing *say*, as it was the second most frequently used tense for *say*.

Figure 4. 33

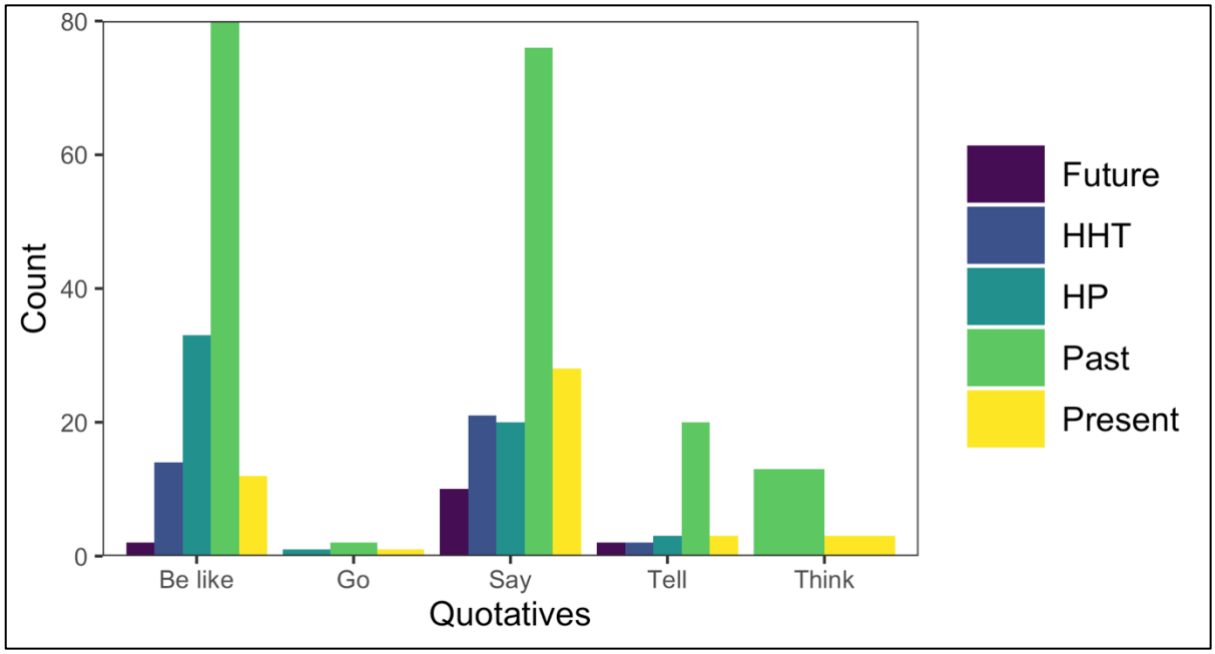
Bar chart of tense and temporal reference for the main quotatives used by young Black participants



Black participants also showed a very high proportion of HP quotative phrases for *be like*. Interestingly, other than for *think* and *go*, present tense seemed the least popular choice among Black speakers, who showed a higher occurrence of HHT and future tense for *say* and *be like*.

Figure 4. 34

Bar chart of tense and temporal reference for the main quotatives used by young Coloured participants

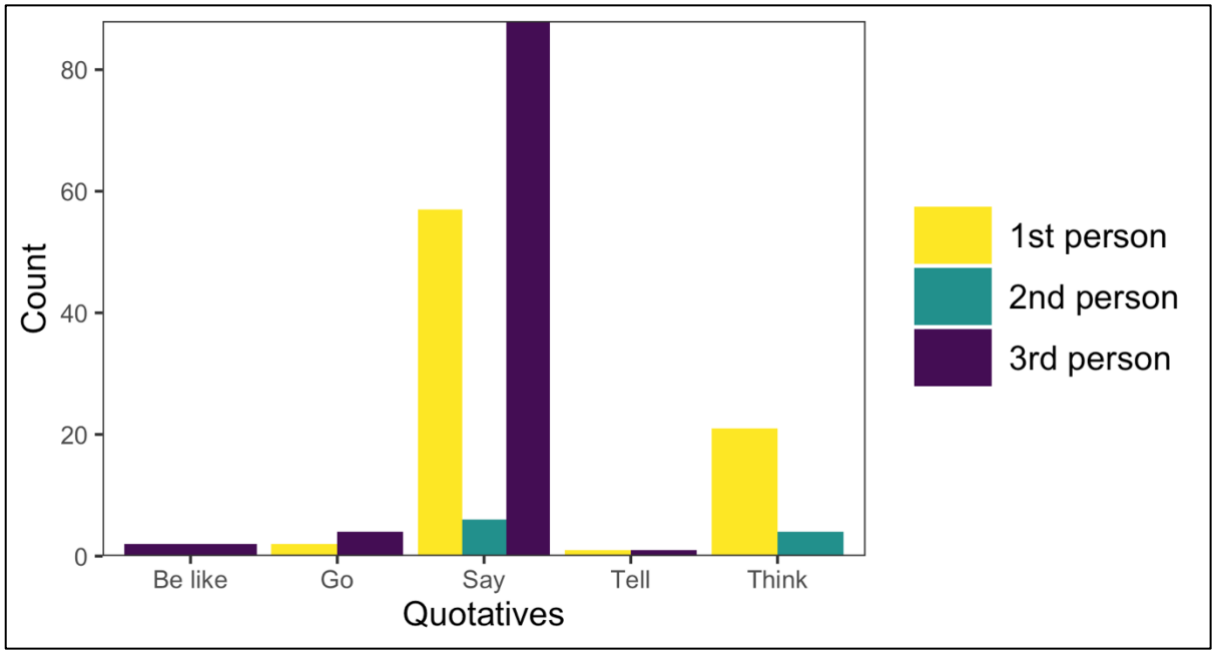


Whilst there was a fairly high count of HP phrases containing *be like* for Coloured speakers (figure 4.34), that number was not as high as those for Black or White speakers. Similarly to Indian speakers, present tense was the second most frequently used tense for quotative phrases containing *say*.

4.5 Grammatical Person

Figure 4. 35

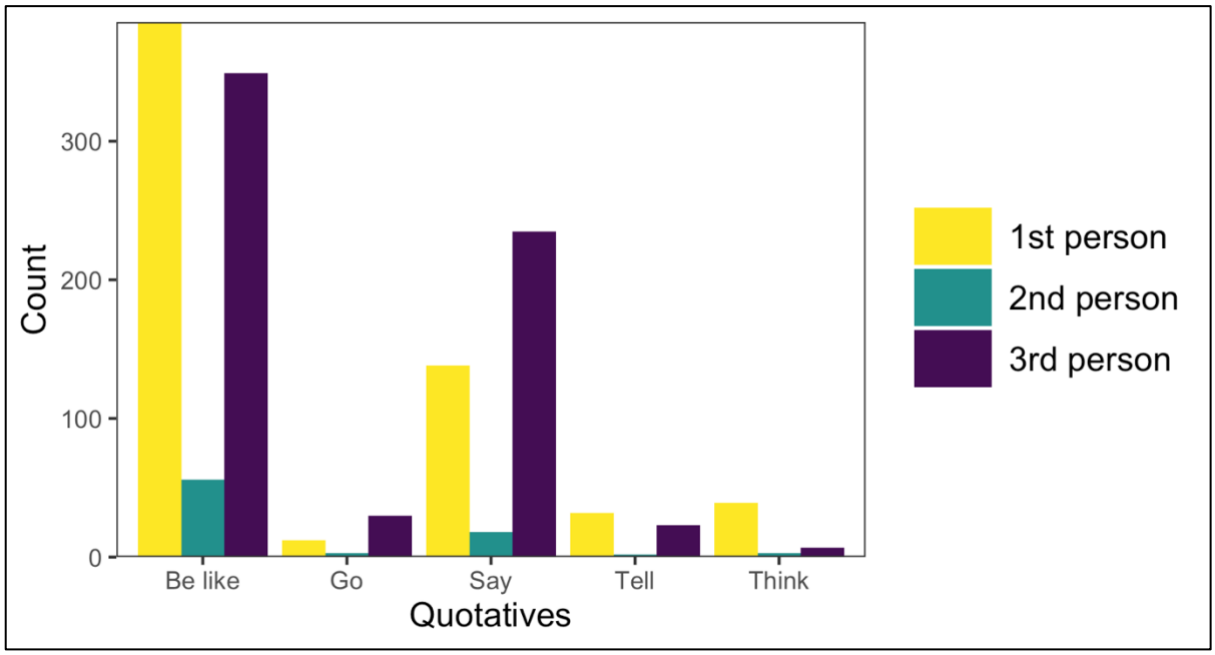
Bar chart of grammatical person for the main quotatives used by older participants



The older generation favoured third-person subjects for *be like*, *go*, *say* and *tell*. As could be expected from quotatives used only to express internal monologue, *think* was not expressed alongside any third-person subjects and favoured the first person instead. There were small numbers of second-person subjects for both *say* and *think*.

Figure 4. 36

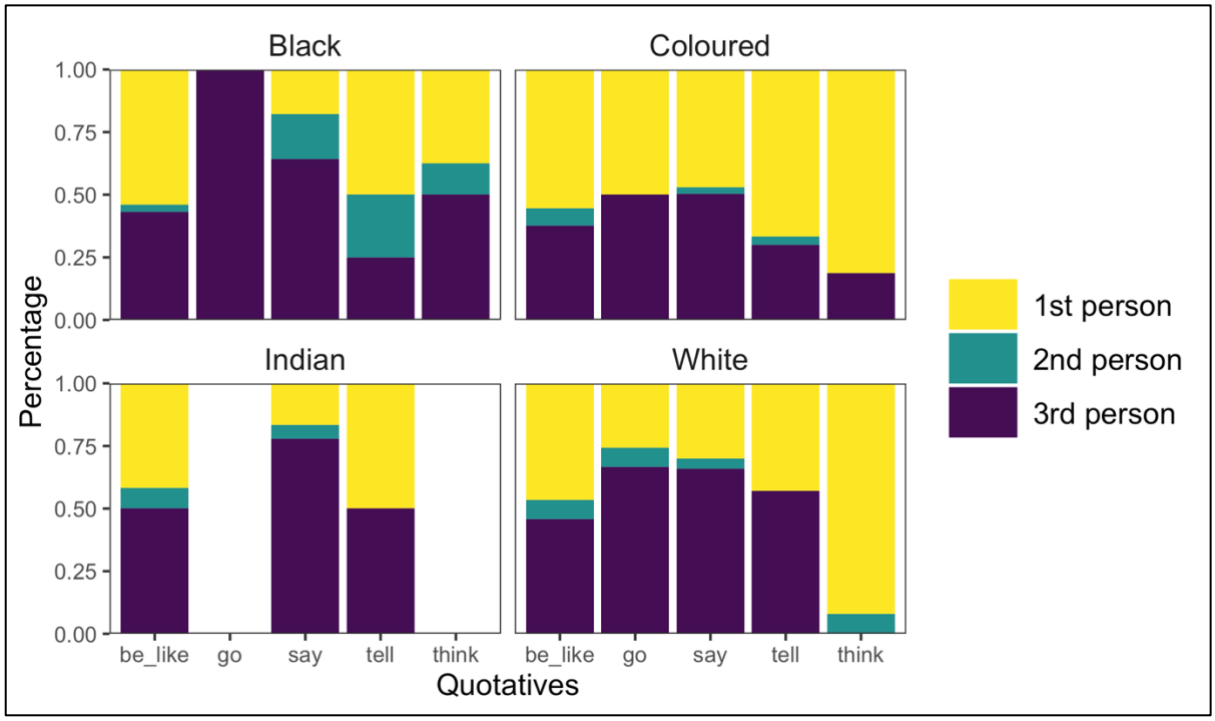
Bar chart of grammatical person for the main quotatives used by younger participants



Among the younger generation, first-person subjects were the preferred choice for *be like*, *tell* and *think*, whilst third-person subjects were favoured for *say* and *go*. For *tell* and *be like* there was only a small difference in the counts of first- and third-person subjects, but there were considerably more first-person subjects than third-person subjects for *think*. Perhaps due to the larger number of young participants, a small number of each of the quotatives had been expressed alongside second-person subjects.

Figure 4. 37

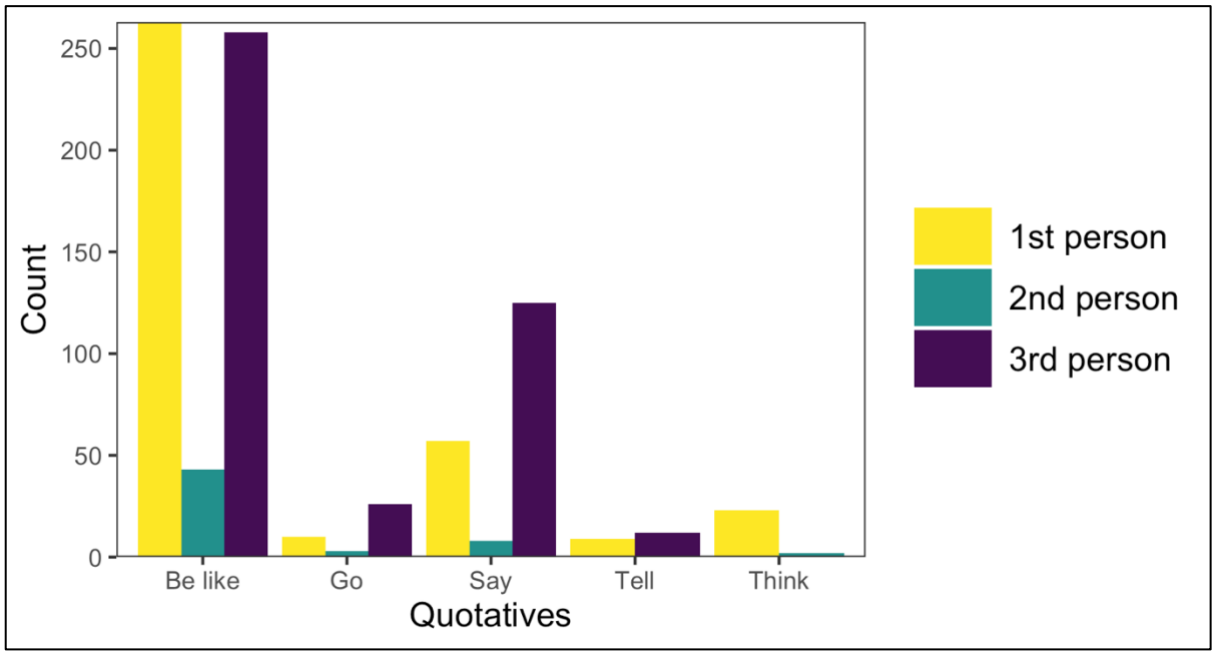
Bar charts showing grammatical person by percentage for each of the main quotatives used by younger participants of different races



Black, Coloured and White speakers all showed a preference for expressing *be like* alongside first-person subjects. In general, Coloured speakers tended to use first-person subjects more frequently, especially for *be like*, *tell* and *think*, whilst Black speakers tended to favour third-person subjects more frequently for *think*, *go* and *say*. Indian speakers also favoured third-person subjects for *be like* and *say*, although *tell* was expressed equally often alongside first and third-person subjects. White and Coloured speakers both showed a strong preference for expressing *think* alongside first-person subjects, with White speakers having no instances of *think* expressed alongside third-person subjects. Black speakers interestingly favoured third-person subjects over both first- and third-person subjects when using *think*. Black speakers also only used third-person subjects when introducing quotes using *go*. White speakers showed a preference for third-person subjects when using *go* and Coloured speakers used both first- and third-person subjects equally alongside *go*.

Figure 4. 38

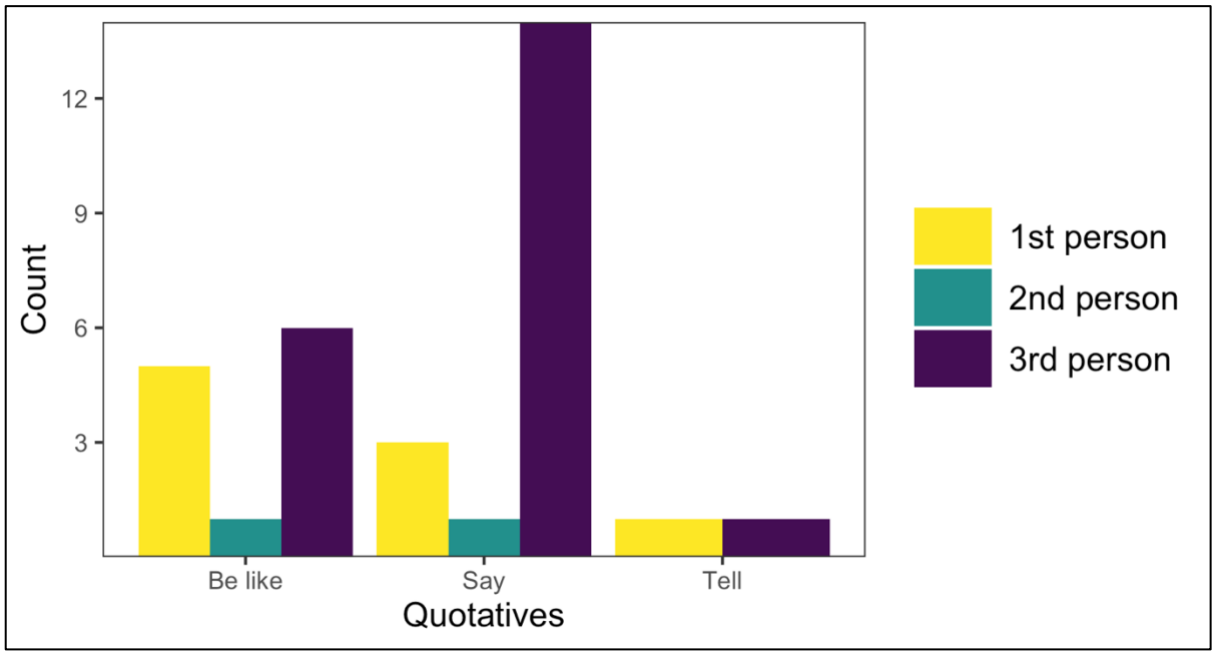
Bar chart of grammatical person for the main quotatives used by young White participants



White speakers used *be like*d slightly more often alongside first-person subjects than they did alongside third-person subjects. When using *think*, there was also a clear preference for first-person subjects and no instances of third-person subjects. First-person subjects were favoured when expressing *go*, *say* and *tell*. All quotatives other than *tell* showed low numbers of the use of second-person subjects.

Figure 4. 39

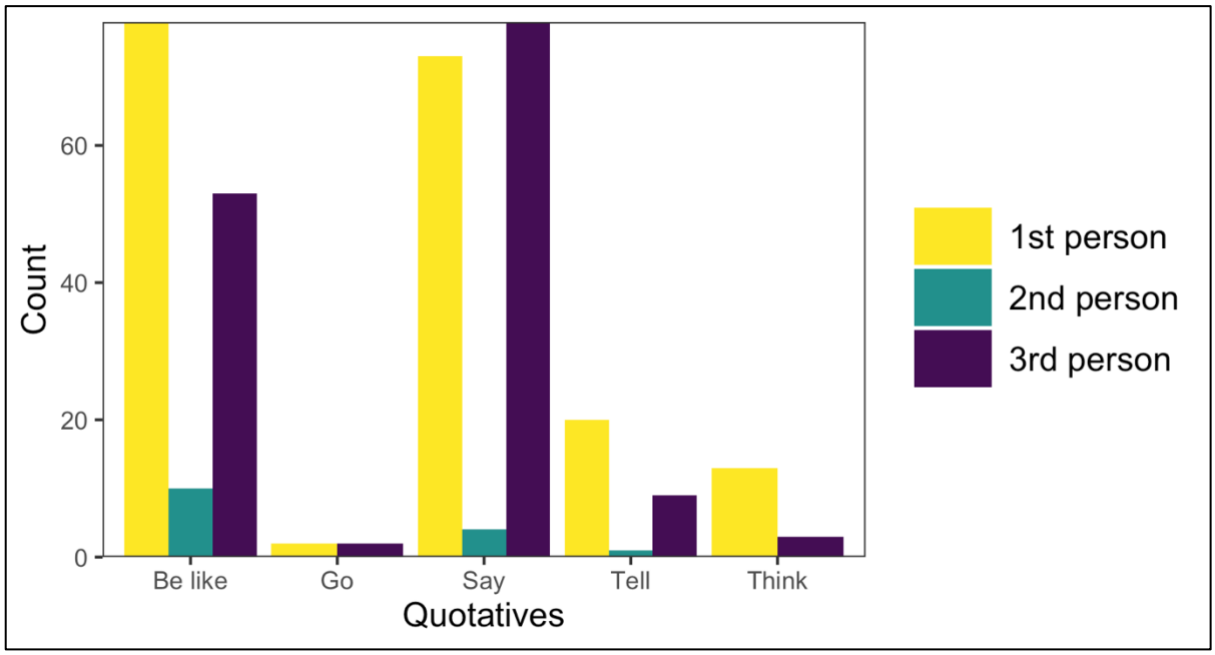
Bar chart of grammatical person for the main quotatives used by young Indian participants



Among Indian speakers, third-person subjects were used more often than first-person subjects and were favoured for the expression of both *say* and *be like*. *Tell* was expressed alongside equal numbers of first and third subjects, with no instances of second-person subjects being used. *Be like* and *say*, however, were used alongside small numbers of second-person subjects.

Figure 4. 40

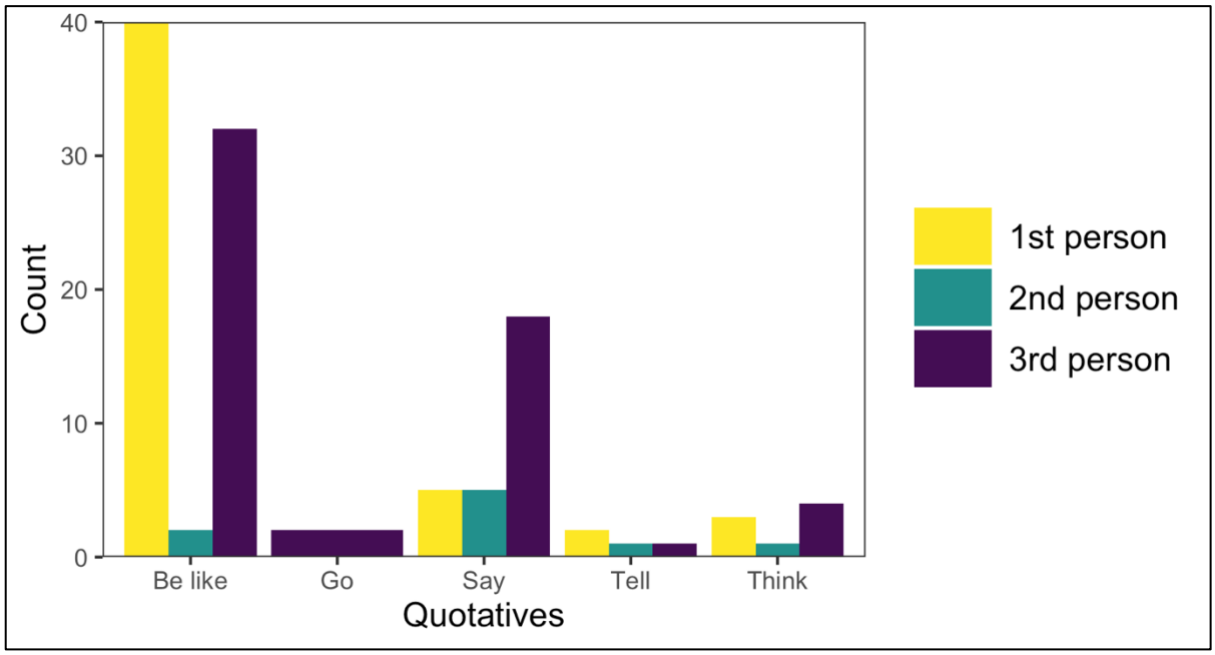
Bar chart of grammatical person for the main quotatives used by young Coloured participants



Among Coloured speakers, there was a strong preference for the use of first-person subjects alongside *be like*, *tell* and *think*. Although *say* was used most often with third-person subjects, this was closely followed by first-person subjects. *Go* occurred equally often alongside first- and third-person subjects. *Be like*, *say* and *tell* also occurred alongside small numbers of second-person subjects.

Figure 4. 41

Bar chart of grammatical person for the main quotatives used by young Black participants



Black speakers showed a higher use of first-person subjects in phrases containing both *be like* and *tell*. However, they showed a higher usage of third-person subjects for both *say* and *think*, and *go* only occurred in phrases with third-person subjects. Other than *go*, all quotatives appeared alongside small numbers of second-person subjects.

4.6 Conclusion

These descriptive statistics have been intended to provide a general impression of the distribution of quotatives in the database according to both social and linguistic factors. They suggest that the *be like* quotative has been enthusiastically adopted by the young speakers within Cape Town, as it is the quotative that occurs most often within the database. The more traditional quotatives, *say* and *think*, have not been rendered obsolete and continue to be used by younger speakers. Changes to the content of the quotes that quotatives introduce have also taken place, implying that the change to the quotative system is not only the introduction of a new quotative. *Go*, zero quotatives and other quotatives are used to quote internal monologue and non-lexicalised sounds in addition to speech. Although, the unequal numbers of participants who attended FMC schools versus FDET schools makes it more challenging to interpret, schooling does seem to have some effect on quotative choice, as *tell* is used more often by speakers with FDET schooling than by those with FMC schooling for both Coloured

and Indian participants. Quotative usage did show some variation according to gender. For example, young White men used *say*, *think* and *tell* more often than young White women did, despite female speakers tending to be responsible for the production of the majority of quotatives. However, men and women tended to share similar patterns of quotative usage, as Black and White men and women both produced more *be like* than *say* quotatives and Coloured and Indian men and women used *say* more frequently than *be like*. When it came to *tense and temporal reference*, Black and White speakers used a greater proportion of HP when expressing *be like*, and Coloured and Indian speakers used a higher proportion of present tense when expressing *say*, than occurred with the expression of other quotatives in the database. White, Coloured and Black speakers also all showed a higher usage of first-person subjects when expressing *be like* quotatives. As no statistically significant observations can be made from the bar charts alone, a limited version of the database, containing only the main quotatives, *be like*, *say*, *go*, *tell* and *think*, is used in the following sections in order to perform logistic regression analyses in order determine the effects that linguistic and social variables may have on quotative choice and the ways they may constrain specifically the use of the *be like* quotative.

5. Inferential Statistics Results

This chapter presents the results of the inferential statistics, which are presented in two parts. The inferential statistics investigate the linguistic variables that may constrain the use of *be like*, and whether the variable constraints on *be like* differ between speakers according to race, gender or schooling. Section 5.1 presents the results from various logistic regression models used to analyse the impact of explanatory variables and section 5.2 presents an alternative analysis using conditional inference trees and random forests.

5.1 Logistic Regression

The descriptive statistics alone make it clear that *be like* has been widely adopted by young Cape Town speakers. This section is an attempt to use logistic regression analysis to determine the variable constraints on the use of *be like* and other quotatives, and then to determine whether those constraints differ according to social variables of race, gender and schooling. The main quotatives database, made up of *be like*, *tell*, *say*, *think* and *go*, was fitted to three different models of logistic regression: a fixed-effects multinomial logistic regression model, a fixed-effects binomial logistic regression model and a mixed-effects binomial logistic regression model.

Levshina (2015:253) explains that logistic regression is used to model "the relationships between a categorical response variable with two or more possible values and one or more explanatory variables, or predictors." Because logistic regression is used specifically for categorical, rather than continuous, response variables, but can be used with both categorical and continuous explanatory variables, it is commonly used in research on synonyms in order to determine on what "conceptual, geographical, social, [or] pragmatic" basis speakers may be choosing between variants (Levshina, 2015:253). Binomial logistic regression is used to predict the choice between two synonyms, or one synonym over a group of other synonyms, whereas multinomial logistic regression assesses the choice between three or more synonyms, focusing on multiple outcomes at once.

The exact definition of fixed and random effects vary (Andrew, 2005), but for the purposes of this study, fixed effects refer to most explanatory variables, which have a measurable effect on the data and random effects refer to factors that are "sampled randomly from the population" (Levshina, 2015: 192) and have an effect that is less easily measured. The individual participants in our study are random effects, as the impact that they have on the data,

due to personal history and conversational style for example, is variable and not easily predicted or measured. A fixed-effects model takes into consideration only fixed effects, whereas a mixed-effects model accounts for both fixed and random effects.

5.1.1 Multinomial logistic regression without mixed effects

A multinomial logistic regression analysis provides the opportunity to assess the behaviour and spread of the entire quotative cohort simultaneously, rather focusing solely on *be like*. There are two types of multinomial logistic regression models. "One type involves fitting two or more binary models, which compare each outcome against some reference level. […] The other approach models the odds of each outcome against all other outcomes." (Levshina, 2015: 277) The multinomial model used below, fitted using the mlogit (Croissant, 2020) package and function, is the first type that Levshina (2015) refers to and, as such, compares the outcomes of multiple binary models with a reference level. Due to convergence errors encountered when random effects were included, it is a fixed-effects model.

Figure 5. 1

R Script showing the results for the later multinomial logistic regression run using mlogit using the younger generation only as data

```
Call:
mlogit(formula = quotative ~ 1 | sex + person + ethnicity + school +
    tense + ContentOfQuote, data = young1, reflevel = "be_like",
    method = "nr")

Frequencies of alternatives:choice
 be_like       go      say     tell    think
0.593398 0.033758 0.293323 0.042761 0.036759

nr method
20 iterations, 0h:0m:2s
g'(-H)^-1g = 7.35E-07
gradient close to zero

Coefficients :
                          Estimate  Std. Error z-value  Pr(>|z|)
(Intercept):go         -2.3595e+00  1.1225e+00 -2.1020  0.035549 *
(Intercept):say        -2.4085e+01  1.5061e+04 -0.0016  0.998724
(Intercept):tell       -2.5562e+01  1.8061e+04 -0.0014  0.998871
(Intercept):think      -2.0628e+01  5.8828e+03 -0.0035  0.997202
sexm:go                 4.7302e-01  3.3123e-01  1.4280  0.153278
sexm:say                5.9211e-01  1.4948e-01  3.9611 7.460e-05 ***
sexm:tell               6.7382e-01  3.2429e-01  2.0778  0.037725 *
sexm:think              7.8648e-01  3.6763e-01  2.1393  0.032411 *
person2nd:go           -1.7685e-01  6.9502e-01 -0.2544  0.799149
person2nd:say           3.2907e-01  3.7422e-01  0.8793  0.379214
person2nd:tell         -1.5532e-01  8.0979e-01 -0.1918  0.847893
person2nd:think        -6.4637e-01  6.9938e-01 -0.9242  0.355377
person3rd:go            5.3295e-01  3.7313e-01  1.4283  0.153205
person3rd:say           3.2843e-01  1.5523e-01  2.1158  0.034360 *
person3rd:tell         -3.6039e-01  3.1132e-01 -1.1576  0.247020
person3rd:think        -8.2489e-02  5.1033e-01 -0.1616  0.871590
ethnicityc:go           5.7201e-01  9.6523e-01  0.5926  0.553437
ethnicityc:say          1.4601e+00  2.9351e-01  4.9744 6.544e-07 ***
ethnicityc:tell         1.9850e+00  6.2593e-01  3.1713  0.001518 **
ethnicityc:think       -8.2053e-01  6.1139e-01 -1.3421  0.179573
ethnicityi:go          -1.7166e+01  1.1270e+04 -0.0015  0.998785
ethnicityi:say          1.3297e+00  5.1619e-01  2.5760  0.009996 **
ethnicityi:tell         1.0322e+00  9.9772e-01  1.0346  0.300854
ethnicityi:think       -1.8380e+01  9.1196e+03 -0.0020  0.998392
ethnicityw:go           1.2798e+00  9.3971e-01  1.3619  0.173234
ethnicityw:say          3.8168e-01  2.9269e-01  1.3040  0.192230
ethnicityw:tell         1.1424e+00  7.0820e-01  1.6131  0.106726
ethnicityw:think       -7.4688e-01  5.9990e-01 -1.2450  0.213125
```

```
schoolnot:go             3.0867e-01  9.6740e-01  0.3191  0.749669
schoolnot:say            7.6585e-01  2.6253e-01  2.9172  0.003531 **
schoolnot:tell           2.2437e+00  4.4746e-01  5.0144 5.321e-07 ***
schoolnot:think          1.8370e+00  6.1256e-01  2.9989  0.002709 **
tenseHHT:go             -7.7692e-01  7.6836e-01 -1.0111  0.311950
tenseHHT:say            -4.0137e-02  3.8521e-01 -0.1042  0.917015
tenseHHT:tell            3.6387e-01  7.8829e-01  0.4616  0.644374
tenseHHT:think           1.5951e+01  5.8828e+03  0.0027  0.997837
tenseHP:go              -1.0461e+00  5.1449e-01 -2.0332  0.042028 *
tenseHP:say             -1.2312e+00  3.2390e-01 -3.8012  0.000144 ***
tenseHP:tell            -8.9851e-01  7.3666e-01 -1.2197  0.222572
tenseHP:think            1.6028e+01  5.8828e+03  0.0027  0.997826
tensepast:go            -3.5806e+00  8.5399e-01 -4.1928 2.756e-05 ***
tensepast:say           -4.1038e-01  3.0787e-01 -1.3330  0.182538
tensepast:tell          -5.1339e-02  6.7832e-01 -0.0757  0.939668
tensepast:think          1.6856e+01  5.8828e+03  0.0029  0.997714
tensepresent:go         -2.0057e-01  5.8672e-01 -0.3418  0.732471
tensepresent:say         2.2659e-01  3.6609e-01  0.6189  0.535959
tensepresent:tell        1.5147e-01  7.9231e-01  0.1912  0.848386
tensepresent:think       1.7132e+01  5.8828e+03  0.0029  0.997676
ContentOfQuoteim:go     -2.2902e+00  8.3375e-01 -2.7469  0.006016 **
ContentOfQuoteim:say     1.8055e+01  1.5061e+04  0.0012  0.999044
ContentOfQuoteim:tell    1.9253e+01  1.8061e+04  0.0011  0.999149
ContentOfQuoteim:think   2.9011e+00  1.0646e+00  2.7250  0.006431 **
ContentOfQuotes:go      -6.5281e-01  4.8767e-01 -1.3386  0.180690
ContentOfQuotes:say      2.3092e+01  1.5061e+04  0.0015  0.998777
ContentOfQuotes:tell     2.1672e+01  1.8061e+04  0.0012  0.999043
ContentOfQuotes:think   -1.7429e+01  2.1604e+03 -0.0081  0.993563
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood: -1026.4
McFadden R^2:  0.25967
Likelihood ratio test : chisq = 720 (p.value = < 2.22e-16)
```

The goodness of fit statistics indicate that this model was fairly well fitted. The low *p*-value (p.value = < 2.22e-16) of the likelihood ratio test showed that the model as a whole was significant and McFadden's R^2 indicated a good fit (between 0.2 and 0.4), with a value of 0.25967 (Levshina, 2015). The log likelihood, however, was quite high at -1026.4, showing that the fit of the model could definitely be improved, as the absolute value of the log likelihood, the better the fit (Levshina, 2015). The mean value of the fitted probabilities, retrieved using the 'fitted()' function in R (not pictured in figure 5.1), was 0.5503572, which also indicated that the predictive power of the model could be improved, as the closer the fitted probability is to 1, the more accurate it is (Levshina, 2015).

The log odds ratios are provided in the second ('estimate') column of the table of coefficients, and the p-values are provided in the final column. Negative log odds ratios indicate a preference for *be like* over the quotative provided, whereas positive log odds ratios

indicate a preference for the quotative listed (*tell*, *think*, *go* or *say*). The significance codes make it easy to identify which of these effects are statistically significant. The intercept rows represent the data when all variables are at their reference levels, which are, namely, Black for race, female for gender, FMC schooling for schooling, future tense for *tense and temporal reference*, exclamations and non-verbalised sounds for *content of quote*, and first person for *grammatical person*.

Reviewing the intercept rows shows that, when all variables are at their reference levels, *be like* is more likely to be used than *go*. Other statistically significant findings showed that men were more likely to use *say*, *tell* or *think* than *be like*; Coloured speakers were more likely to use *say* or *tell* than *be like;* Indian speakers were more likely to use *say* than *be like*; and speakers that did not attend a model-C school were more likely to use *tell* and *think* than *be like*. Speakers expressing a quote in HP were more likely to use *be like* than *say* or *go*; speakers expressing a quote in past tense were more likely to use *be like* than *go*; speakers using a third-person subject in a quotative phrase were more likely to use *be like* than *say*; and speakers expressing a quote containing inner monologue were more likely to use *be like* than *go*, but more likely to use *think* than *be like*. Of these findings, the ones that showed the greatest statistical significance, were the male speakers' and Coloured speakers' preference of *say* over *be like*, the preference speakers with FDET schooling showed for *tell* over *be like*, the general preference for the use of *be like* over *say* when a quotative phrase was in HP and the preference for the use of *be like* over *go* when a phrase was in past tense.

A model was also fitted for the main quotatives of both age groups which, although excluded, is worth mentioning for the strong effects it showed for the age variable. The older generation was the reference level and the results showed that *be like* was strongly preferred over all other quotatives, *go*, *think*, *tell* and *say*, for the younger generation. This model, with a similarly significant *p*-value (p.value = < 2.22e-16), but higher McFadden's R^2 at 0.35211 and log likelihood of -1077, did not have as good a fit as the model resulting from only the data of only the young speakers.

5.1.2 Binomial logistic regression without mixed effects

In order to create a binomial model, the main quotatives used by young speakers were divided into two groups, one containing only *be like* and another containing *say*, *think*, *go* and *tell*, which I will refer to as 'other quotatives' in this section (4.2.1), not to be confused with

the other quotatives referred to throughout section 4.1. Initially, all explanatory variables were included in the model. Two models were generated using both the 'lrm()' and 'glm()' functions from the rms (Harrell Jr, 2020) and base R (R Core Team, 2019) packages, in order to find the tests and results recommended by Levshina (2015).

The results of this initial run, however, showed that of all the linguistic and social variables, *grammatical person* was the only explanatory variable that did not display any significance. Stepwise selection was then employed to determine whether or not to exclude it from the model. Stepwise selection is an alternate method for selecting which variables to include in a model (Levshina, 2015:266-267). I used the 'step()' function on R to run bidirectional, forward and backward stepwise selection, all of which resulted in the inclusion of all variables other than *grammatical person*. Finally, I performed a chi-squared test ($p = 0.2442$) using the 'anova()' function in R to determine whether the change would significantly affect the deviance of the model. There was no significant reduction in deviance (1540.8 vs 1543.6) when *grammatical person* was part of the model, indicating that it was not a useful variable in this case and could be excluded (Levshina, 2015:267).

Figure 5. 2

R Script showing the results for the later binomial logistic regression run using lrm, with grammatical person excluded

```
Logistic Regression Model

lrm(formula = binomial ~ sex + ethnicity + school + tense + ContentOfQuote,
    data = young)

                      Model Likelihood     Discrimination    Rank Discrim.
                         Ratio Test           Indexes           Indexes
Obs           1333    LR chi2     257.56    R2       0.237    C       0.743
 be_like       791    d.f.            11    g        1.165    Dxy     0.486
 other         542    Pr(> chi2) <0.0001    gr       3.205    gamma   0.507
max |deriv| 2e-11                          gp       0.237    tau-a   0.234
                                           Brier    0.197

                  Coef    S.E.   Wald Z Pr(>|Z|)
Intercept         -2.4015 0.5237 -4.59  <0.0001
sex=m              0.5861 0.1287  4.55  <0.0001
ethnicity=c        1.1081 0.2530  4.38  <0.0001
ethnicity=i        0.8946 0.4690  1.91  0.0564
ethnicity=w        0.3820 0.2523  1.51  0.1300
school=not         1.0988 0.2278  4.82  <0.0001
tense=HHT         -0.1945 0.3500 -0.56  0.5784
tense=HP          -1.0850 0.2926 -3.71  0.0002
tense=past        -0.5207 0.2820 -1.85  0.0648
tense=present      0.2151 0.3231  0.67  0.5055
ContentOfQuote=im  0.6920 0.4215  1.64  0.1006
ContentOfQuote=s   1.9299 0.3956  4.88  <0.0001
```

Both the 'lrm()' and 'glm()' models were equally fitted and produced the same results. However, the call for the 'lrm()' model (figure 5.2) contains useful information for interpreting the fit of the model. The significant p-value of the Model Likelihood Ratio Test which informs us that the model itself, as well as one or more of its explanatory variables deviates significantly from zero (Levshina, 2015). The model also displays acceptable, if not ideal, discrimination according to Hosmer and Lemeshow (2000:162 in Levshina, 2015:259), who define acceptable discrimination as a *C* value between 0.7 and 0.8. As can be seen in the Rank Discrimination Indexes column, at *C* = 0.743, the model correctly predicts the correct quotative (*be like* or other) 74% of the time. The Nagelkerke pseudo-R2 also indicates the predictive ability of the model, ranging between 0 and 1. The closer to 1 the more accurate the prediction and at 0 there is no predictive power. At R2 = 0.237 this seems to indicate that the model has little predictive power, but this statistic is better suited to linear regression models and frequently reports lower

values for logistic regression models (Hosmer and Lemeshow, 2000:167 in Levshina, 2015:259).

The model was also validated with bootstrapping, using the 'validate()' function in R, to ensure that it was not overfitted. Low optimism values in the bootstrap validation implied that the model was not overfitted, and would therefore perform adequately if given different data. It is worth noting, however, that natural stylistic variation between participants may still result in a degree of overfitting that could be addressed by using a model that includes random effects, such as the one used in section 4.2.3 (Levshina, 2015).

Figure 5. 3

R Script showing the results for the later binomial logistic regression run using glm, with person excluded

```
Call:
glm(formula = binomial ~ sex + ethnicity + school + tense + ContentOfQuote,
    family = binomial, data = young)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0413  -0.9315  -0.5409   1.0000   2.5093

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      -2.4015     0.5236  -4.586 4.51e-06 ***
sexm              0.5861     0.1287   4.552 5.30e-06 ***
ethnicityc        1.1081     0.2530   4.380 1.19e-05 ***
ethnicityi        0.8946     0.4690   1.908 0.056445 .
ethnicityw        0.3820     0.2523   1.514 0.129950
schoolnot         1.0988     0.2278   4.824 1.41e-06 ***
tenseHHT         -0.1945     0.3500  -0.556 0.578443
tenseHP          -1.0850     0.2926  -3.709 0.000208 ***
tensepast        -0.5207     0.2820  -1.846 0.064827 .
tensepresent      0.2151     0.3231   0.666 0.505534
ContentOfQuoteim  0.6920     0.4215   1.642 0.100606
ContentOfQuotes   1.9299     0.3956   4.879 1.07e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1801.1  on 1332  degrees of freedom
Residual deviance: 1543.6  on 1321  degrees of freedom
AIC: 1567.6

Number of Fisher Scoring iterations: 4
```

The ‘glm()’ model (figure 5.3) returned the same correlation and significant variables as the ‘lrm()’ model (figure 5.2), but may be easier to interpret due to the use of significance codes. The estimate column of the table of coefficients supplies log odds ratios, which reflect the odds of the outcome for each level of the explanatory variables, with the reference level of that variable. The reference levels are represented by the intercept row. The levels were automatically selected according to alphabetical order and were assigned as follows: Black for the race variable, female for the gender variable, future for the *tense and temporal reference* variable, and exclamations and non-lexicalised sounds for the *content of quote* variable.

Negative log odds ratios indicate that a variable level is correlated with the use of *be like*, whilst positive log odds ratios indicate that a variable level is correlated with the use of other quotatives. As such, the results show that HHT, past tense and HP were positively correlated with the use of *be like*. However, of these, only the intercept and HP were statistically significant. The other statistically significant variables (those negatively correlated with the use of *be like*) were spoken *content of quote*, male speaker gender, Coloured speaker race and non-ex-Model-C schooling. This indicates that, when direct speech is being quoted, *be like* is less likely to be used. *Be like* is also less likely to be used by male speakers, Coloured speakers and students who attended FDET schools. However, *be like* is more likely to be used for expressing quotes in the HP and more likely to be used by female Black speakers who attended FMC schools and are expressing quotes in the future tense containing exclamations or non-lexicalised sounds.

Figure 5. 4

R Script showing the results for glm binomial logistic regression with log odds calculated using sum contrasts and with person included

```
Call:
glm(formula = binomial ~ sex + person + ethnicity + school +
    tense + ContentOfQuote, family = binomial, data = young)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0910  -0.9657  -0.5353   1.0178   2.4591

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)     -0.43670    0.20318  -2.149 0.031609 *
sex1            -0.29099    0.06446  -4.514 6.36e-06 ***
person1         -0.04454    0.11787  -0.378 0.705528
person2         -0.11950    0.18930  -0.631 0.527861
ethnicity1      -0.59153    0.19985  -2.960 0.003078 **
ethnicity2       0.52872    0.15123   3.496 0.000472 ***
ethnicity3       0.27503    0.32526   0.846 0.397794
school1         -0.55390    0.11406  -4.856 1.20e-06 ***
tense1           0.30021    0.22350   1.343 0.179189
tense2           0.12792    0.19724   0.649 0.516618
tense3          -0.77592    0.12740  -6.090 1.13e-09 ***
tense4          -0.19265    0.11303  -1.704 0.088291 .
ContentOfQuote1 -0.86808    0.26600  -3.263 0.001101 **
ContentOfQuote2 -0.13942    0.17318  -0.805 0.420776
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1801.1  on 1332  degrees of freedom
Residual deviance: 1540.8  on 1319  degrees of freedom
AIC: 1568.8

Number of Fisher Scoring iterations: 4
```

Whilst the above results (figures 5.2 and 5.3) are useful, in order to best compare the findings of this study to those in the literature, another version of the ‘glm’ model (figure 5.4) was generated. This model differs from the previous ‘glm’ model (figure 5.3) in that sum contrasts, also known as contrastive or deviation coding, were used to calculate the log odds ratios, allowing for the calculation of centred factor weights (or probabilities) similar to the uncentred factor weights used in programmes such as VARBRUL and GoldVarb (D.E. Johnson, personal communication, June 14, 2021). For the purposes of comparability, *grammatical person* has been included in this model.

Table 5. 1

Factor weights calculated using sum contrasting for the fixed-effects binomial logistic regression model

Explanatory variables	Log odds (sum contrast)	Factor weights	%	N
Speaker gender				
female	0.29099	0.5722385	64.45	553
male	-0.29099	0.4277615	50.11	238
range		0.2946235		
Speaker race				
Black	-0.21222	0.4471432	63.79	74
Coloured	-0.59153	0.3562839	40.75	141
Indian	0.52872	0.6291845	37.5	12
White	0.27503	0.5683273	67.22	564
range		0.2729006		
Speaker schooling				
FMC	0.55390	0.6350399	63.18	743
FDET	-0.55390	0.3649601	30.57	48
range		0.2700798		
Content of quote				
exclamations/sounds	1,0075	0.7325306	87.69	57
inner monologue	-0.86808	0.295654	78.8	197
speech	-0.13942	0.4652013	52.75	537
range		0.3675705		
Grammatical person				
first person	0.16404	0.5409183	63.59	386
second person	-0.04454	0.4888668	68.29	56
third person	-0.11950	0.4701605	54.19	349

	range	0.0707578		
Tense and temporal reference				
future	0.53984	0.6317752	41.54	27
HHT	0.30021	0.5744939	45.65	42
HP	0.12792	0.5319365	74.26	300
past	-0.77592	0.3151999	57.12	349
present	-0.19205	0.4521345	45.34	73
	range	0.3165753		

As can be seen in a comparison of figures 5.3 and 5.4, the changes to the coding affect the results of the model. However, table 5.1 demonstrates the way that this allows for a closer examination of the reference levels of the explanatory variables, as log odds and therefore factor weights can be calculated for all variables. The factor weights in table 5.1 show that future tense, HHT, HP, first-person grammatical subjects, quotes containing exclamations and non-lexicalised sounds, speakers from FMC schools, Indian speakers, White speakers and female speakers are correlated with the use of *be like* quotatives. Past tense, present tense, second and third grammatical person subjects, quotes containing internal monologue and speech, Black speakers, Coloured speakers, speakers who attended FDET schools and male speakers were correlated with the use of other quotatives. The more a factor weight differs from 0.5, the greater its effect is, indicating that male speakers, Coloured speakers, FDET schooling, use of past tense and expression of internal monologue were all correlated with the use of other quotatives, whilst Indian speakers, female speakers, use of future tense, expression of exclamations and non-lexicalised sounds, and FMC schooling were correlated with the use of *be like*.

Figure 5. 5

Variable effect plots for the fixed-effects binomial logistic regression model

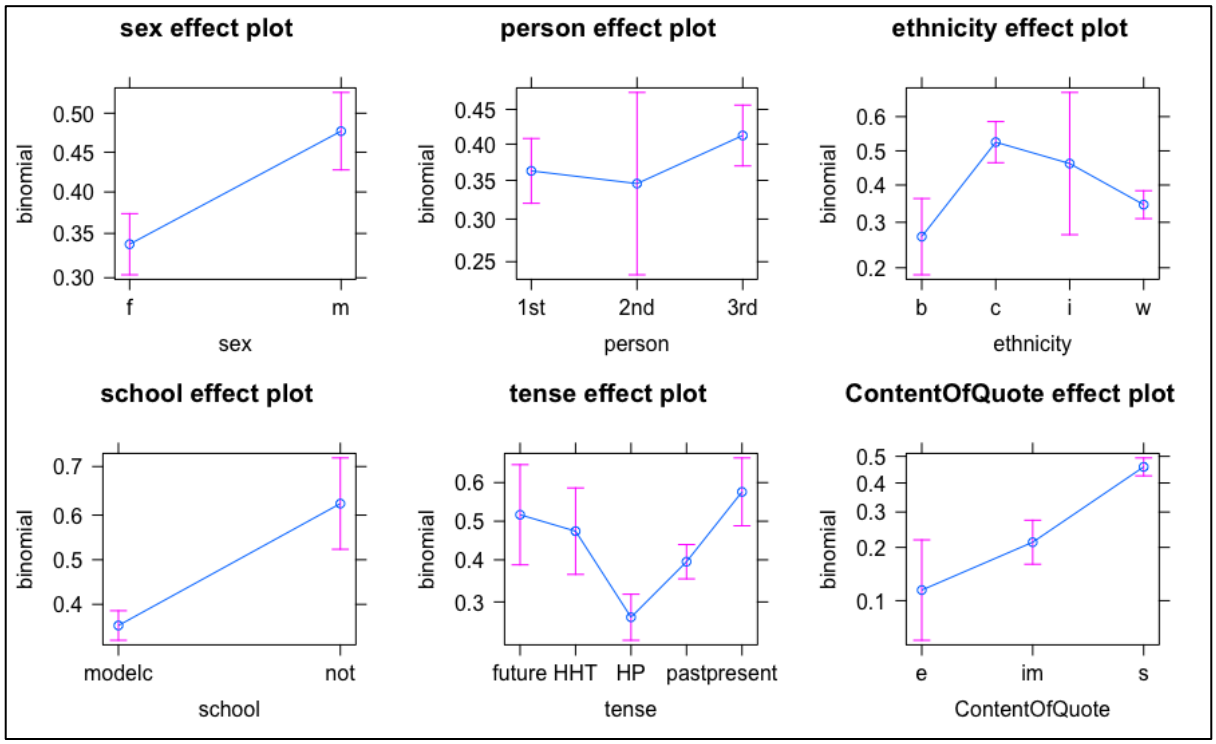


Table 5. 2

Variable effects for the fixed-effects binomial logistic regression model

Social variables	Fitted probability	Linguistic variables	Fitted probability
Gender effect		**Tense and temporal reference effect**	
male	0.4780685	future	0.5224646
female	0.3376289	HHT	0.4738802
Race effect		HP	0.2699075
Black	0.2660880	past	0.3939400
Coloured	0.5233657	present	0.5756807
Indian	0.4700441	**Grammatical person effect**	
White	0.3469303	first person	0.3627931
Schooling effect		second person	0.3456488
FMC	0.3556155	third person	0.4122478
FDET	0.6234817	**Content of quote effect**	

	exclamations/sounds	0.1121829
	inner monologue	0.2015520
	speech	0.4653632

In addition to the calculation of factor weights a variable effect analysis (figure 5.5 and table 5.2), generated using the 'allEffects()' function from the effects package (Fox, 2003; Fox & Weisberg, 2018; Fox & Weisberg, 2019), provides the fitted probabilities for the different levels of all variables. This allows for easier interpretation of their impact on quotative choice. The probability shown for each variable is the probability of another quotative being chosen rather than *be like*. Therefore, the lower the probability shown, the greater the chances of *be like* being chosen is. This is best seen in figure 5.5, where the lower a point on each diagram is (and the closer its value is to zero), the more likely it is that *be like* will be used in those conditions. For example, in the first row, we see that women are more likely to use *be like* than, first and second-person subjects are more likely to be used alongside *be like* than a third-person subject, and Black and White speakers are more likely than Coloured and Indian speakers to use *be like*. The second row shows that speakers from FMC schools are more likely to use *be like*, whilst speakers with FDET schooling are more likely to use other quotatives; phrases in the future or present tense are more likely to feature other quotatives, whereas HHT, past and especially HP phrases are more likely to feature *be like*; and quotes containing speech are less likely to be introduced by *be like* than quotes containing internal monologue, and exclamations or non-lexicalised sounds are most likely to be introduced by *be like*.

5.1.3 Binomial logistic regression with mixed effects

A mixed-effects binomial logistic regression was then performed in order to take into consideration the impact individual speakers may have on the data. The model was generated using the 'glmer()' function in the lme4 (Bates et al., 2015) and lmerTest (Kuznetsova, Brockhoff & Christensen, 2017) package on R and, as with previous models, used the database of main quotatives expressed by younger speakers. All explanatory variables were included. This decision was informed by the Lower Akaike Information Criterion (AIC in figure 5.6 below) and Bayesian Information Criterion (BIC in figure 5.6 below) values in the model, which reflect goodness of fit and are thus used to inform the choice of model (Levshina, 2015). They assess the fit of a model by reflecting the model's ability to explain as much of the variation with the use of as few variables as possible (Bevans, 2020). *Grammatical person* was

included in this model as neither the AIC nor BIC, both of which penalise the additional inclusion of variables, decreased significantly when it was excluded.

Figure 5. 6

R Script showing the results for the mixed-effects binomial logistic regression model generated with glmer

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: binomial ~ sex + ethnicity + school + tense + person + ContentOfQuote +
    (1 | speaker)
   Data: young2
Control: glmerControl(optimizer = "bobyqa")

     AIC      BIC   logLik deviance df.resid
  1356.0   1433.9   -663.0   1326.0     1318

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.2317 -0.5841 -0.1662  0.5606  4.6843

Random effects:
 Groups  Name        Variance Std.Dev.
 speaker (Intercept) 1.452    1.205
Number of obs: 1333, groups:  speaker, 52

Fixed effects:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)      -2.229096   0.796208  -2.800  0.00512 **
sexm              0.611581   0.405698   1.507  0.13169
ethnicityc        0.677969   0.605660   1.119  0.26297
ethnicityi        0.306907   0.772726   0.397  0.69124
ethnicityw        0.078109   0.641059   0.122  0.90302
schoolnot         1.420629   0.556264   2.554  0.01065 *
tenseHHT         -0.366282   0.397524  -0.921  0.35684
tenseHP          -1.365267   0.332443  -4.107 4.01e-05 ***
tensepast        -0.753851   0.325252  -2.318  0.02046 *
tensepresent      0.124303   0.371564   0.335  0.73797
person2nd        -0.004372   0.335437  -0.013  0.98960
person3rd         0.203941   0.156457   1.303  0.19240
ContentOfQuoteim  0.753965   0.474037   1.591  0.11172
ContentOfQuotes   1.968906   0.448093   4.394 1.11e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

As in previous regression models, the intercept represents the reference levels of each variable, which remain Black for the race variable, female for the gender variable, first person for the *grammatical person* variable, future for the *tense and temporal reference* variable, and exclamations and non-lexicalised sounds for the *content of quote* variable. Negative log odds ratios indicate a preference for *be like* and positive log odds ratios indicate a preference for other quotatives.

Similar to the findings of the fixed-effects binomial regression, the intercept, FDET schooling, HP, past tense, and quotes containing speech were all found to be statistically significant. However, male speakers and Coloured speakers were not found to be statistically significant predictors in the mixed-effects model. Thus, the mixed-effects regression shows that speakers who attended FDET schools and speakers expressing quotes containing direct speech were more likely to use other quotatives than *be like* and speakers using the HP or past tense were more likely to use *be like* than other quotatives. Female Black speakers who attended FMC schools expressing quotes containing exclamations or non-lexicalised sounds in the future tense with first-person subjects, as represented by the intercept level, also showed a preference for the use of *be like* over other quotatives.

Figure 5. 7

Variable effect plots generated using the allEffects() function from the effects and jtools packages

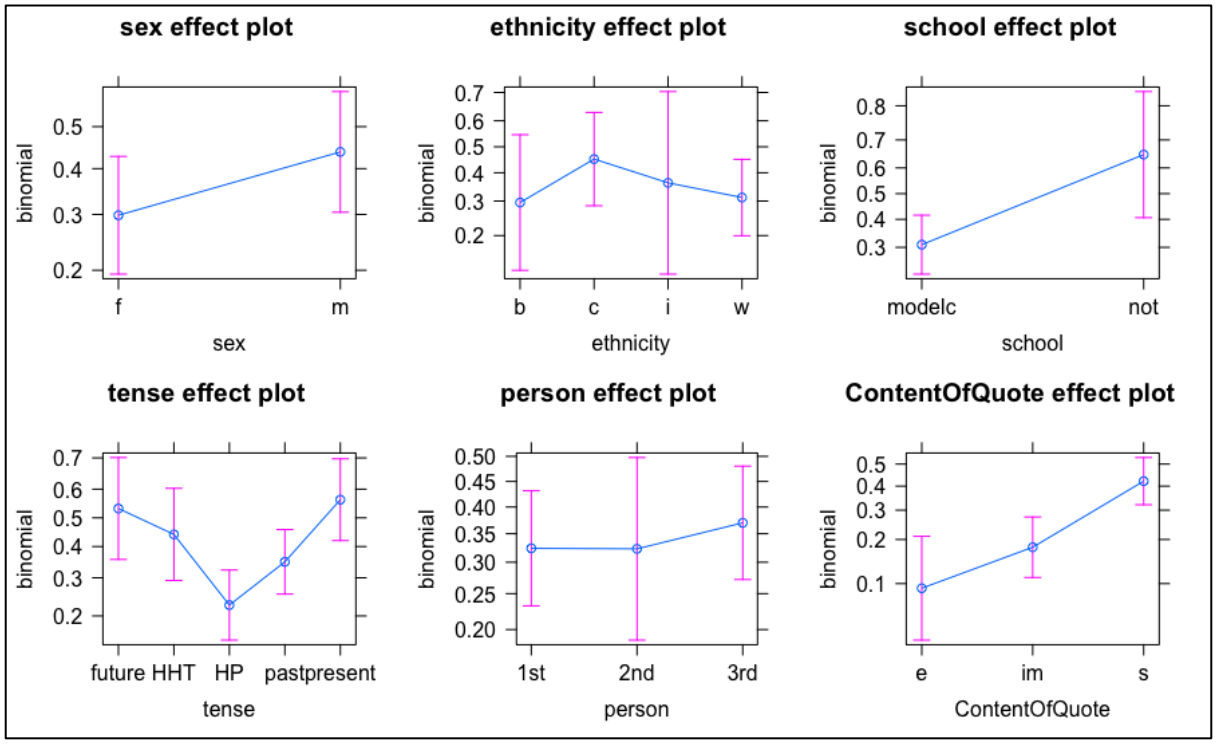


Table 5. 3

Variable effects for the mixed-effects binomial logistic regression model

Social variables	Fitted probability	Linguistic variables	Fitted probability
Gender effect		**Tense and temporal reference effect**	
male	0.4393075	future	0.5329424

female	0.2982691	HHT	0.4416862
Race effect		HP	0.2256032
Black	0.2952271	past	0.3493516
Coloured	0.4521068	present	0.5637177
Indian	0.3628017	**Grammatical person effect**	
White	0.3117347	first person	0.3239051
Schooling effect		second person	0.3229483
FMC	0.3089688	third person	0.3700641
FDET	0.6492365	**Content of quote effect**	
		exclamations/sounds	0.09256898
		inner monologue	0.17818415
		speech	0.42219793

The variable effects analysis of the mixed-effects binomial regression (figure 5.7 and table 5.3) reveals a similar pattern to the variable effects analysis of the fixed-effects binomial regression (figure 5.5 and table 5.2). This is best seen in figures 5.5 and 5.7 in which the graphs of each variable show similar patterns, with the points lowest on the graphs being more likely to occur with *be like* and the points higher on the graphs being more likely to occur with other quotatives. However the strengths of the effects (the further from 0.5, the stronger the effect) changed once participants were included as a random variable in the mixed-effects model. The effect of gender, *grammatical person*, schooling and *content of quote* all increased. Within the race variable, the effect of a speaker being Black decreased, but the effect of being Indian or White increased. The effect of a speaker being Coloured also increased but, more interestingly, the fitted probability changed from 0.5233657 to 0.4521068, showing that when participants are treated as random variables, a speaker being Coloured changes from being more likely to result in the expression of other quotatives to being more likely to result in the use of *be like*. The effects of future, HHT, HP and past tense phrases increased in the mixed model, whilst the effect of present tense phrases decreased.

5.1.4 Interactions within the models

Discussions of interactions between variables within the data seldom occur within the literature. Levshina (2015:268) defines interactions in the data as "when the effect of one predictor depends on the value of another variable." With the exception of Tagliamonte et al.'s (2016) inclusion of an interaction between cities and *tense and temporal reference*, *grammatical person* and *content of quote* in their mixed-effects binomial regression model (created in order to assess the effect of different cities on the use of *be like*), there are few examples of interactions within the literature.

It is likely that there are interactions with these data. For example, during apartheid, only White students were permitted to attend model-C schools, so we find that all White students in our data attended FMC schools. Therefore, schooling is to some degree dependent on race, and we could assume that there is an interaction between the schooling and race social variables. It is also likely that there are interactions within the linguistic variables. For example, prior to the introduction of *be like*, when *think* was the only quotative used to quote internal monologue, one would primarily quote their own inner monologue, rather than anyone else's, thus using first-person subjects for internal monologue. Therefore it is possible that there was an interaction between *grammatical person* and *content of quote*.

After generating exploratory line graphs in order to identify potential interactions between explanatory variables, experimental binomial logistic regression models were created to test the inclusion of all possible interactions in our data. With the use of anova's chi-squared test, the inclusion of each of the following interactions was found to result in a model that differed significantly from the model that included no interactions (figure 5.6): race and school; race and gender; race and *content of quote*; *grammatical person* and race; *grammatical person* and schooling; gender and *tense and temporal reference*; and gender and schooling. However, it was not possible to include all seven interactions in one model without encountering error codes.

When interactions were included in the mixed-effects model, a lower AIC resulted from the inclusion of the following interactions: *grammatical person* and *content of quote*; *tense and temporal reference* and *content of quote*; gender and schooling; gender and *content of quote*; schooling and *tense and temporal reference*; gender and *grammatical person*; gender and *tense and temporal reference*; race and *grammatical person*; and schooling and *content of quote*. However, the inclusion of all other interactions resulted in error codes. That a smaller AIC

results from the inclusion of all of the above interactions, suggests that the fit of the model is improved when interactions are taken into account but, as with the fixed-effects model, it was not possible to include all interactions in one mixed-effects logistic regression model. If the data are highly interactive, they may be better explained and analysed using conditional inference trees than logistic regression.

5.2 Random Forest and Conditional Inference Trees

As discussed in the previous chapter, both random forests and conditional inference trees, or ctrees, (Hothorn & Zeileis, 2015; Hothorn, Hornik & Zeileis, 2006) use permutation to select and classify explanatory variables that are highly correlated with response variables, as well as to determine the conditional importance of the different explanatory variables. They are particularly suited to data with a large variety of potential explanatory variables and contain complex interactions between variables (Levshina, 2015). A random forest (Hothorn et al., 2006; Strobl et al., 2007; Strobl et al., 2008) is be used to determine the conditional importance of the explanatory variables for predicting quotative choice among the younger speakers, before we turn to ctrees for an analysis of the main quotatives used by speakers of both generations.

Figure 5. 8

Dot chart of the conditional importance of the explanatory variables for the younger generation

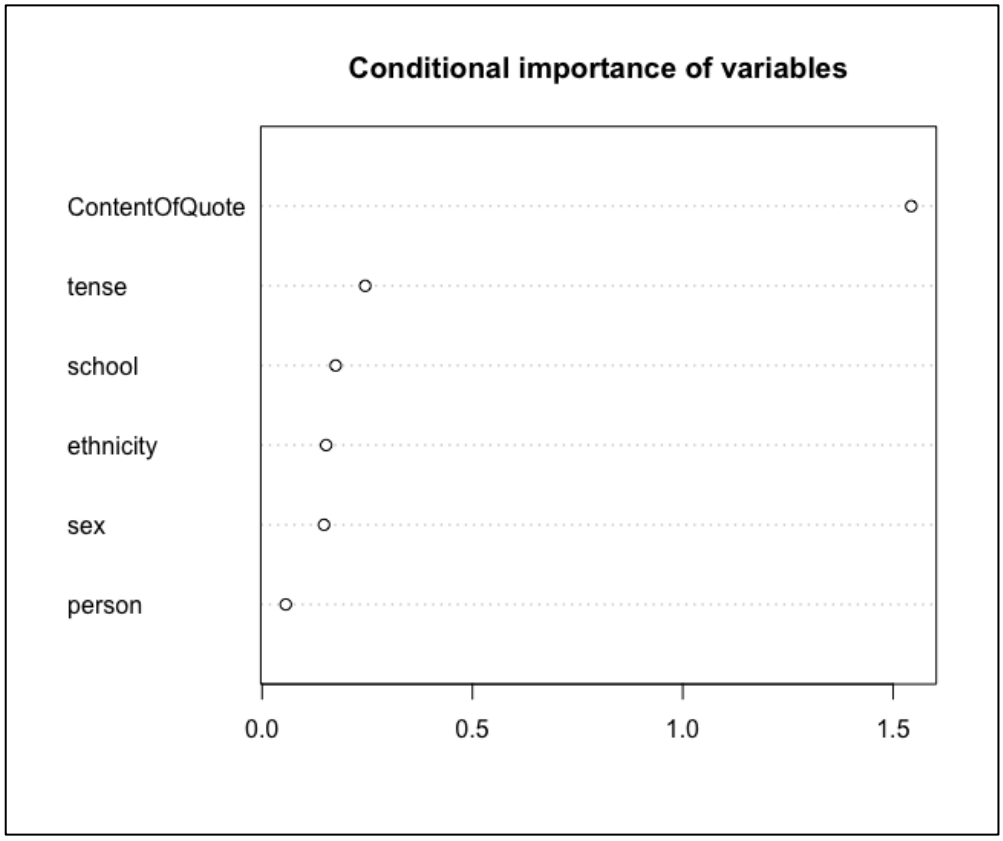


Table 5. 4

Conditional importance values for explanatory variables for the younger generation (accuracy: 69.69%)

Variable	Speaker gender	Grammatical person	Speaker race	Speaker school	Tense & temporal reference	Content of quote
Conditional importance	0.147	0.056	0.152	0.175	0.245	1.54

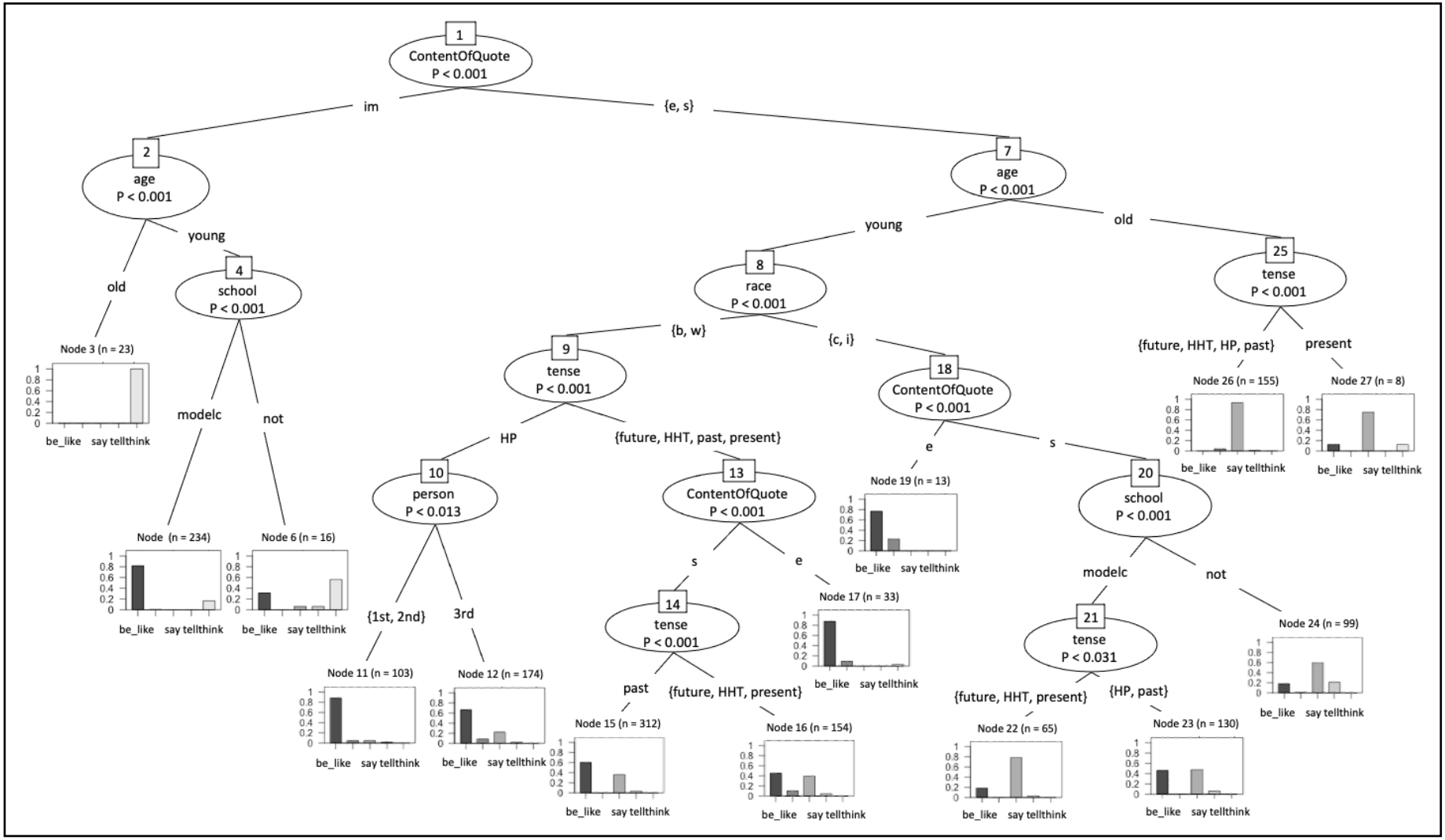
The random forest generated using the database of main quotatives produced by the younger generation (in figure 5.8 and table 5.4) presents the conditional importance of the five linguistic and social explanatory variables. The variable with the highest conditional importance, and thus the greatest impact on quotative choice, is that of *content of quote*. *Tense and temporal reference* is the second most important variable when determining which of the main quotatives will be expressed. It is followed by speaker schooling, speaker race and

speaker gender, with *grammatical person* displaying the lowest value of conditional importance.

As discussed previously, ctrees do not require pruning (Hothorn, Seibold & Zeileis, 2019). However, I found that the ctrees generated using our data were too large to fit one page, and I experimented with various models to determine whether it was possible to present a smaller tree without sacrificing explanatory power. Each of these models was generated multiple times with different seeds, to ensure they returned similar results each time. I calculated the accuracy of each model and compared the complexity, accuracy and size of all the models. All of the ctrees were generated using the database of main quotatives which included *be like*, *go*, *say*, *tell* and *think*, but excluded zero quotatives. However, for some of the ctrees the older generations' quotatives were included in this database, whilst others included only the main quotatives produced by the younger generation.

Figure 5. 9

Ctree generated using all main quotatives produced by all speakers



The first ctree I created (figure 5.9) was very large and contained both the younger and older generations and required $p < 0.05$ in order to create a split. It had 69.19% accuracy. I then generated a binomial version of this same tree (distinguishing between *be like* and the other main quotatives as I did for the binomial logistic regression models), which had a 75.71% accuracy rate. I then created ctrees using only the data from the younger generation. The ctree containing all main quotatives had an accuracy rate of 65.79%, whereas the binomial ctree had an accuracy rate of 72.47%. However, even when I included only the younger generation, and did not include all the main quotatives, the ctrees remained too large to read easily on a page.

I then attempted to decrease the size of the ctree by restricting not only the database to only young speakers and the result variable to only *be like* or *other*, but also by restricting the explanatory variables to include only to only the *content of quote*, speaker gender, race and school (excluding *grammatical person* and *tense and temporal reference*). This tree had an accuracy percentage of 69.69%. Whilst this ctree was smaller, when contrast with the previous binomial ctree that included all explanatory variables, the slightly smaller size did not seem to be worth the (72.47%-69.69%) 2.78% decrease in predictive power. Applying the same restricted explanatory variables to a ctree with a multinomial result variable, also resulted in an accuracy roughly 2% lower than the initial model (69.19%-66.32%=2.87%) which included all explanatory variables.

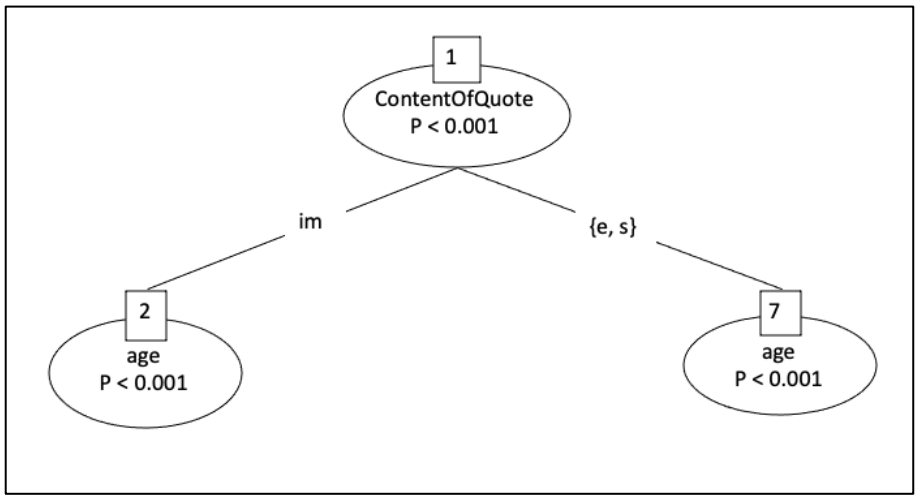
I then tried limiting the required p-value for a split to $p < 0.001$, but the size of the resulting tree was not much smaller than the initial tree and decreased in accuracy to 65.79%. I then programmed the ctree to require more observations per node, in addition to limiting the p-value, but this had no effect on either the size or accuracy of the ctree. It remained the same as when only the p-value was limited. Similarly, the ctree didn't reduce in size when programmed to require more observations per subset, in addition to limiting the p-value. However, requiring more observations per subset resulted in the accuracy decreasing to 65.49%. Finally, I found that limiting the branch depth, in addition to limiting the p-value, resulted the smallest ctree. The accuracy of a ctree with a maximum branch depth of 3 was 65.79%.

Whilst the use of a binomial result variable resulted in a ctree with the highest accuracy, I decided to use the initial ctree, as I felt it offered the most complexity with a high level of accuracy, at 69.19%. This decision was made so that the as much of quotative cohort could be examined at once, in order to gain understanding about how the quotative system as a whole is

changing due to the introduction of *be like*. For ease of reading, enlarged segments of the ctree in figure 5.9 are presented alongside an analysis of the ctree.

Figure 5. 10

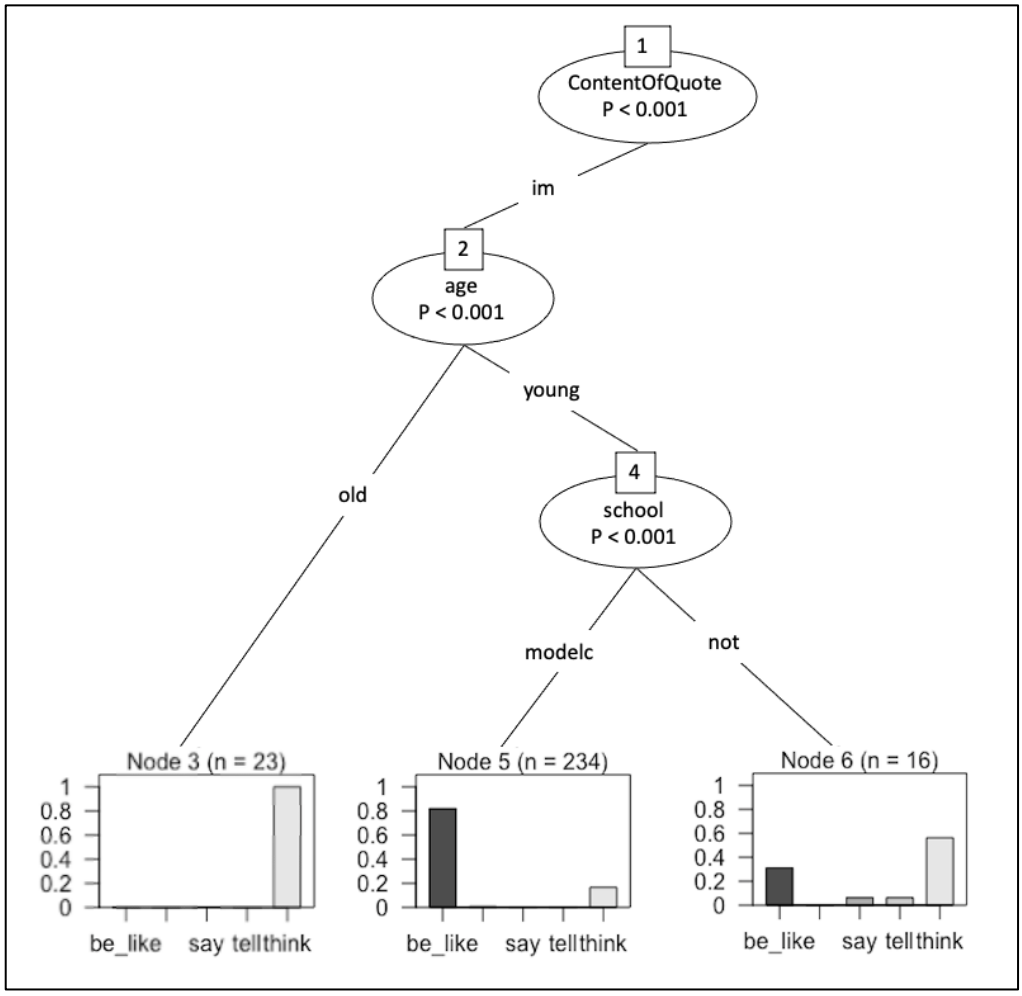
The initial split of the ctree (1)



The initial split ($p < 0.001$) divides the data according to content of quote (figure 5.10). It separates the data containing spoken quotes from data containing internal monologue, exclamations and non-lexicalised sounds. The next split is identical on both sides of the tree, as each branch is divided according to age, separating the younger generation from the older generation ($p < 0.001$ for nodes 2 and 7).

Figure 5. 11

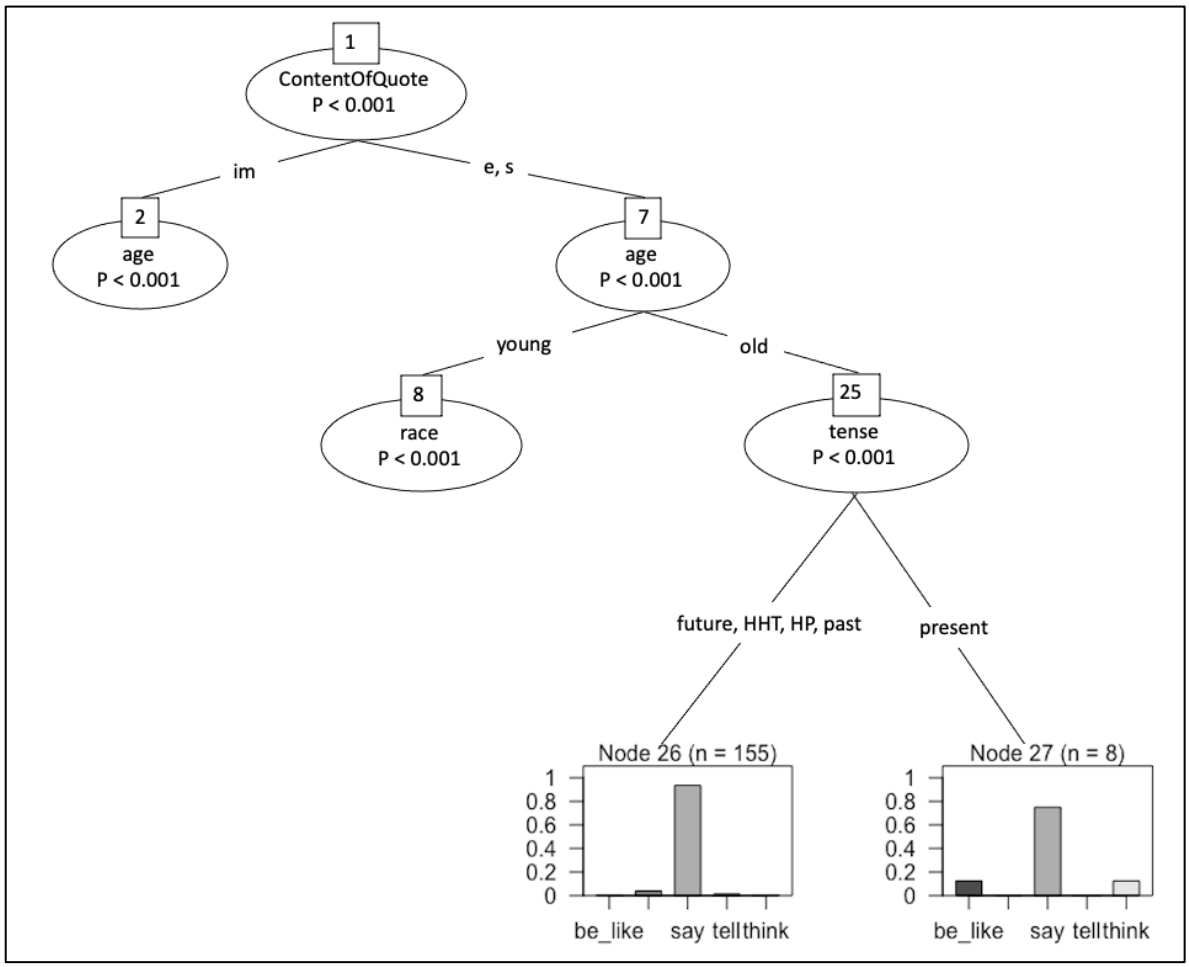
The left-hand side of the ctree, showing the distribution of quotatives for quotes containing either inner monologue (nodes 2-4)



The left-hand branch of the ctree represents all quotes containing internal monologue. At the second node, quotes expressed by older speakers are separated from those expressed by younger speakers ($p < 0.001$). Node three shows that the only quotes containing internal monologue spoken by older speakers were introduced by *think*. Quotes containing internal monologue expressed by younger speakers are then divided according to speakers' schooling at node four ($p < 0.001$). The fifth node represents the internal monologue quotes that were spoken by speakers who attended FMC schools, who primarily used *be like*, but also used *think* and two instances of *go* to quote internal monologue. The sixth node on the other hand shows the quotatives used by speakers who did attend FDET schools and used *think* more often than *be like* to quote internal monologue. Unlike the speakers who went to FMC schools, there were also instances of them using *say* and *tell* to quote internal monologue.

Figure 5. 12

The right-hand side of the ctree, showing the distribution of quotatives used by the older generation to express quotes containing speech or exclamations and non-lexicalised sounds, according to tense and temporal reference (nodes 25-27)



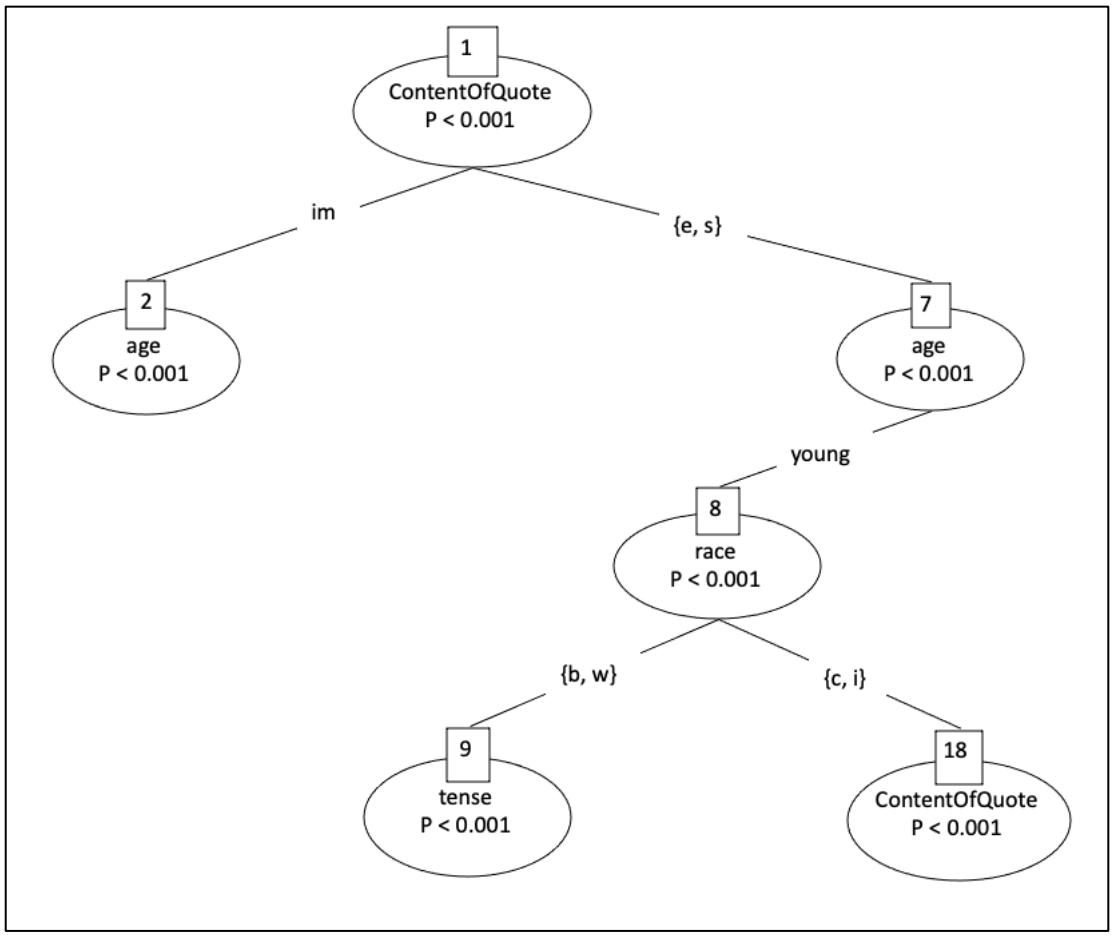
The right-hand half of the tree depicts quotes containing speech or exclamations and non-lexicalised sounds. The seventh node divides them according to age ($p < 0.001$), after which quotes containing speech or exclamations and non-lexicalised sounds expressed by participants from the older generation are further divided according to *tense and temporal reference* and those expressed by the younger generation are further divided according to the race of the speakers.

At the 25th node, quotes expressed by the older generation which contain speech or non-lexicalised sounds are divided according to *tense and temporal reference* ($p = 0.002$). The quotative phrases that were spoken in the past, future, HHT or HP are depicted in node 26 and primarily use *say*, although there are several uses of *go* and *tell*, and one each of *be like* and *think*. Node 27 represents those quotative phrases expressed in the present tense and also shows

a preference for the use of *say*, with one use each of *be like* and *think*, but no instances of *go* or *tell*. Because 155 quotatives are represented in node 26 and only 8 in node 27, there appear to be more instances of *be like* and *think*. However, this is only due to the fact that a smaller sample results in single quotatives making up a greater percentage than they would if they were part of a larger sample.

Figure 5. 13

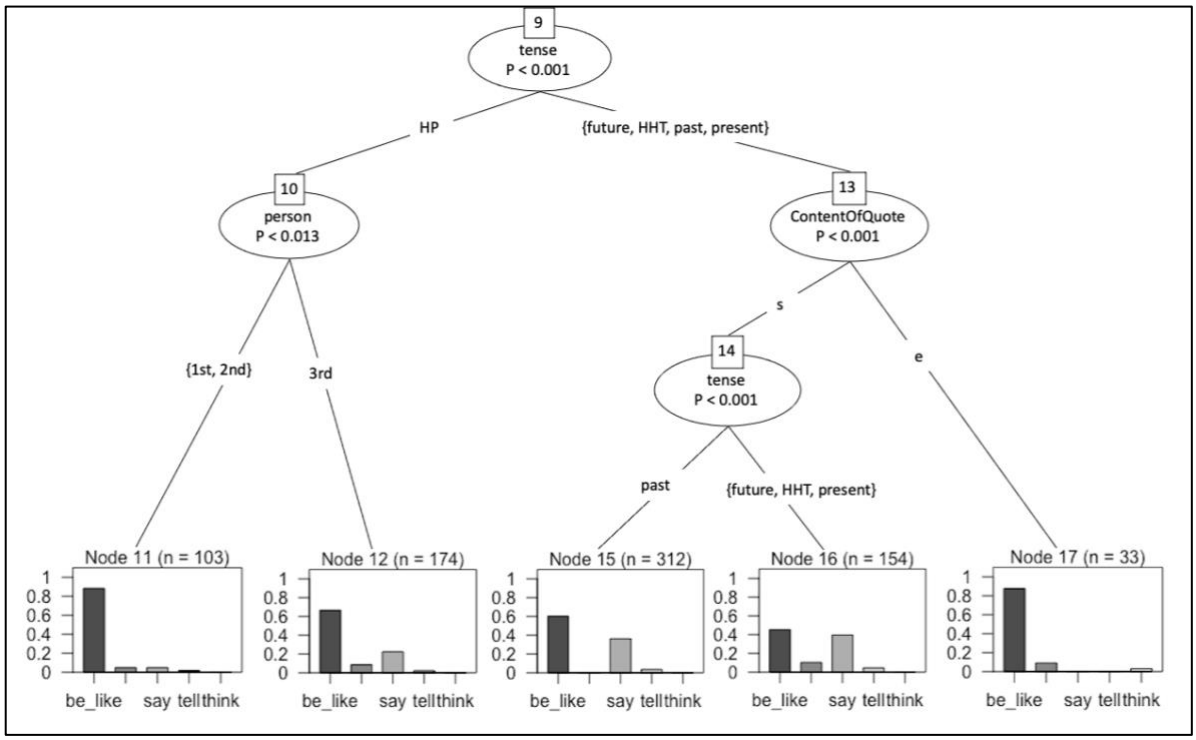
The left-hand segment of the right-hand side of the ctree, showing the distribution of quotatives used by the younger generation, to express quotes containing speech or exclamations and non-lexicalised sounds (nodes 8-24)



The quotes containing speech or non-lexicalised sounds and exclamations expressed by younger speakers are divided at node eight according to the speakers' race ($p < 0.001$). Quotes spoken by Black and White speakers, on the left, are separated from quotes spoken by Coloured and Indian speakers, on the right.

Figure 5. 14

The left half of the left-hand segment of the right-hand side of the ctree, showing the distribution of quotatives used by Black and White speakers of the younger generation to express quotes containing speech or exclamations and non-lexicalised sounds (nodes 9-17)



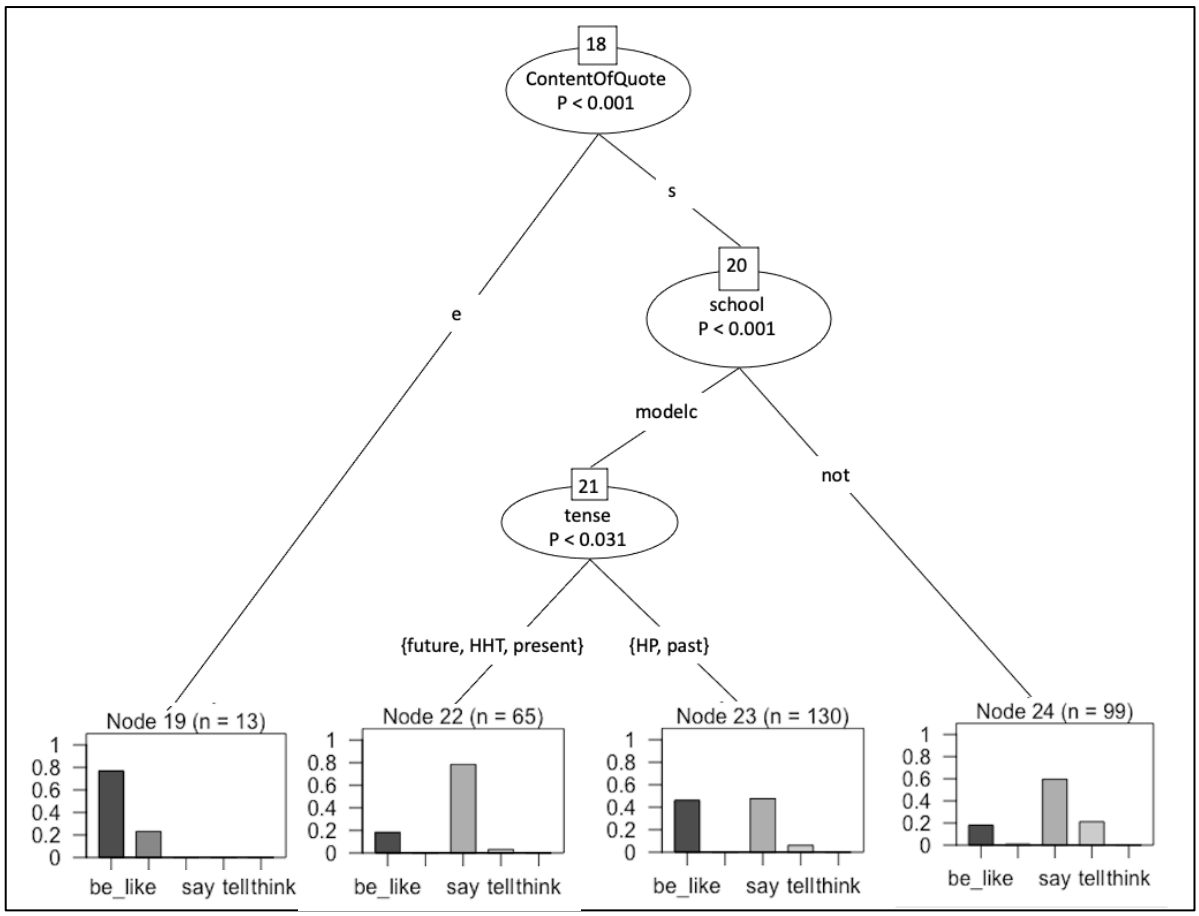
The quotes expressed by Black and White participants are then further divided according to *tense and temporal reference* at the ninth node ($p < 0.001$), in which HP quotative phrases are separated from future, HHT, past and present tense phrases. HP phrases are then split according to the *grammatical person* of the subject at the tenth node ($p = 0.013$). Node 11 depicts *be like* as the primary quotative used when HP quotative phrases contain first- or second-person subjects, although a few instances of *go*, *say* and *tell* are also shown. The same quotatives are used to express HP phrases with third-person subjects at the 12th node. However, although *be like* is still the primary quotative employed, it is used less frequently than for phrases with first- and second-person subjects, and there is a noticeably higher percentage of use of *say*, followed by *go* and a small number of *tell*.

Quotative phrases in future, HHT, past and present tense are divided at node 13 according to *content of quote* ($p < 0.001$). Node 17 shows that the quotatives used to express future, HHT, past and present tense phrases with quotes containing exclamations and non-lexicalised sounds are a very high percentage of *be like*, a small percentage of *go* and one instance of *think*. Quotes containing speech, however, undergo another division according to

tense and temporal reference at node 14 ($p < 0.001$). Those that are expressed in past tense phrases are depicted in node 15, showing use of *be like*, *say* and *tell*, whereas those expressed in future, HHT and present tense phrases are depicted in node 16, showing use of *be like*, *go*, *say* and *tell*. The majority quotative used in each is *be like*, followed by *say*, with lower occurrences of *go* and *tell*. However, both nodes 15 and 16 show lower percentages of *be like* than all other nodes containing quotatives expressed by young Black and White speakers, with node 16 showing the highest use of *say* and lowest use of *be like*.

Figure 5. 15

The right half of the left-hand segment of the right-hand side of the ctree, showing the distribution of quotatives used by Coloured and Indian speakers of the younger generation to express quotes containing speech or exclamations and non-lexicalised sounds (nodes 18-24)



Quotes expressed by Coloured and Indian speakers, on the other hand, are split according to *content of quote* once more. At node 18, quotes containing exclamations and non-lexicalised sounds are separated from those containing speech ($p < 0.001$). Node 19 depicts the quotes containing exclamations and non-lexicalised sounds, showing that *be like* and *go* were the only quotatives used in order to express them. Quotes containing speech are further divided

according to speakers' schooling at the 20th node. Quotatives expressed by Coloured and Indian speakers who went to FDET schools are shown in node 24 to primarily be *say*, followed by *tell*, *be like* and a couple instances of *go*. Quotatives expressed by speakers who received FMC schooling are split a final time according to *tense and temporal reference* at node 21. Node 22 shows quotative phrases in future, HHT and present tense primarily use *say*, followed by smaller numbers of *be like* and *tell*, whereas node 23 shows quotative phrases in HP and past tense only slightly favour the use of *say* over *be like* and also show a higher percentage of the use of *tell*.

5.3 Conclusion

The descriptive statistics, as well as the results of the ctree analysis clearly indicated that a clear generational change has taken place as younger speakers have adopted the *be like* quotative. The descriptive statistics showed that not only did the different generations use different proportions of the main quotatives, but used the quotatives differently, with the younger generation quoting more internal monologue, exclamations and non-lexicalised sounds. Age was the second most significant variable on the ctree generated from the database, dividing quotatives according to whether or not they were spoken by a participant from the older or younger generation.

The logistic regression analyses indicated that quotes containing speech and use of HP and past tense were significant linguistic variables that favoured the use of *be like*, whereas speaker schooling was marked as a significant social variable, with students who attended FDET schools favouring the use of quotatives other than *be like*. The fixed-effects multinomial and binomial models also indicated that speaker race and gender were significant social variables, with men, Coloured speakers and Indian speakers favouring the use of quotatives other than *be like*. However, when speaker was treated as a random variable in the mixed-effects model, these were no longer marked as significant. The use of intercepts did make some of the data challenging to read, as all intercepts were marked as significant, but p-values and log odds ratios were not provided for the reference levels of each variable.

The random forest generated supported the findings of the logistic regression analyses, showing that even in more interactive models *content of quote* was the most important variable by far for determining quotative choice, followed by *tense and temporal reference*, speaker schooling, speaker race, speaker gender and finally the *grammatical person* of the subject. The

ctree, made up only of statistically significant splits, echoed these findings and offered a more complex analysis of the variables determining quotative choice. The first split to occur was according to *content of quote*, separating quotes containing internal monologue from those containing speech or exclamations and non-lexicalised sounds. On both sides of the ctree the data was then divided according to age, showing the clear difference in quotative behaviour in younger and older speakers.

Among younger speakers, *think* and *be like* were the most commonly used quotatives for expressing inner monologue, but students who attended FDET schools used more *think* than *be like* and used small proportions of *say* and *tell*, whilst students who did attend FMC schools used more be like than think and used *go* twice. However, unlike with quoted internal monologue, when it came to quoted speech, exclamations and non-lexicalised sounds the behaviour of the younger generation was divided according to race.

Black and White speakers' behaviour was split according to *tense and temporal reference*, in which quotatives expressed using HP were separated from those in other *tense and temporal references*. When quotative phrases in HP were accompanied by a first- or second-person subject, *be like* was primarily used with very small proportions of *say*, *tell* and *go*. When they were accompanied by third-person subjects, *be like* was still the preferred quotative, but higher proportions of *say* and *go* were used alongside a similarly low proportion of *tell*. The quotatives not expressed using HP were then split according to whether they were quoting speech or exclamations and non-lexicalised sounds. Black and White speakers primarily used *be like* to express exclamations and non-lexicalised sounds, as well as smaller proportions of *go* and *think*. When they quoted speech, a smaller percentage of *be like* and a higher percentage of *say* was used. Speech quoted in the past tense used *be like* more frequently than *say* and used a small number of *tell* quotatives. Whereas speech quoted in future, HHT or present tense used a similar amount *be like* and *say* quotatives, as well as small quantities of *go* and *tell*.

Coloured and Indian speakers, on the other hand, all favoured *be like* and *go* for quoting exclamations and non-lexicalised sounds, but seemed to quote speech differently depending whether or not they had attended an FMC school. Those who had FDET schooling used *say* and *tell* more often than *like*. Those who attended FMC schools used *be like* more often, *tell* only very slightly more often and *say* less often when they were speaking in HP and past tense than when they were speaking in any other *tense and temporal references*.

6. Discussion and Concluding Remarks

This chapter considers the results of the study in contrast with other research discussed in the literature review. The findings clearly show the enthusiasm with which *be like* has been adopted into the quotative cohorts of Cape Town English speakers. Younger speakers use *be like* significantly more often than older speakers, whilst also quoting internal dialogue and using first-person subjects more frequently. The results indicate that the linguistic constraints affecting the use of *be like* are *content of quote*, *tense and temporal reference* and the *grammatical person* of the subject, with *content of quote* being the most significant and *grammatical person* the least. Not only did the use of quotatives differ between the younger and older speakers, amidst the younger generation there seemed to be differences according to the speakers' race and schooling. This chapter begins by reviewing the quotatives found in the data and proceeds to examine each of the linguistic and social variables individually.

6.1 Quotatives

The main quotatives were largely similar to those in the literature, although *go* seemed to be used more frequently and *ask* less frequently than they were reported to have been used in the literature. The quotatives *ask* and *tell* seemed to be more commonly used by Cape Town speakers than they were by speakers from other countries. *Tell* especially was seldom mentioned in the literature and neither were commonly included in the main quotatives category. *Ask* made up 1.8% of the quotatives and *tell* was used 3.3% of the time and included as one the main quotatives in the Cape Town data. Whereas *go* didn't seem to be as popular a choice among Cape Town speakers as it was in the literature. Making up only 2.8% of the quotatives, it was used least often of the main quotatives. Kohn and Franz (2009) were one of the few researchers to have included *tell* among the main quotatives they studied. They focused specifically on quotatives used by Black and Latino speakers and included multilingual speakers in their sample. The Cape Town database also included multilingual speakers and Coloured and Indian speakers expressed the highest proportion of *tell* quotatives, whilst White speakers expressed the lowest proportion of *tell* quotatives (see figure 4.7 in results). Further research could investigate whether there are any correlations between race, ethnicity or multilingualism and the use of *tell* as a quotative. Such a static correlation is not to be pursued for its own sake, but rather to confirm that (a) some groupings are quicker than others in

adopting global changes and (b) that high rates of adoption of the new quotatives comes at the cost of use of erstwhile standard forms like *tell* (at least in informal conversational style aimed at in the interviews).

Two of the innovations noted in the literature, both of which came into use at the same time as *be like* also appeared in the database of Cape Town English speakers. In the USA, the quotatives *be all* and *all like* were used to communicate speakers' internal monologues or attitudes in a similar manner to *be like*, but neither of them was adopted as widely and they seemed to decline in usage after the 1990s (Singler, 2001; Barbieri, 2005; Kohn & Franz, 2009). Each quotative, *be all* (56) and *all like* (57) was used only once and by different young, White, female speakers. It is possible that White speakers may feel most immersed in and connected to American media and ways of speaking, and could be the first to adopt and spread American innovations.

56. Ja, I had a friend and she would bring us all the leftovers after Eid, and we ***were all***, "Aah, this is so cool!"

57. So, she left on Friday ***all like***, "What am I gonna do?' 'Cause we see each other literally every single Monday."

Within the interviews, speakers also used quotative verbs as lexical verbs or for functions other than introducing a quote. The word *like* was used frequently in the interviews and its use was not restricted to the expression of *be like* quotatives. The examples below show words that are used as quotatives being used for other functions, including *like* being used according to its approximating function or its 'similar to' sense (58), according to its 'as if' sense (59) and according to its 'for example' sense (60) (Meehan, 1991).

58. That's the only memory I have from my primary - my first primary school. That was when I was <u>like</u> five, I think.

59. They think I act <u>like</u> I'm better than them, 'cause I go to UCT. Which is true.

60. She and another nurse, they do check-ups there, <u>like</u> deworming and vaccinations.

In addition to these approximating (this would be the 'similar to' specifically in reference to numbers) and comparison functions, *like* can also be a focus marker (61) or a means of hedging (62).

61. She was my standard one and standard two teacher. I was, <u>like</u>**,** her star student and I also loved her to bits and pieces 'cause I was her star student.

62. She enjoyed it, but it just wasn't in the right, like, it was in Brackenfell, so it just wasn't in the right area.

Like also occurred alongside other quotatives, for example in (63), but when such 'double quotatives' occurred, the first quotative was considered the primary one.

63. I said, like, "I'm really sorry I was very drunk and it's not -, don't take it to mean anything."

Other quotatives, such as *say* and *ask* in examples (64) and (65) were also used in other ways, such as introducing reported, rather than direct, speech.

64. He didn't want the nurses or doctors to touch him. He just said everyone must leave him alone.

65. Like this, two weeks ago, we had this conversation. She asked me when is my birthday, and I said, "Now in a few day's time," and, um, she asked, "What day?"

6.2 Explanatory Variables

This section reviews the results for each variable individually, beginning with the linguistic variables and following with the social variables, prior to a discussion of interactions between the variables. A random forest generated using the younger speakers' data determined the most significant variables for determining quotative choice, in descending order, to be *content of quote*, *tense and temporal reference*, speaker schooling, speaker race, speaker gender and *grammatical person*. These findings are in contrast with the findings from Tagliamonte et al.'s (2016) fixed-effects logistic regression models generated in GoldVarb, in which *tense and temporal reference* was the most significant linguistic variable, followed by *grammatical person* and then *content of quote*. It is possible that the inclusion of each of the main quotatives other than zero quotatives as well as the recursive partitioning technique employed in ctrees and random forests contributed to this difference. However, the logistic regression models returned similar findings. Although speaker age is not included in the logistic regression models and the random forest, the ctree generated using data from both generations of speakers showed age to be the second most significant split as, once the quotatives were divided according to *content of quote*, they were divided according to speaker age.

6.2.1 Linguistic variables

The linguistic variables are presented in order of significance, beginning with *content of quote*. *Content of quote* was found to be a significant variable, as the use *be like* was negatively correlated with quotes containing speech and more likely to be used to introduce quotes containing exclamations and non-lexicalised sounds. *Tense and temporal reference* too was an important variable, with the use of HP and past tense positively correlated with the use of *be like*. The logistic regression findings did not indicate that the *grammatical person* of the subject in the quotative phrase was a significant variable, but ctree results show that there was one instance in which it was significant. Among the Coloured and Indian speakers, first and second *grammatical person* resulted in higher percentages of *be like* among quotes expressing speech or exclamations and non-lexcalised sounds in HP.

6.2.1.1 Content of Quote

In the early stages of the grammaticalisation of *be like*, it was associated with quoting internal monologue (Ferrara & Bell, 1995; Blyth et al., 1990; Tagliamonte & Hudson, 1999). *Be like* continued to be highly correlated with the use of internal monologue as it diffused, but it also began to be used to introduce greater proportions of quotes containing speech and non-lexicalised sounds (Tagliamonte, 2004). In Tagliamonte et al. (2016), *content of quote* remained one of the three significant factors correlated with the use of *be like*, but it was the weakest of the three, with *tense and temporal reference* having the strongest effect, followed by *grammatical person* of the subject in the quotative phrase.

One of the changes observable in the data was the way in which both the quotatives and the content of the quotes they introduce differ between the younger and older generation. Among the younger generation, *be like*, *go*, zero and other quotatives were used to introduce all three kinds of quoted content: speech, internal monologue and exclamations and non-lexicalised sounds. The multinomial and binomial logistic regression models found *content of quote* to be a significant factor. The multinomial model found that, when quoting internal monologue, *think* was favoured over *be like* ($p < 0.01$) and *be like* was favoured over *go* ($p < 0.01$) and the binomial model found that quotatives other than *be like* were favoured when quoting direct speech ($p < 0.001$). Whilst the logistic regression models don't return p-values for the reference levels of each variable, the factor weights and effect plots (figures 5.5 and

5.7) implied that *be like* was highly correlated with the quotes containing exclamations and non-lexicalised sounds.

The ctree provided more information on how all quotatives were selected according to the content of the quotes they introduced. As the first split on the ctree, content was found to be the most significant variable for determining quotative choice ($p < 0.001$). *Be like* was used approximately 80% of the time that exclamations and non-lexicalised sounds were quoted. The use of *be like* being was similarly high for quotes containing internal monologue, reaching significance in the speech of young participants who attended FMC schools ($p < 0.001$).

Whereas young people with different schooling showed a preference for *think* and the older White participants solely use *think* to express internal monologue. The probability of using *be like* was considerably lower for quotes containing direct speech, unless they happen to be in the HP.

6.2.1.2 Tense

Whilst the initial studies of the grammaticalisation of *be like* focused primarily on its use for quoting internal monologue and non-lexicalised sounds, for dramatic effect and its occurrence alongside first-person subjects (Blyth et al., 1990; Ferrara & Bell, 1995; Tagliamonte & Hudson, 1999), it was also observed early on that *be like* tended to occur in present tense phrases (Blythe et al., 1990; Romaine & Lange, 1991). Later, there was greater emphasis on and investigation into the ways in which *tense and temporal reference* constrained the use of *be like*, specifically noting the correlation between the use HP and the use of *be like* (Tagliamonte & D'Arcy 2007; Buchstaller & D'Arcy, 2009; D'Arcy, 2010; Tagliamonte et al., 2016). However, similar to the gender effect, this has not been a universal trend as some speech communities, such as Maori men in D'Arcy's (2010) study and the British English speakers in D'Arcy and Buchstaller's (2009) study disfavoured the use of *be like* with HP. The correlation of *be like* with HP quotative phrases does, however, appear in the Cape Town data.

The descriptive statistics indicated that past tense was the most commonly used *tense and temporal reference* for all quotatives other than *go*, which primarily occurred in HP. However, Black and White speakers used a noticeably high proportion of HP when expressing *be like* quotatives, and Coloured and Indian participants seemed to use a higher proportion of present tense when expressing *say* than Black and White participants. The multinomial and binomial logistic regression analyses found *tense and temporal reference* to be a significant

variable. The multinomial model showed that *be like* was significantly preferred over *say* ($p < 0.001$) and *go* ($p < 0.05$) when the quotative phrase was in HP and *be like* was preferred over *go* ($p < 0.001$) when the quotative phrase was in the past tense. The binomial model found that *be like* was significantly more likely to be chosen over other quotatives when the phrase was in past tense ($p < 0.05$) and especially when the phrase was in HP ($p < 0.001$).

The ctree also indicated that *tense and temporal reference* was a significant variable for quotes containing speech or exclamations and non-lexicalised sounds, although it didn't significantly constrain quotes containing internal monologue. The findings show that young Black and White speakers expressing speech or exclamations and non-lexicalised sounds used the highest quantities of *be like* when using HP, especially if they also used a first- or second-person subject, in which case *be like* was used almost 90% of the time ($p < 0.001$). When expressing speech, Black and White participants also used a greater proportion of *be like* when using past tense than when they were using future, HHT or present tense ($p < 0.001$). *Tense and temporal reference* was only significant ($p < 0.031$) among young Coloured and Indian participants who had been to an FMC school. These speakers showed a significantly higher use of *be like* when using with HP or past tense than when they used other *tense and temporal references* ($p = 0.031$).

During the data entry phase a number of quotative phrases using habitual, hypothetical or timeless forms were noticed and coded as 'HHT' tense, forming what might otherwise have been called an 'other' category. As the possibility of an association of *be like* with hypothetical speech (Romaine & Lange, 1991; Tagliamonte & D'Arcy, 2004) or the use of the continuous aspect (Blyth et al., 1990), these were included in order to investigate possible correlations with *be like*. Although both binomial logistic regression models indicated that HHT was positively correlated with *be like*, neither of these results were statistically significant ($p = 0.35684$ and $p = 0.578443$).

6.2.1.3 Grammatical Person

In the initial grammaticalisation of *be like*, the innovative quotative was marked by its use alongside first-person subjects (Ferrara & Bell, 1995; Blyth et al., 1990; Tagliamonte & Hudson, 1999). As *be like* was more widely adopted, there was a degree of semantic bleaching and expansion as it came to be used more often alongside third-person subjects. However, despite this, first-person subjects remained the preference, possibly due to changes to narrative

styles which depended on enacted recollections of a speaker's thoughts (Buchstaller & D'Arcy, 2009; Tagliamonte et al., 2016).

Despite the emphasis on *grammatical person* within the literature, this variable didn't display the same significance in the Cape Town data. Whilst the data displayed a greater number of first-person subjects accompanying quotatives expressed by the younger generation than the older generation and *tell* seemed to be used more often with first-person subjects than third-person subjects, *grammatical person* did not significantly affect the binomial and multinomial logistic regression analyses and was thus omitted from those models. The random forest showed *grammatical person* to be the least important of the variables for predicting quotative choice, with a conditional importance of only 0.056. However, *grammatical person* did have a significant effect for young Black and White speakers who, as mentioned above, used a higher proportion of *be like* quotatives when expressing speech, exclamations and non-lexicalised sounds in HP when the quotative phrase included a first- or second-person subject than when it included a third-person subject ($p = 0.013$).

6.2.1.4 Mimetic Effect

Initial observations about early grammaticalisation of *be like* noted its frequent occurrence alongside first-person subjects, internal monologue and non-lexicalised sounds, as well as its use for 'dramatic effect' (Blyth et al., 1990; Ferrara & Bell, 1995; Tagliamonte & Hudson, 1999). Both HP and mimetic re-enactment are used in order to create or enhance the dramatic effect of a narrative. As the data shows that young Cape Town speakers do tend to use HP with *be like* very often, it would be interesting to investigate to what degree they also use mimetic re-enactment.

6.2.2 Social variables

The social variables of gender, race and schooling are discussed below in order to determine whether there are differences in quotative usage between specific groups or communities of young speakers. Speaker age was included in order to differentiate between the younger speakers, all of whom were below 35 years old, and the older generation (over 50 years old) who represented speakers less likely to have participated in the quotative usage. When age was included as a variable it was marked as significant. Gender was significant in the fixed-effect logistic regression models, but not in the mixed-effect model. Race and

schooling were both found to be significant variables in the fixed-effect logistic regression models and the ctree model. However, in the mixed-effects model, race was not found to be significant, and schooling decreased in significance.

6.2.2.1 Age

In the 1990s, it was reported that speakers decreased their use of *be like* after the age of 25 (Blythe et al., 1990) and for a time it was considered likely that *be like* would prove to be an age-graded change (Barbieri, 2009). However, speakers were soon reported to continue the change into their thirties (Singler, 2001) and Barbieri (2009) found that speakers, particularly women, maintained their use of *be like* into their forties. Unfortunately, the Cape Town data did not include any speakers in their forties but, as *be like* was used once by two different members of the older group and was used often by younger speakers in their thirties, the results seem to support the hypothesis that the adoption of *be like* is not an age-graded change. One of the speakers who used *be like* in the older group works at a university and it is possible they could have been influenced by their students. The second use of *be like* by a speaker in the older group was accompanied by a dummy *it*, a combination which Fox and Robles (2010) note is particularly associated with mimetic re-enactment.

There is a clear difference in quotative usage between the younger and older generation. There is little further insight to be gained, as age was entered as a categorical, rather than continuous, variable. The original survey was interested in young people's speech, and older speakers were only inserted as a contrast to confirm the differences in young people's speech. Whilst the younger speakers continued to use the traditional quotatives, *say* and *think*, the proportion of these quotatives was significantly decreased from 71% to 29% and 12% to 4%, respectively. *Be like* was the primary quotative used by younger speakers but, making up only 59% of all quotatives, it was not favoured to the extent that *say* was favoured by older speakers at 71%. Crucially, the younger group of speakers seemed to be using quotatives to fulfil different functions to the previous generation. There is an observable increase in the use of quotatives to cite or re-enact internal monologue, non-lexicalised sounds and exclamations among younger speakers. Internal monologue makes up 19% of the quotes expressed by the younger group and are introduced using *be like*, *think, say,* zero quotatives and other quotatives, whereas it made up 11% of the quotes expressed by the older group and was only expressed using *think* and zero quotatives. Non-lexicalised sounds and exclamations made up 5% of the

quotes expressed by the younger generation, using *be like*, *go*, *think*, zero quotatives and other quotatives, and 1% of the quotes expressed by the older generation, using only *say* and *think*. The significance of the difference between generations is confirmed in the ctree, where age is the second most important variable for determining quotative choice and forms the second split on both branches of the ctree ($p < 0.001$).

6.2.2.2 Gender

There is no consensus about the effect of gender on *be like* within the literature. Whilst there are those such as Romaine and Lange (1991), Ferrara and Bell (1995), Tagliamonte and Hudson (1999) and Singler (2001) who found that female speakers used *be like* more often than male speakers and seemed to be leading the change, there are also accounts of men using more *be like* than women (Blyth et al., 1990; Daily-O'Cain, 2000). In the data used for this dissertation, *be like* does appear to be more popular among female speakers, but not significantly so.

The descriptive statistics indicated that, although the majority of quotatives were expressed by female speakers, the quotatives *say*, *think* and *tell* were more often expressed by White male speakers than White female speakers. The bar charts in chapter 4 seemed to indicate that patterns of quotative usage differed according to social factors like the speakers' race or schooling rather than gender. For example, Black and White speakers of both genders produced more *be like* quotatives than *say* quotatives, and Coloured and Indian speakers of both genders produced more *say* quotatives than *be like* quotatives.

The fixed-effects multinomial and binomial logistic regression analyses indicated that gender had a significant effect on quotative use. The multinomial model found that men favoured *say* ($p < 0.001$), *tell* ($p < 0.05$) and *think* ($p < 0.05$) over *be like* and the binomial model found that men tended to prefer the other main quotatives over *be like* in general ($p < 0.001$). However, these models do not consider the random effect of individual participants, so an individual who produced a lot of quotatives could skew the results. In the mixed-effects binomial regression model, which did include speaker as a random variable, gender was not identified as a significant variable. The ctree similarly did not include gender, indicating that there were no significant splits made in the data according to the gender of speakers (therefore $p > 0.05$).

6.2.2.3 Race

Neither the race nor the ethnicity of speakers was commonly included as a social variable within the literature, although there were a handful of studies that focused explicitly on ethnicity or race. Cukor-Avila (2012) hypothesised in her Springville study that White teenagers were the first to adopt *be like* and from there, through shared environments like schools, it spread into both urban and rural AAVE speaker communities, and *be like* was adapted according to the structural norms of AAVE. D'Arcy (2010), on the other hand, found that there were significant differences in the ways that White Pakeha men used quotatives and the way that Maori men used quotatives, with Pakeha men preferring *be like* and Maori men preferring zero quotatives. The context in which they each used zero quotatives was also different, with Maori men relying on mimetic effect to convey different speakers' identities as they recounted a series of dialogue, whereas Pakeha men tended to use a lexical verb prior to the zero quotative that still implied quotation. Kohn and Franz's (2009) study also focuses specifically on speakers of different races: a Latino community and an African American community. They found that within those communities *be like* wasn't as constrained by *tense and temporal reference* as it was in the rest of the literature at the time. As mentioned above, their study is also one of the few others that included *tell* as one of the main quotatives examined.

As mentioned above (in 6.2.2.2), an examination of the descriptive statistics seemed to indicate that patterns of quotative usage differed more according to race than gender. The use of quotatives when expressing speech, exclamations and non-lexicalised sounds, did seem to differ quantitatively according to speakers' race, with the behaviour of Black and White speakers being different to Coloured and Indian speakers. Both show the highest proportion of *be like* when using HP. Black and White use more *be like* in general, whereas Coloured and Indian use more *say*. Coloured and Indian use more *tell* and show more specific use of *go* which is used in a high proportion for exclamations and non-lexicalised sounds but very seldom used for speech, particularly among those who had FDET schooling.

The fixed-effects binomial logistic regression model found that Coloured speakers seemed to favour the other main quotatives over *be like* (0.001) and the multinomial model found that both Coloured ($p < 0.01$) and Indian ($p < 0.01$) speakers preferred the use of *say* over *be like*. Coloured speakers also preferred the use of *tell* ($p < 0.01$). Similarly to gender, these race effects were no longer reported as significant when participant was included as a random variable in the mixed-effects binomial logistic regression model. However, unlike gender, race

did seem to be a significant social factor in the more interactive ctree model. When it came to the expression of quotes containing speech, non-lexicalised sounds and exclamations, the ctree distinguished between the behaviour of Black and White participants and the behaviour of Coloured and Indian participants, showing that Black and White speakers used a higher percentage of *be like* quotatives.

For Black and White speakers *be like* was the quotative of preference in HP ($p < 0.001$) and first- or second-person environments ($p = 0.013$). In HP and third-person environments *be like* was still the most commonly used, but there were higher proportions of *say* and *tell* used than with first and second person. In the other *tense and temporal references*, there was a split between speech and exclamations and non-lexicalised sounds ($p < 0.001$). *Be like* was the preferred quotative for expressing exclamations and non-lexicalised sounds but it was also accompanied by small numbers of *go* and *think*. Even for expressing quoted speech, *be like* was the quotative of choice, although there was a higher proportion of say and lower proportion of *be like* used. When in past tense environments, there were more instances of *be like* than *say*, as well as a few instances of *tell* ($p < 0.001$). Quoted speech in future, HHT and present tense environments was expressed in similar quantities by both *be like* and *say* quotatives, as well as small numbers of *go* and *tell* ($p < 0.001$).

Coloured and Indian speakers primarily used *be like* (but also use a smaller quantity of *go*) when expressing non-lexicalised sounds and exclamations ($p < 0.001$). There seemed to be a preference for using *say* to express speech, although this behaviour differed according to schooling. Those who went to FDET schools express speech using a majority of *say* followed by *tell* which is closely followed by *be like* and a very small number of *go* ($p < 0.001$). Speakers who did go to FMC schools use a similar amount of *be like* and *say* quotatives if they're expressing quoted speech in either HP or past tense, but use a significantly higher proportion of *say* if they are expressing speech in HHT, future or present tense ($p = 0.031$). In general, those who went to FMC schools use a smaller proportion of *tell* quotatives to express quoted speech than those who went to FDET schools.

6.2.2.4 Schooling

In the descriptive linguistics, the most discernible difference between participants who attended an FMC school and those who didn't was that the former expressed fewer instances of *tell*. Both the multinomial and binomial logistic regression analyses indicated that speakers'

schooling was a significant variable. The multinomial model reported that that *say* ($p < 0.01$), *tell* ($p < 0.001$) and *think* ($p < 0.01$) were favoured over *be like* by speakers with FDET schooling. Schooling was also significant in the binomial models, both with FDET students preferring the other main quotatives over *be like*, although the degree of significance decreased in the mixed-effects model ($p < 0.001$ and $p < 0.05$, respectively). Schooling was also significant in the ctree model. When quoting inner monologue, speakers who attended FMC schools used *be like* roughly 80% of the time and *think* roughly 20% of the time, whereas speakers who had different schooling used *think* roughly 60% of the time and *be like* about 30% of the time ($p < 0.001$). There was further differentiation according to schooling when Coloured and Indian speakers quoted speech ($p < 0.001$). Those who had FMC schooling showed a slight preference for *say* but used *say* and *be like* each just below 50% of the time when they used HP. When they were using any of the other *tense and temporal references*, speakers with FMC schooling used *say* roughly 80% of the time and *be like* 20% of the time, with an even smaller proportion of *tell*. Speakers who attended FDET schools primarily used *say*, which made up 60% of the quotatives in this context, whilst *tell* was used just over 20% of the time and *be like* just below 20% of the time.

6.3 Possible Implications

The use of *be like* as a quotative is a global change that shows adaptation to each local system it is used within. Whilst there seems to be a global shift toward a more self-referential and dialogic style of narration (Tagliamonte and D'Arcy, 2007) among younger and possibly (now) middle-aged speakers, as this trend is adopted by speakers it is still affected by their pre-existing style and functions for which they use quotatives. *Be like* is used differently based on a community's pre-existing narrative and quotation style. For example, in Tagliamonte and Hudson (1999), British speakers' more introspective style and Canadian speakers' more action-oriented styled impacted the quotatives that were more popularly used in each region. Similarly, Davydova (2019) found that the different ways English was perceived and the uses it was put to in India and Germany influenced the quotatives that were used in each community. D'Arcy (2010) and Cukor-Avila (2012) both showed the ways that *be like* adapted according to the pre-existing configurations of their local varieties, AAVE, Pakeha and Maori English. Naturally, a similar process of adaptation will have taken place in Cape Town.

It is possible that the greater use of *tell* and lower use of *go* are representative of a pre-existing local style. As they seem to occur particularly among Coloured and Indian speakers who attended FDET schools, it would be useful to assess how frequently they are used in older generations of those communities. Not unlike the Maori men who used more zero quotatives and fewer instances of *be like*, due to their pre-existing quotative style (D'Arcy, 2010), Coloured speakers also used the highest percentage of zero quotatives, followed by Indian speakers, whilst White and Black participants used the fewest.

Kohn and Franz (2010) also found that *tense and temporal reference* was not as significant a variable in their populations as it was in the literature. Whilst *tense and temporal reference* is undoubtedly a significant variable in these data, *grammatical person* did not impact quotative choice as significantly for Cape Town speakers as it did elsewhere. Interestingly, Davydova (2019) suggested that *grammatical person* and mimetic re-enactment were more easily acquired constraints, as they weren't as semantically and pragmatically complex as tense and *content of quote*, yet Cape Town speakers appear to have adopted *tense and temporal reference* and *content of quote* constraints more easily than *grammatical person*.

The *grammatical person* of the subject in a quotative phrase was only significant for young Black and White speakers when they were using HP to express quoted speech, non-lexicalised sounds or exclamations. In other contexts and for other speakers, *grammatical person* did not significantly affect quotative choice. Young Black and White speakers used the highest proportion of *be like*, as well as the most significant HP effect and the only instance of significant correlation between an increase in *be like* and the use of first grammatical person. Speakers who had attended an FMC school also showed greater use of *be like* according to the variable constraints reported in the literature, using it as the primary quotative for introducing inner monologue and among Coloured and Indian FMC students there was *tense and temporal reference* differentiation which was not present in the speech of those from FDET schools. However, all young participants showed a preference for *be like* when quoting exclamations and non-lexicalised sounds.

It is possible that, similarly to the students in Springville (Cukor-Avila, 2012), White speakers initiated the adoption of *be like* in Cape Town. It is likely that American and British English carry a degree of prestige in South Africa that could be compared to the prestige Davydova (2019) found English carried in Germany. Through international interactions and media exposure, White South Africans may have acquired the use of *be like*. The emerging Black middle class, who use the new crossover varieties (Mesthrie, 2010; 2017), have likely

acquired *be like* by similar means to White speakers, and possibly also through interactions with those White South Africans. The acquisition of *be like* and its documented linguistic constraints is most apparent among speakers who attended FMC schools. This could be because FMC schools offer an advantage for those hoping to attend prestige higher education institutions as well as the fact that they provide a space of daily interaction between pupils who would not have interacted during apartheid.

D'Arcy discusses the way in which quotatives are used to index ethnic identities and belonging, whilst clearly stating that her intention was to discuss a shared repertoire of features that is used differently by different speakers in order to index ethnic identities or characteristics of ethnic identities, and not to imply there are fixed, separate ethnolects. Likewise, this study does not intend to overemphasise race or ethnicity as static and dichotomous, but to find explanations in the availability of new global trends and the extent to which different students feel a readiness to identify with them. Quotatives, particularly *be like* and *tell*, appear to be used similarly in Cape Town. Whilst the social meaning of variables is ever shifting, one could speculate about the characteristics within the indexical field of *tell* and *be like* (Eckert, 2012). It is plausible that speakers, whether consciously or not, have come to associate frequent use of *be like* quotatives, especially in HP environments or prior to quotes containing inner monologue or exclamations and non-lexicalised sounds, with characteristics evocative of America, success, middle-class belonging, mobility, or wealth. Whereas use of *tell* may be associated with local community belonging or, as it is an older verb, its use among Coloured and Indian participants may indicate a degree of resistance to using new, global styles of quotation.

However, the changes seen in the mixed-effects model demonstrate that social variables, particularly gender and race, were not as significant as the fixed-effects models implied. The use of a mixed-effects model, as well as the incorporation of models that use binary recursive partitioning, allow for a clearer understanding of the importance of each variable, whilst minimising skewing effects that individual participants may have on the data. The way in which trees are able to process data that includes many potential interactions between variables allows for researchers to develop an understanding of the specific context and speaker groups in which certain quotatives or quotative constraints are most present. The results of the ctree revealed that *be like* was the most common quotative used for the expression of quotes containing exclamations and non-lexicalised sounds by all participants. Most participants, with the exception of Coloured and Indian participants who attended FDET

schools, also showed a marked increase in the proportion of *be like* to other quotatives when using HP to quote speech. When expressing quotes containing inner monologue, all participants who attended FMC schools favoured the use of *be like*, whereas participants who attended different schools favoured *think*.

6.4 Conclusion

The quotative *be like* has been enthusiastically adopted by young speakers, especially Black and White participants and participants who attended FMC schools. The younger speakers show a much higher usage of *be like* and a decrease in the use of *say* and *think*, whilst also increasing the proportion of quotes containing their own thoughts and non-lexicalised sounds and exclamations. The use of *be like* by participants in their thirties, as well as two instances of the quotative in the older control group indicated that this is not an age-graded change.

When all the main quotatives are considered together as a quotative cohort, the most important variable for determining quotative choice is *content of quote*. Among all participants, quotes containing exclamations and non-lexicalised sounds were paired most often with *be like*. Among participants who had attended FMC schools, quotes containing inner monologues were primarily introduced by *be like*. However, when *be like* was contrasted with the other main quotatives as a group, the most important variable was *tense and temporal reference*. The *tense and temporal reference* effect was observed primarily in the expression of quotes containing speech. Among Black and White participants, as well as Coloured and Indian participants who had attended FMC schools, quotative phrases in HP contained a higher proportion of *be like* quotatives than quotative phrases in the other *tense and temporal reference*s. *Grammatical person* was not as significant in the Cape Town data as it was reported to be in other studies in the literature. It only had a significant effect on quotative choice among Black and White speakers using HP, for whom the inclusion of a first- or second-person subject resulted in a higher proportion of *be like* quotatives.

Although the linguistic constraints on *be like* differed somewhat between speakers of different races, they did not seem to differ greatly between speakers of different genders. Of the social variables, the gender of the speaker proved to be the least important variable for determining the use of *be like*. Whilst the fixed-effects logistic regression models reported gender as significant, neither the mixed-effects model nor the ctree included speaker gender as

a significant factor. The results for speaker race were similar in that race was reported as significant in the fixed-effects models, but not in the mixed-effects model. However, the greater ability of the ctree to analyse data containing multiple interactions revealed that there was a significant difference, at least for quotes containing speech, exclamations and non-lexicalised sounds, between the quotative practice of Black and White speakers on the one hand, and Coloured and Indian speaker on the other. Black and White participants used *be like* more prolifically than Coloured and Indian participants, whereas Coloured and Indian participants used a greater percentage of *tell* and zero quotatives.

This distinction was not based solely on the race of speakers, as FMC schools appeared to be grounds for disseminating and accelerating the acquisition of *be like*. The difference between participants who had attended FMC schools and those who did not is observed most clearly in the expression of quotes containing inner monologue, for which FMC students showed a strong preference for the use of *be like* and students from FDET schools showed a strong preference for the use of *think*. Among Indian and Coloured participants, those who had attended an FMC school used *be like* more often and *tell* less often than those who had attended FDET schools.

There are rich opportunities for further studies of quotatives within South Africa. Future research could further investigate the use of zero quotatives among Cape Town speakers. The inclusion of mimetic effect as a linguistic variable could also provide valuable insight into the use of quotatives in South Africa. An analysis of the sociolinguistic interviews with participants from different South African cities within Professor Rajend Mesthrie's Database of South African English would allow for an understanding of the regional variation quotative practices. The collection of more recent data could be used for a real-time study of the acquisition and grammaticalisation of *be like* and the inclusion of ethnographic research methods may allow for further investigation into the indexical fields of quotatives such as *be like* and *tell*.

Bibliography

Barbieri, F. 2005. Quotative use in American English: A corpus-based, cross-register comparison. *Journal of English Linguistics*. 33:222-256

Barbieri, F. 2007. Older men and younger women: A corpus-based study of quotative use in American English. *English World-Wide*. 28(1):23-45

Barbieri, F. 2009. Quotative be like in American English: Ephemeral or here to stay? *English World-Wide*. 30(1):68-90

Bates, D., Maechler, M., Bolker, B. and Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*. 67(1): 1-48

Bell, A., & Sharma, D. (Eds.). 2014. Debate: Media and language change. *Journal of Sociolinguistics*. 18(2): 213–286.

Benor, S. 2010. Ethnolinguistic repertoire: Shifting the analytic focus in language and ethnicity. *Journal of Sociolinguistics.* 14(2): 159-183

Bevans, R. 2020, March 26. An introduction to Akaike information criterion. *Scribbr.* Available from https://www.scribbr.com/statistics/akaike-information-criterion/

Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. 1999. *The Longman Grammar of Spoken and Written English*. London: Longman.

Blommaert, J. 2003. Commentary: A sociolinguistics of globalization. *Journal of Sociolinguistics*. 7(4):607-623

Blommaert, J. 2010. *The Sociolinguistics of Globalization*. Cambridge University Press.

Bloomfield, L. 1933. *Language*. New York, NY: Holt, Rhinehart and Winston.

Blyth, C., Recktenwald, S. & Wang, J. 1990. I'm like, "Say What?!": A new quotative in American Oral Narrative. *American Speech*. 65(3):215-227

Bogetić, K. 2014. *Be like* and the quotative system of Jamaican English: Linguistic trajectories of globalization and localization. *English Today 119.* 30(3):5-12

Bucholtz, M., & Hall, K. 2005. Identity and interaction: a sociocultural linguistic approach. *Discourse Studies. 7*(4/5): 585–614

Buchstaller, I. 2001. An alternative view of 'like': Its grammaticalization in conversational American English and beyond. *Edinburgh Working Papers in Applied Linguistics*. 11:21-41

Buchstaller, I. 2004. The sociolinguistic constraints on the quotative system – U.S. English and British English compared. Unpublished PhD dissertation. Edinburgh, U.K.: University of Ediburgh.

Buchstaller, Isabelle. 2006a. Diagnostics of age-graded linguistic behaviour: The case of the quotative system. *Journal of Sociolinguistics*. 10: 3–30.

Buchstaller, Isabelle. 2006b. Social stereotypes, personality traits and regional perception displaced: Attitudes towards the 'new' quotatives in the U.K. *Journal of Sociolinguistics*. 10: 362–381.

Buchstaller, I. 2008. The localization of global linguistic variants. *English World-Wide*. 29(1):15-44

Buchstaller, I. 2015. Exploring linguistics malleability across the life span: Age-specific patterns in quotative use. *Language in Society*. 44: 457-496

Buchstaller, I. & D'Arcy, A. 2009. Localized globalization: A multi-local, multivariate investigation of quotative be like. Journal of Sociolinguistics. 13(3):291-331

Butters, R. 1982. Editor's note [on "be+like"]. *American Speech*. 57:149

Chambers, J.K. 1992. "Linguistic correlates of gender and sex". *English World-Wide* 13: 173–218.

Coupland, Nikolas. 2009. The mediated performance of vernaculars. *Journal of English Linguistics*. 37:284–300.

Croissant, Y. 2020. mlogit: Multinomial Logit Models (R package version 1.1-0) [Computer software]. *The Comprehensive R Archive Network.* Available from https://CRAN.R-project.org/package=mlogit

Cukor-Avila, P. 2002. *She say, she go, she be like*: Verbs of quotation over time in African American Vernacular English. *American Speech*. 77:3-31

Cukor-Avila, P. 2012. Some structural consequences of diffusion. *Language in Society*. 41(5):615-640

Daily-O'Cain, J. 2000. The sociolinguistic distribution of and attitudes toward focuser *like* and quotative *like*. *Journal of Sociolinguistics*. 4(1):60-80

D'Arcy, A. 2004.The development of linguistic constraints: Phonological innovations in St. John's English. *Language Variation and Change*.

D'Arcy, A. 2010. Quoting ethnicity: Constructing dialogue in Aotearoa/New Zealand. Journal of Sociolinguistics. 14(1):60-88

Davydova, J. 2019. Quotative *like* in the Englishes of the Outer and Expanding Circles. *World Englishes*. 38: 578– 592

Dictionary Unit for South African English. 2020. Smokkel. In *Dictionary of South African English*. Available from https://dsae.co.za/entry/smokkel/e06641

Eckert, P. 2008. Where do ethnolects stop? *International Journal of Bilingualism*. 12:25-42

Eckert, P. 2012. Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation. *Annual Review of Anthropology*. 41(1): 87–100

Erman, B. 2001. Pragmatic markers revisited with a focus on *you know* in adult and adolescent talk. *Journal of Pragmatics*.33:1337-1359

Ferrara, K. & Bell, B. 1995. Sociolinguistic variation and discourse function of constructed discourse introducers: The case of *be + like*. *American Speech*. 70(3):265-290

Fleischman, Suzanne, and Marina Yaguello. 2004. "Discourse Markers Across Languages: Evidence from English and French." In *Discourse across Languages and Cultures*, ed. Carol Lynn Moder and Aida Martinovic-Zic. Amsterdam: Benjamins. 129–148.

Fox, B. & Robles, J. 2010. It's like mmm: Enactments and it's like. *Discourse Studies*. 12(6):715-738

Fox, J. 2003. Effect Displays in R for Generalised Linear Models. Journal of

Statistical Software. 8(15): 1-27

Fox, J. and Weisberg, S. 2018. Visualizing Fit and Lack of Fit in Complex Regression Models with Predictor Effect Plots and Partial Residuals. *Journal of Statistical Software*. 87(9): 1-27

Fox, J. and Weisberg, S. 2019. *An R Companion to Applied Regression* (3rd ed.). Sage Publications.

Gelman, A. 2005, January 25. Why I don't use the terms 'fixed and random effects'. *Statistical Modelling, Causal Inference and Social Science*. Available from https://statmodeling.stat.columbia.edu/2005/01/25/why_i_dont_use/

Golato, Andrea. 2000. "An Innovative German Quotative for Reporting on Embodied Actions: *Und ich so/und er so* 'and I'm like/and he's like'." *Journal of Pragmatics*. 32: 29–54.

Harrell, F.E. Jr. 2020. rms: Regression Modeling Strategies (R package version 6.0-1) [Computer software]. *The Comprehensive R Archive Network*. Available from https://CRAN.R-project.org/package=rms

Hopper, P. 1991. On some principles of grammaticalization. In Elizabeth Traugott and Bernd Heine (eds.) *Approaches to Grammaticalization*. Amsterdam and Philadelphia: John Benjamins. 17-35

Hopper, P.J. & Traugott, E. C. 2003. *Grammaticalization* (2nd ed.). Cambridge University Press.

Hosmer, D.W., & Lemeshow, S. 2000. *Applied Logistic Regression*. New York: Wiley.

Hothorn, T., Buehlmann, P., Dudoit, S., Molinaro, A. and Van Der Laan, M. 2006. Survival Ensembles. *Biostatistics*. 7(3): 355-373

Hothorn, T., Hornik, K. and Zeileis, A. 2006. Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics*. 15(3): 651-674

Hothorn, T. and Zeileis, A. 2015. partykit: A Modular Toolkit for Recursive Partytioning in R. *Journal of Machine Learning Research*. 16: 3905-3909

Hothorn, T., Seibold H. and Zeileis, A. 2019. Partykit Reference Manual: A Toolkit for Recursive Partytioning. *The Comprehensive R Archive Network*. Available from https://cran.r-project.org/web/packages/partykit/partykit.pdf

Johnson, D.E. n.d. Rbrul Manual. *Daniel Ezra Johnson*. Available from http://www.danielezrajohnson.com/Rbrul_manual.html

Johnson, D.E. 2009. Getting off the GoldVarb Standard: Introducing Rbrul for Mixed-Effects Variable Rule Analysis. *Language and Linguistics Compass*. 3: 359-383

Johnstone, B. 1987. "He says… so I said": Verb tense alternations and narrative depictions of authority in American English. *Linguistics*. 25:33-52

Johnstone, B. et al. 2006. Mobility, Indexicality, and the Enregisterment of 'Pittsburghese'. *Journal of English Linguistics*. 34(2): 77–104

Klewitz, G. & Couper-Kuhlen, E. 1999. Quote-unquote? The role of prosody in the contextualization of reported speech sentences. *Pragmatics*. 9(4):459-485

Kohn, M. & Franz, H. 2009. Localized patterns for global variants: The case of quotative systems of African American and Latino speakers. *American Speech*. 84(3):259-297

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*. 82(13): 1-26

Labov, W. 1994. *Principles of Linguistic Change*. Vol. 1: Internal Factors. Oxford and Cambridge: Blackwell.

Labov, W. 2001. *Principles of Linguistic Change*. Vol. 2: Social Factors. Malden/MA: Blackwell.

Labov, W. 2003. 'Pursuing the cascade model', in D. Britain and J. Cheshire (eds), *Social Dialectology: In Honor of Peter Trudgill*. Amsterdam: Benjamins. 9-22

Labov, W. 2007. Transmission and Diffusion. *Language*. 83(2):344-387

Laserna, C., Seih, Y. & Pennebaker, J. 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

Levey, S., Groulx, K. & Roy, J. 2013. A variationist perspective on discourse-pragmatic change in a contact setting. *Language Variation and Change*. 25:225-251

Levshina, N. 2015. *How to do Linguistics with R: Data exploration and statistical analysis*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Louro, C. 2013. Quotatives down under: *Be like* in cross-generational Australian English speech. *English World-Wide*. 34(1):48-76

Macauley, R. 2001. *You're like 'why not?'*: The quotative expression of Glasgow adolescents. *Journal of Sociolinguistics*. 5:3-21

Meehan, T. 1991. It's like, "What's happening in the evolution of *like*?": A theory of grammaticalization. *Kansas Working Papers in Linguistics*. 16:37-51

Mesthrie, R. 2010. Socio-phonetics and social change: Deracialisation of the GOOSE vowel in South African English. *Journal of Sociolinguistics*. *14*(1): 3–33

Mesthrie, R. 2012. Ethnicity, substrate and place: The dynamics of Coloured and Indian English in five South African cities in relation to the variable (t). *Language Variation and Change*. *24*(3): 371–395

Mesthrie, R. 2017. Class, gender, and substrate erasure in sociolinguistic change: A sociophonetic study of schwa in deracializing South African English. *Language*. 93(2): 314–346

Mesthrie, R. & Deumert, A. 2009 'Language Variation and Change' in R. Mesthrie, J. Swann, A. Deumert & W. Leap (eds), *Introducing Sociolinguistics* (2nd ed.). Edinburgh: Edinburgh University Press.

Meyerhoff, M. & Niedzielski, N. 2003. The globalization of vernacular variation. *Journal of Sociolinguistics*. 7: 534–555.

Milroy, L. 1980. *Language and Social Networks*. London: Basil Blackwell.

Milroy, L., & Milroy, J. 1992. Social network and social class: Toward an integrated sociolinguistic model. *Language in society*. 21(1): 1-26

Milroy, L. 2007. Off the shelf or over the counter? On the social dynamics of sound changes. In C. Cain and G. Russom (eds.) *Studies in the History of the English Language 3*. Berlin/New York: Mouton de Gruyter. 149–172.

Nerbonne, J. 2010. Measuring the diffusion of linguistic change. *Philosophical Transactions: Biological Sciences*. 365(1559): 3821-3828.

R Core Team 2019. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*. Available from https://www.R-project.org/

Romaine, S. & Lange, D. 1991. The use of *like* as a marker of reported speech and thought: A case of grammaticalization in progress: *American Speech*. 66:227-279

RStudio Team 2019. RStudio: Integrated Development for R. *RStudio, Inc.* Available from http://www.rstudio.com/.

Perse, E. & Powell, R. 1985. Loneliness, parasocial interaction, and local television news viewing. *Human Communication Research*. 12:155–180

Sanchez, T. & Charity, A. 1999. "Use of *be like* and Other Verbs of Quotation in a Predominantly African-American Community." Paper presented at the 28th annual conference on New Ways of Analyzing Variation (NWAV 28), Toronto, Oct. 14–17.

Sankoff, D., Tagliamonte, S. and Smith, E. 2005. Goldvarb X: A variable rule application for Macintosh and Windows [Computer software]. Department of Linguistics, University of Toronto.

Sankoff, D., Tagliamonte, S. and Smith, E. 2012. Goldvarb Lion: A variable rule application for Macintosh [Computer software]. Department of Linguistics, University of Toronto.

Sankoff, G., Thibault, P., Nagy, N., Blondeau, H., Fonollosa, M., & Gagnon, L. 1997. Variation in the use of discourse markers in a language contact situation. *Language Variation and Change*. 9:191–217.

Sayers, D. 2014. The mediated innovation model: A framework for researching media influence in language change. *Journal of Sociolinguistics*. 18(2):185-212

Schiffrin, D. 1981. Tense variation in narrative. *Language*. 57:45-62.

Schourup, L. 1982. "Common Discourse Particles in English Con- versation." Ph.D. diss., Ohio State Univ. *Working Papers in Linguistics* (Dept. of Linguistics, Ohio State Univ.) 28 [1983]. Available at http://linguistics.osu.edu/research/publications/workingpapers/files/osu_wpl_28.pdf.

Sheth, F. A. 2009. *Toward a Political Philosophy of Race* . SUNY Press.

Silverstein, M. 2003. Indexical order and the dialectics of sociolinguistic life. *Language & Communication*. *23*(3): 193–229

Singler, J. 2001. Why you can't do a VARBRUL study of quotatives and what such a study can show us. *University of Pennsylvania Working Papers in Linguistics*. 7:257-278

Strobl, C., Boulesteix, A., Zeileis, A. and Hothorn, T. 2007. Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution. *BMC Bioinformatics*. 8(25)

Strobl, C., Boulesteix, A., Kneib, T., Augustin, T. and Zeileis, A. 2008. Conditional Variable Importance for Random Forests. *BMC Bioinformatics.* 9(307)

Tagliamonte, S. 2014. Situating media influence in sociolinguistic context. *Journal of Sociolinguistics*. 18(2):223-232

Tagliamonte, S. & Hudson, R. 1999. *Be like* et al. beyond American: The quotative system in British and Canadian youth. *Journal of Sociolinguistics*. 3(2):147-172

Tagliamonte, S. & D'Arcy, A. 2004. *He's like, she's like*: The quotative system in Canadian youth. *Journal of Sociolinguistics*. 8(4):493-514

Tagliamonte, S. & D'Arcy, A. 2007. Frequency and variation in the community grammar: Tracking a new change through the generations. *Language Variation and Change*. 19:1-19

Tagliamonte, S. & D'Arcy, A. 2009. Peaks Beyond Phonology: Adolescence, Incrementation, and Language Change. *Language*. 85(1):58-108

Tagliamonte, S. 2005. *So* who? *Like* how? *Just* what? Discourse markers in the conversations of young Canadians. *Journal of Pragmatics*. 47: 1896-1915

Tagliamonte, S., D'Arcy, A. & Louro, C. 2016. Outliers, impact and rationalization in linguistic change. *Language*. 92(4):824-849

Tannen, D. 1986. Introducing constructed dialogue in Greek and American conversational and literary narrative. In *Direct and indirect speech*, edited by Florian Coulmas, 311-324. New York: de Gruyter.

Titilola, O. 2017. *I was like* as a quotative device: Implications for indirect or reported speech in Nigerian English usage. *Asia Pacific Journal of Multidisciplinary Research*. 5(2):94-103

Trudgill, P. 1974. Linguistic change and diffusion: Description and explanation in sociolinguistic dialect geography. *Language in Society*. 3:215-246.

Trudgill, P. 2014. Diffusion, drift and the irrelevance of media influence. *Journal of Sociolinguistics*. 18(2):214-222

Underhill, R. 1988. *Like* is, like, focus. *American Speech*. 63:234-246

Waters, C. & Tagliamonte, S. 2017. Is one innovation enough? Leaders, covariation, and language change. *American Speech*. 92(1):23-49

Wickham, H. 2007. Reshaping Data with the reshape Package. *Journal of Statistical Software*. 21(12): 1-20

Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.