Pragmatic "like" in the South African English of Port Elizabeth

YM Coetzee

orcid.org/0000-0003-3082-1152

Dissertation accepted in fulfilment of the requirements for the degree *Master of Arts in Linguistics and Literary Theory* at the North-West University

Supervisor: Prof I Bekker

Co-supervisor: Prof YV Botha

Graduation: June 2025

DEDICATION

I dedicate this dissertation to my parents whose love and support were indispensable throughout my entire academic years. To my mother who took me to the public library at a young age where I fell in love with books which in turn ignited my passion for language.

ACKNOWLEDGEMENTS

I would like to thank God for giving me the ability, strength and support in order to further my studies as well as for giving me my friends and family who gave me immense support throughout this journey.

I would like to thank the NWU financial support, the School of Language Department and the Department of Arts and Culture for their various bursaries without which pursuing a master's degree would not have been possible for me. I also thank my supervisor Prof. Ian Bekker for funding my trips to the SALALS and SAMWOP conferences of 2024.

I thank the organisers of the SALALS conference for providing me with an amazing first conference experience and I thank the audience members who were willing to listen to my presentation, ask questions and provide feedback. I also thank the organisers of the SAMWOP conference for accepting my abstract and inviting me to present my paper at Wits University.

I would like to thank my two supervisors, Prof Ian Bekker and Dr Lande Botha for providing me with guidance and feedback throughout this entire process.

I thank my friend Celia who could give me advice and encouragement because of her experience of obtaining a master's degree and for her friendship, visits, and after she moved to Cambodia, for her catch-up phone calls. I also thank my friend Cisca for coffee and dinner outings, for providing me with anything I needed help with while writing the dissertation and for her close friendship and especially for being willing to edit my dissertation. In addition, I thank my boyfriend Erich, who has been an immense mental and emotional support system, for helping me feel better when the work seemed formidable. I also thank him for his help with programming in R and the maths side of statistics; without his help I would not have progressed as quickly in those areas.

I thank the Ferdinand Postma Library for surrounding me with books and especially for their linguistics section which I love browsing.

Finally, a huge thanks to the participants who were willing to be interviewed for this study. Without your help this entire study would not have happened.

ABSTRACT

Key terms: Quotative *be like*, discourse-marker *like*, variationist sociolinguistics, language change in progress.

Investigations into quotative *be like* focus on its role in introducing reported speech, where *like* combines with a form of the verb *be*, for example: "I'm **like**, 'Get the hell out of here'" (Tannen, 1986:321). Studies also examine the variable use of *be like* within the English quotative system, comparing it to the use of other quotatives such as *go* and *say* as an example of language change in progress. Investigations into discourse-marker *like* (DM-*like*), for example, "She **like** covered the mirror" (Meehan, 1991:45), in turn, question the view that it is merely a filler, instead suggesting that its use is systematic and constrained by internal factors such as its position in the grammar, where it has three basic positions – clause-initial, clause-medial and clause-final. Both quotative *be like* and DM-*like* have been studied in terms of the influence that social factors such as sex, age and social class have on their frequency (and position) of use. While these issues have been moderately studied in the major English varieties, limited investigation has been conducted for first language (L1) White South African English (WSAE). This M.A. dissertation aimed to contribute to the variationist study of pragmatic *like* within WSAE by conducting a statistical analysis using c-trees in R. The data analysed are transcripts of sociolinguistic interviews from 19 native inhabitants of Port-Elizabeth/Gqeberha, aged 18–48.

The social factors that were expected to influence the use of *be like* and DM-*like* were age, gender and social class, where younger, middle-class female speakers were expected to be the leaders in the use of both types of pragmatic *like*. In terms of linguistic factors, quotative *be like* was expected to be used the most with the present tense, reporting thought (as opposed to speech), as well as with first-person subjects (as opposed to the third person). DM-*like* was expected to be found mostly in clause-medial rather than clause-initial position.

The results for this study indicated that younger speakers are the most frequent users of both quotative *be like* and DM-*like*, which is similar to results found for international varieties such as American, Canadian, Australian and New Zealand English. Middle-class speakers were the most frequent users of both *be like* and DM-*like*. However, DM-*like* was more frequent among upper-middle class men compared to their female counterparts. Contrary to expectation, *be like* was used more with the past tense (vs the present tense) while the most frequent tense reported for *be like* in varieties such as Australian, British and Canadian English were the present tense (which includes the historical present). DM-*like* was more frequent in clause-medial as compared to clause-initial position.

LIST OF ABBREVIATIONS

NWU	North-West University
AAVE	African American Vernacular English
AI	Artificial Intelligence
AmE	American English
AMF	Anglophone Montreal French
AusE	Australian English
BNC	British National Corpus
BrE	British English
BSAE	Black South African English
CAMSNAE	Cambridge Corpus of Spoken North American English
CanE	Canadian English
COGTA	Department of Cooperative Governance and Traditional Affairs
COLT	Bergen Corpus of London Teenage Language
CSAE	Coloured South-African English
C-tree	Conditional Inference Tree
DM	Discourse Marker
EngE	England English
HP	Historical Present
ICE	International Corpus of English
ICE-GB	International Corpus of English, British variety
IMT	Institute of Maritime Technology
IrE	Irish English
ISAE	Indian South African English
LM	Lower-middle class
LSWE	Longman Spoken and Written English Corpus
LWC	Lower-working class
MA	Magister Artium

MFY	Mobilization for Youth Survey
MIT	Massachusetts Institute of Technology
MM	Middle / middle-middle class
MU	The Mobile Unit (Copus)
NE	Natal English
NITCS	Northern Irish Transcribed Corpus of Speech
NMBM	Nelson Mandela Bay Metropolitan Municipality
NP	Noun Phrase
NWAVE	New Ways of Analyzing Variation in English
NWU	North-West University
NYC	New York City
NZE	New Zealand English
OED	Oxford English Dictionary
PNC	Philadelphia Neighborhood Corpus
RP	Received Pronunciation
S.A.	South Africa
SABC	South African Broadcasting Corporation
SAE	South African English
SALALS	Southern African Linguistics and Applied Linguistics Society
SAMWOP	Southern African Microlinguistics Workshop
UM	Upper-middle class
UTF-8	Unicode Transformation Format – 8
UWE	University of the West of England
VOC	Dutch East India Company (*Vereenigde Oostindische Compagnie)*
VP	Verb Phrase
VS	Variationist Sociolinguistics
WESSA	White English-speaking South Africans
WSAE	White South African English
WWI	World War 1

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: CONTEXTUALISATION AND PROBLEM STATEMENT

1.1 Introduction

According to Meehan (1991:37), the word *like* is listed in the OED from the 14th century. D'Arcy (2017:4) states that the verb use of *like* is derived from the Old English word *lician*, which meant "to please": "Ge noldon gode lician". 'You would not please God'" (*Ibid*.:4), but in contemporary use it means *to like* something (*Ibid*.:4). Meehan (1991:39) explains that *like* used for similarity or comparison is an adverbial extension of the Old English adjective *gelic* (meaning "having the form of"; *Ibid*.:39), e.g. "She sings **like** her, laughs **like** her, talks **like** her" (Meehan, 1991:39). D'Arcy (2017:6) states that the comparative use is dated from Early Middle English (circa 1200). Another use listed by Meehan (1991:40) means "approximately", which relates to the sense of "something like", e.g. "I wanted it **like** about this short (*Ibid*.:40) and I wrote it in **like** ten minutes (*Ibid*.:41; listed in Biber *et al*. (1999:557) as a hedge). Meehan (1991:41) states that the approximative use is linked to the "similar to" meaning, while the conjunction has the meaning of "as if", e.g. "**Like** you could afford the house I want" (*Ibid*.:41) and "…it was **like** I was watching someone else do it" (*Ibid*.:41). According to D'Arcy (2017:7), the conjunctive use dates from the 15th century in Early Modern English. Another use listed by Meehan (1991:42) (which according to her, dates from the 19th century), are instances in which *like* means "for example": "**Like** what?" (*Ibid*.:42) or "Do you have **like** a mint or something?" (*Ibid*.:42). Biber *et al*. (1999:557) categorises the latter, along with examples as in: "They'd bring **like** a little flaming fire thing" (*Ibid*.:552) as epistemic stance adverbials or hedges, which are similar to the approximative use. Uses of *like* that involve hedging are described by Meehan (1991:43) as the most recent in the OED. She states that this use has lost a large amount of its semantic meaning and would not change the meaning of the sentence if *like* were removed.

According to Meehan (1991:44), *like* became more pragmatic the more it was generalised to new contexts and became more grammatical rather than lexical. The following examples are uses of *like* as a discourse marker that is entirely pragmatic with little semantic meaning left: "She's **like** really pregnant" (*Ibid*.:44); "I want to **like** minor in communications" (*Ibid*.:44); "**Like** it was 25 years ago when I had him as an instructor" (*Ibid*.:45). Lastly, *like* as a quotative in the construction *be like* is described by Meehan (1991:46) as "more grammaticalized" than the discourse marker in the examples above. It is different because its position in a sentence is always the same – before a quotation of reported speech (Meehan, 1991:48). In Meehan's *(Ibid*.:49) trajectory of *like*, the discourse-marker uses date from 1959 and quotative *be like* is the newest use, dating from the 1960s.

In the literature, there are a few different terms used for pragmatic instances of *like*, such as "focus *like*" (e.g. Meehan, 1991:49) and "discourse marker" (e.g. D'Arcy, 2017:14; Romaine & Lange, 1991:258). There is thus some disagreement about exactly which instances of pragmatic *like* fit into the category of discourse marker (e.g. D'Arcy, 2017:10 vs Andersen, 2001:260). For this study, all instances of pragmatic *like* which are not quotative *be like* are referred to as "discourse-marker *like*" (DM-*like*).

1.1.1 Discourse-marker *Like*

Schourup (1982:28) states that the following are standard dictionary uses of *like*:

(1) A transitive verb: I'd **like** to have my back rubbed. (*Ibid*.:28)

(2) A noun: Everyone has **likes** and **dislikes**. (*Ibid*.:28)

(3) An adjective, meaning "the same" or "likely": Tables of **like** colour; I have prepared dishes that are **like** to please you. (*Ibid*.:28)

(4) A preposition with meanings such as "similar to" and "such as": His typewriter is **like** a small airplane; a subject **like** physics. (*Ibid*.:28)

(5) A conjunction, meaning "as" or "as if": She holds her pencil **like** most people hold their toothbrush. (*Ibid*.:28)

Schourup (1982:28) describes *like* as an adverb in examples where it is used with numerals and other expressions to indicate impreciseness equivalent to an adverb such as *roughly* or *approximately*:

(6) "The actual interest is more **like** 18 percent" (*Ibid*.:28).

However, Schourup (1982:30) also states that in some examples, *like* cannot be replaced by an adverb such as *approximately*. For example, its meaning in (7) is not imprecise:

(7) Sue: You know that um – they've been livin(g) in this big three-story house with basements – **like** four floors y'know – gigantic house on Summit. (*Ibid*.:30)

Schourup (1982:30) states that in examples such as (7), the speaker is not merely saying that the house has somewhere between three and five floors, but that from the context, it is clear that the speaker is "concerned with the largeness of this house" (because of words such as *big* and *gigantic*). Thus, Schourup (1982:30) states that in these kinds of examples, *like* carries a special meaning. In addition, Schourup (1982:31) states that in instances where *like* co-occurs with adverbs such as *approximately* or *about*, it does not have a redundancy effect and is therefore evidence that it is not equivalent to those adverbs:

(8) …the little boy is **like** approx- about ten years old. (Schourup, 1982:31)

It is because of these types of examples that Andersen (1996:41) states that *like* behaves "as a borderline case between discourse marker and adverbial". Thus, even though *like* in cases with

numeral and quantifiable expressions is used similarly to adverbs such as *roughly*, *approximately*, or *about*, it still has a pragmatic effect and is therefore included in the count and analysis of the discourse marker in this study.

The purely discourse marker uses of *like* are instances in which it can be interpreted as "for example" (Schourup, 1982:36):

(9) … a lotta people **like** in business or other – uh things like that, they get internships y'know for the summers. (*Ibid*.:36)

In other instances, Underhill (1988:238) describes it as a marker of focus, where it puts emphasis on "the most significant new information in a sentence--often, the point of the sentence":

(10) Then, all of a sudden, he **like** stopped producing. (*Ibid*.:238)

(11) She had a research proposal **like** submitted to SIH. (*Ibid*.:238)

It is also used as a hedge, of which its uses with numerical/quantificational expressions (as in example ((6)) is included. In these cases, it often means something similar to "sort of" (Underhill, 1988:240) and expresses a degree of uncertainty. Thus, it is described by Biber *et al.* (1999:557) as an epistemic stance adverb, often used to modify adjectives and noun phrases and can express imprecision of word choice:

(12) We always dragged it out, it's **like** gross and decrepit. (Biber et al., 1999:859)

There are also instances of *like* in which it is used as a hesitation marker, where, according to Schourup (1982:40) it usually occurs "preclausally but after prefatory material, before filled and unfilled pauses, and before restarts":

(13) Y'know an(d) – **like** he buzzed her. (*Ibid*.:39)

(14) It has to do with **like** uh things. (*Ibid*.:42)

(15) it w- **like** bled a little bit. (*Ibid*.:42)

One other use of *like* is when it occurs as a fixed expression, described by Andersen (1996:45) as consisting of *it* as a dummy pronoun collocated with the verb *be*, as in:

(16) It's **like**, what's happened, is Jim still like that? (*Ibid*.:42)

All the uses of *like* demonstrated by examples (6)–(15) were included in the analysis of the discourse marker because of their shared pragmatic effect. The traditional and non-pragmatic uses of *like* such as in examples (1)–(5) were excluded.

There are three structural positions for DM-*like* that have been investigated in the literature. Clause-initial position is defined by D'Arcy (2005:81-82) as *like* appearing before matrix or main clauses, as in:

(17) She likes to travel round. **Like**, she came back here, she was working at a hotel in Jesmond yeah, and then she decided to go to Australia. (D'Arcy 2017:14)

(18) Nowadays there's all this technical stuff and **like** we never had computers. (D'Arcy 2017:14)

and where *like* appears before subordinate clauses, as in:

(19) I just remember, **like** there were tons and tons of papers and garbage all over the floor. (D'Arcy 2005:82)

Clause-medial position is defined as anywhere inside a clause (D'Arcy, 2005:116–171), for example:

(20) Before determiner phrases: I remember there being **like** a solar eclipse. (*Ibid*.:115)

(21) Before adjective phrases: She's **like** dumb or something. (*Ibid*.150)

(22) Before noun phrases: I used that idea of Lucia's **like** struggle and everything. (*Ibid*.:133)

(23) Before verb phrases: They **like** wanna see like how we talk and all that. (*Ibid*.:181)

Clause-final position is described by Schourup (1982:35) as appearing after questions, such as:

(24) Last Bonus Weekend **like**? (*Ibid*.:32)

Schourup (1982:35) states that in such instances, the speaker is seeking clarification about something said previously. It also appears after statements, such as:

(25) If your honour were among us, there might be more discipline **like**. (*Ibid*.:36)

Schourup (1982:36) explains that in such instances, the meaning attributed to *like* is usually "as it were", "so to speak", "in short" or "in a way". D'Arcy (2017:12) states that clause-final *like* is a "traditional and long-standing feature of dialects of English" and that it is not widely spread across varieties of English. For instance, she states that it is declining with North American speakers where it is only found among older speakers and it is most frequently found in Ireland and Northern Ireland (D'Arcy, 2017:13; see also Corrigan, 2015:57; Schweinberger, 2013:24; Schweinberger, 2015:248).

1.1.2 Quotative *Be Like*

Tannen (1986:321) defines quotative *be like* as an introducer of "constructed dialogue". Examples are:

(26) It **was like** [spoken with foreign accent] "Would you stay. Why you why you leaving Santorini?" (Tannen, 1986:321)

(27) He **was like** [accent] "I give you a week of my life. Will you stay." (Tannen, 1986:321)

I **was like** -"Nice sister, right?" (Tannen, 1986:321)

Vandelanotte and Davidse (2009:779) categorise *be like* as one of a group of "innovative intransitive quotatives", because such verbs are formally intransitive and they "lack any obvious reporting sense". Other quotative verbs in this group include *go* and *be all*:

(28) And she **was all**, ''Um, yah, I know, but there's going to be wine there." (Vandelanotte & Davidse, 2009:778)

(29) 'n I look at it 'n I **go**, "Gee I don't know what the sign is either". (Tannen, 1986:319)

Vandelanotte and Davidse (2009:779) state that *go* is "the older innovation. It seems to have emerged in the US about 40 years ago, where it was replaced with *be like*, at least among younger speakers, fairly soon". This observation is supported by the studies discussed below.

According to Blyth *et al.* (1990:215), *be like* is versatile because it can introduce either direct speech or thought, whereas most quotatives introduce only direct speech (such as *say*) or only thought (such as *think*). It can also enable someone to "express an attitude or reaction" (*Ibid*.:215). According to Labov (2018:2), there is a consensus that *be like* originated in the US. It is generally linked with teenage girls from California known as "Valley Girls" (Buchstaller & D'Arcy, 2009:295; Labov, 2018:5), who are described as "suburban and middle-class" (OED, 2024). Tagliamonte and Hudson (1999:149) state that until the early 1990s, *be like* seemed to only occur in the United States and early studies of the quotative were focused on American English (AmE).

Both quotative *be like* and discourse-marker *like* have been studied in terms of the influence that social factors such as sex, age and social class have on their frequency (and position) of use. While these issues have been moderately studied in the major English varieties, limited investigation has been conducted for first language (L1), White South African English (WSAE) (e.g. Lochner, 2019; Horn, 2022).

1.2 Research Objectives

The general research objectives of this M.A. dissertation is to contribute to the variationist study of pragmatic *like* within WSAE by investigating the social and linguistic (internal) factors that have an impact on its use in the WSAE speech community in Port Elizabeth. More broadly, this study adds to the investigation of the spread of *be like* and DM-*like* to global English varieties as well as their development in WSAE.

1.3 Contribution to the Study Field

This research study contributes to the growing spoken language corpus data for WSAE, to the sociolinguistic study of first language (L1), White South African English and how it compares to the other L1 English varieties of the world, and to the global study of pragmatic *like* in general.

1.4 Research Questions and Hypotheses

The research questions and hypotheses for this study are the following:

1.4.1 Research questions

1. How does pragmatic *like* (quotative *be like* and discourse-marker *like)* usage in White South African English (WSAE) in the Port Elizabeth region pattern according to external and internal factors?

2. Once revealed, do the internal and external factors of pragmatic *like* in WSAE in Port Elizabeth English suggest anything about its current stage of development as compared to international English varieties?

3. Are there any similarities or differences in the social constraints on *be like* and DM-*like*?

1.4.2 Basic hypotheses

1. Regarding age, the expectation is that the quotative in the WSAE of Port Elizabeth will be found most frequently in the speech of young people in the age range of 18 to 39 and will be on the decline in the speech of older speakers. The same is expected for DM-*like*.

2. In the case of gender distribution, judging by many of the previous studies, it is expected that *be like* and *like* will be used more by women than by men.

3. For social class, some studies have found that *be like* tends to be used more by the middle class. However, previous studies of DM-*like* have indicated that this form tends to be used more by the lower middle classes in some cases, while it is used more by the upper middle and middle classes in other cases. For this study, DM-*like* is expected to be used more by the middle class than the lower middle and upper middle classes.

4. Regarding grammatical factors of quotative *be like*, WSAE might be in an early stage of development since previous studies have indicated that English varieties further away from America tend to lag behind. In consideration of this, the hypothesis for internal factors is that it would mainly be used with first-person subjects, internal thought reporting and the present tense: these were the features reported to correlate with *be like* in the early AmE

studies. These internal factors are considered to be signs of an early stage of development for *be like*. In terms of the distribution of *be like* as compared with other quotatives, it is expected that it has not yet overtaken other quotatives as the dominant form introducing constructed dialogue.

5. The clause position of DM-*like* is expected to be the following, based on previous studies. First, that clause-medial *like* is the most frequent structural position vs clause-initial *like*. Since clause-final *like* is strongly associated with Irish English and is considered dialectical, it is expected to be non-existent or rare in WSAE.

1.4.3 Operational hypotheses

1. H1 If quotative choice influences the use of *be like*, then it will be the most frequent quotative compared to the other quotatives in use.
   H0 If quotative choice influences the use of *be like*, then it will not be most frequent quotative compared to the other quotatives in use.

2. H1 If the social factors of gender, social class and age influence the use of *be like*, then women will use it more than men, the higher social classes will use it more than the lower social classes and the younger age group will use it more than the older age group.
   H0 If the social factors of gender, social class and age influence the use of *be like*, then women will not use it more than men, the higher social classes will not use it more than the lower social classes and the younger age group will not use it more than the older age group.

3. H1 If content of quote, grammatical person and tense influence the use of *be like*, then it will be used more with internal monologue than direct speech, it will be used more with third-person than first-person subjects, and it will be used more with the past than the present tense.
   H0 If content of quote, grammatical person and tense influence the use of *be like*, then it will not be used more with internal monologue than direct speech, it will not be used more with third-person than first-person subjects, and it will not be used more with the past than the present tense.

4. H1 If the social factors of gender, social class and age influence the use of DM-*like*, then women will use it more than men, the higher social classes will use it more than the lower social classes and the younger age group will use it more than the older age group.
   H0 If the social factors of gender, social class and age influence the use of DM-*like*, then women will not use it more than men, the higher social classes will not use it more than the

lower social classes and the younger age group will not use it more than the older age group.

5. H1 If clause position influences the use of DM-*like*, then clause-medial position will be the most frequent structural position for *like*.
   H0 If clause position influences the use of DM-*like*, then clause-medial position will not be the most frequent structural position for *like*.

1.5 Research Aims

1. This study aims to discover which social and internal factors of *be like* and DM-*like* influence its use in the Port-Elizabeth WSAE speech community.

2. It also aims to discover any social links between *be like* and *like* in the L1 English of Port-Elizabeth.

1.6 Problem Statement

The problem to be addressed throughout this study is to discover which social and internal factors influence the use of quotative *be like* and DM-*like* in the WSAE speech community of Port-Elizabeth and how these findings compare to those of international English varieties such as American, Canadian and New Zealand English.

CHAPTER 2: THEORY

2.1 Introduction

This chapter provides the theoretical background to the study. It first introduces the variationist sociolinguistic (VS) approach to the analysis of natural language as set out in Weinrich *et al.* (1968) and the theoretical assumptions that form the basis of this approach. It then gives a brief overview of the important studies that established the methods of VS, where these methods were first applied in the field of phonology as developed by William Labov and later adapted for other linguistic fields such as the analysis of syntax, especially in the work of Gillian Sankoff. The key concepts relevant to this study are also introduced and discussed in this chapter. Finally, it gives an overview of previous studies conducted for quotative *be like* and discourse-marker *like* in L1 English varieties and their important findings.

2.2 Variationist Sociolinguistics

Sociolinguistics is the study of language in its social context. Tagliamonte (2012:1) describes it as investigating "the interaction between language, culture, and society". This moves away from one of the main areas of linguistic theory, namely generative grammar, as originally formulated by Noam Chomsky, and its predecessors. Generative grammar focuses on the "ideal speaker-listener, in a completely homogeneous speech-community…" (Chomsky, 1964:1). It is primarily concerned with how the grammar of a language is processed in the brain, using the principles and parameters of Universal Grammar to illustrate such processes. Chomsky (1964:2) divided language into two elements, "competence" (the knowledge that a speaker has of their native language) and "performance" (the use of that language in specific situations).

Sociolinguistics takes up the formal study of performance and investigates the language internal and external factors that interact with language use in context. However, it moves away from the division of language between *langue* and *parole* (Saussure, 1916:xxxvii) or *competence* and *performance*, established by Saussure and reinforced by Chomsky (Chambers, 2002:27, 35). According to Chambers (2002:35–36), studying natural language data in context does not fit in with theories that divide language into such distinct categories. Sociolinguistics rather sees *langue* and *parole* as representative of poles on a continuum rather than separate categories that can be studied in isolation from each other (Chambers, 2002:36).

According to Mesthrie (2009:3), scholars such as Franz Boas, Leonard Bloomfield and Edward Sapir introduced an anthropological perspective into the study of language through their knowledge of cultural and social patterns and, in effect, established along with others the more general field of sociolinguistics. Variationist Sociolinguistics (VS) developed into a subfield distinct

from the rest of Sociolinguistics with the innovative research of William Labov (discussed in more detail in sections 2.2.2 and 2.2.3. Tagliamonte (2012:2), explains that this subfield attempts to discover how language changes through time by systematically describing the inherent variability of language. According to Chambers (2002:12), Sapir, for example, acknowledged that language is variable throughout his research, but until the start of VS in the 1960s, it was never formally studied. The main approach to linguistic study before the advent of VS was to "regularize" linguistic data for analysis and "to eliminate real-world variability" (Chambers, 2002:12). Linguistic study was approached in this way because it was believed that including such real-world variability would make the data analysis unmanageable. Labov's early studies, such as his New York survey, proved that such data could be manageable as well as scientific (Chambers, 2002:15–16). The assumption in VS is that variability in language is governed by certain rules and does not consist of random occurrences. Because of this focus, the main object of study in VS is the linguistic variable.

2.2.1 Assumptions in VS

Variationist sociolinguistics moves away from approaches to language study which require the object of study to be "homogeneous" (Weinrich *et al*., 1968:121), such as structuralist linguistics (developed by Saussure) and generative grammar: both in the "neogrammarian" tradition (Weinrich *et al*., 1968:125). Weinrich *et al*. (1968:125) state that such theories concern themselves with the "linguistic competence" of the individual (or the language of the idiolect, *Ibid*.:129) and only consider the speech community to the extent that it is homogenous and to the extent that the individual as speaker is "a perfect representative of it". Such theories do not consider the "observed diversity of speech behaviour" and Weinrich *et al*. (1968:125) state that Chomsky considers such diversity to be irrelevant to language theory. Where VS differ is that it makes linguistic diversity the object of study and assumes that

> continuous variation exists within each dialect as a structural element, correlated with some other linguistic or non-linguistic factor, and that the steady movement of tokens from one categorical class to another is part of the underlying structure (Labov, 1966, cited in Weinrich *et al*., 1968:150).

Another assumption is that "language must be structured to function efficiently" and the main concern that arises from this is how language functions as the structure changes (Weinrich *et al.*, 1968:150). Thus, Weinrich *et al.* (1968:131) propose to describe "orderly differentiation within language". The main viewpoint of VS is therefore that language is a heterogeneous system in which variability has a systematic structure (Weinrich *et al.*, 1968:162). In addition, it is not possible to "understand the development of a language change apart from the social life of the community in which it occurs" (Labov, 1972:3). One of the main facts about the study of variation is that variation in language is most easily observed in everyday speech, or the vernacular

(Tagliamonte, 2012:2). Labov (1972:208) defines the vernacular as “the style in which the minimum attention is given to the monitoring of speech”. Tagliamonte (2012:2) states that subsequent discussions verified that the vernacular is the ideal style for sociolinguistics to investigate.

The main object through which variability is studied is the linguistic variable. Weinrich *et al.* (1968:167) define the linguistic variable as “a variable element within the system controlled by a single rule”. Tagliamonte (2012:2) describes it as the “alternation of forms in language”; this means two or more linguistic forms that either have a similar meaning or similar function within the same grammatical system. The criteria for linguistic variants (the alternative forms for one meaning/function) are that the choice of one variant over another must vary systematically; variants must have “a structurally defined relationship in the grammar” (Tagliamonte, 2012:5) and “they must also co-vary, correlating with patterns of social and/or linguistic phenomena” (*Ibid*.:5). An example of linguistic variants that have a similar function is the choice between the relative pronouns *that* or *who*. For instance, speakers refer to humans with either *that*: “The people **that** did it were brainwashed” (Tagliamonte, 2012:3), or *who*: “I’m only exposed to the people **who** speak the same way that I do” (*Ibid*.:3). Tagliamonte (2012:5) states that linguistic variables can be found at all levels of grammar, which means that a variable can be “a phoneme, lexical item, structural category, natural class of units, syntactic relationship” or “the permutation or placement of items” (*Ibid*.:5). According to Tagliamonte (2012:2,120), VS has two components, a “descriptive-interpretive” component that requires “detailed, critical observation of variation and change” (*Ibid*.:2) and a statistical component, of which the quantitative methods were specifically developed for VS research.

Weinrich *et al.* (1968:100) assert that the idea of describing language as a “homogenous object” (*Ibid*.:100), as found in generative grammar, is unrealistic since it ignores the inherent heterogeneity in language and thus cannot account for how a language changes from one state to another. According to Tagliamonte (2012:71), VS studies the linguistic variable to answer the five problems of linguistic change as put forth in Weinrich *et al.* (1968:101):

1. What are the constraints on language change as it moves from one state to another? (the “problem of constraints” (*Ibid*.:101)).

2. How does language change? (the “transition problem” (*Ibid*.:101)).

3. How are observed language changes embedded in the linguistic and social systems of those changes? (the “embedding problem” (*Ibid*.:101)).

4. How can the effect of observed language changes be evaluated, in the sense of how they impact the linguistic structure, the efficiency of communication (i.e. their “functional load”

(*Ibid*.:101)) and the wide variety of other factors involved in speaking? ("the evaluation problem"). According to Tagliamonte (2012:71), this evaluation includes how members of a speech community evaluate a given change and the effect of their evaluation on the change.

5. "What factors can account for the actuation of changes?" (*Ibid*.:102). In other words, why did a change occur in a structural feature in a particular language at a certain time and place, but not in another language with the same structural feature or in the same language at a different time? (the "actuation problem" (*Ibid*.:102)).

The current study is concerned with the embedding problem: how an observed language change is embedded in the linguistic and social systems of those changes. The following section reviews some of the important studies in VS and the conclusions drawn from them. According to Ball (2010:2), the two studies that developed and refined the methods of VS were Labov's study of Martha's Vinyard (Labov, 1963) and his study of New York City (Labov, 1966a). The review of the seminal VS studies below will commence with these two.

2.2.2 Martha's Vinyard (Labov, 1963)

According to Pope *et al.* (2007:615), Labov's (1963) study of language variation and change in progress on the island of Martha's Vinyard was revolutionary and established the fundamentals of VS. They also state that this study proved that the link between social and linguistic variables could be studied systematically. They also introduced the method of apparent-time. With this method, "diachronic change in progress" (Pope *et al.*, 2007:615) could be extrapolated via synchronic patterns. Labov (1963:275) states that it was a preliminary investigation into the social patterns of linguistic change. Labov (1963:275) described Martha's Vinyard as a "self-contained unit", relatively far from the mainland and divided into two sections; "up-island" and "down-island" (*Ibid*.:276), where "up-island" represents town life where most of the community live. "Down-island", on the other hand, is mainly rural, consisting of a few villages and farms. Labov (1963:278) describes the area as an "important relic area of American English: an island of *r*-pronouncers in a sea of *r*-lessness" and notes that it maintained numerous archaic linguistic characteristics originating from New England before 1800, of which one feature is the pronunciation of the consonant /r/ in speech (or rhotic /r/).

An important concept springing from this study is the method of identifying the linguistic variable. Labov (1963:279) explains that the linguistic variable must be frequent, something that "occurs so often in the course of undirected natural conversation that its behavior can be charted from unstructured contexts and brief interviews". The second requirement is that it must be structured, meaning that it should be integrated into "a larger system of functioning units". The third requirement is that it must be socially stratified to the point that initial investigations should

suggest an “asymmetrical distribution over a wide range of age levels or other ordered strata of society”. In Labov’s study of Martha’s Vinyard, he identified two such variables, the first was pronunciation of the phoneme /r/. Labov (1966a:33) describes the (r) [1]variable as “the presence or absence of final and pre-consonantal /r/ in words such as *car* or *card*, *bare* or *bared*”, also known as “rhotic /r/” or “non-prevocalic /r/” (Trudgill, 2003:112). The second variant of (r) is the absence of *r*-pronunciation, also known as “non-rhotic” (Trudgill, 2003:65) pronunciation. Thus, the variants of the variable (r) are examples such as the word *star* (Trudgill, 2003:112), where /r/ is heard vs a newer pronunciation, *stah* (*Ibid*.:112) where, according to Trudgill (2003:112), *r*-pronunciation has been lost.

Labov (1963:280) describes rhotic *r* as following a social pattern contrary to the norm in Martha’s Vinyard, where *r*-pronunciation (retroflexion) was increasing in some areas and decreasing in others. The other variable involved changes in the pronunciation of the diphthongs /ai/ (as in *right* and *pride* (*Ibid*.:281)) and /au/ (as in *house* and *out* (Ibid.:281)), where, for example, the standard sound for /ai/ - [aɪ] is changing to [əɪ] (the vowels are becoming centralised (*Ibid*.:280)). Labov (1963:280) comments that speakers are not aware of these vowel changes, which is another necessary characteristic of a linguistic variable.

An important feature of this study regarding the collection of natural language is that Labov (1963:283) devised an interview schedule to capture as many instances of the relevant diphthongs as possible. The structure of the Labovian interview schedule consists of two parts: a list of words to capture a careful reading style (formal speech) and a list of discussion questions formulated with the aim of eliciting the variable in question in casual speech. The variants in formal vs casual speech would then be compared to each other. This interview schedule became a standard for linguistic data collection in VS and is discussed in more detail in Chapter 3, section 3.6.

The distribution of vowel centralisation was investigated across age and time (speakers were recorded over three different periods). One of the main questions relevant for many studies of language change in progress is whether a difference in the frequency of the linguistic variable over successive ages is an example of a linguistic change, or a normal change in speaking patterns usually related to age (Labov, 1963:291). For example, vowel centralisation in Martha’s Vinyard indicated a “regular increase in successive age levels, reaching a peak in the 31–45” (*Ibid*.:291) age group. Labov (1963:293) states that in the case of the centralisation of /au/, the evidence confirms a sound change in progress since old recordings from 1933 showed no instances of such centralisation and thus the feature did not exist at that time. However, in Labov’s recordings from 1961–1962 (*Ibid.*:284), vowel centralisation showed a regular increase in

[1] Variables are represented between brackets while phonemes are represented between slashes.

frequency from older to younger speakers (except for the oldest and youngest age group which produced minimal frequencies) (*Ibid*.:291). Labov (1963:294) concluded that no change correlated with age could explain the results, which is evidence for a sound change in progress. However, he states that the minimum frequencies in the youngest and oldest age groups suggests that the effect of age cannot be discounted entirely and might be a secondary influence (*Ibid*.:294).

The study of Martha's Vinyard proved that, by correlating a linguistic variable to social factors, it was possible to directly observe historical processes such as a sound change (previously thought to be impossible (Labov, 1963:291, n.35, 293)). Two key concepts in VS investigation were touched upon in this review: age-grading and the apparent-time hypothesis. These concepts are briefly described below.

An apparent-time study is the method of observing the linguistic behaviour of different age groups at the same time and making inferences about the patterns that can be discerned from the data. Unlike real-time studies in which a participant needs to be observed through several months or years, it saves time and resources (Chambers, 2002:212). Chambers (*Ibid*.:212) notes that conjectures made from apparent-time studies are dependable in general, but such conjectures hinge on the hypothesis that the language usage of one age group will not change much as they advance in maturity. One of the obstacles to apparent-time studies is age-grading, defined as the change in frequency of a linguistic variable at a later age of a person's life. Chambers states that age-grading is rare and thus does not negate the hypothesis. However, it does add a precaution to it (Chambers, 2002:213; Labov, 1966a:199).

2.2.3 The social stratification of English in New York (Labov, 1966)

Labov's study of Martha's Vinyard was the precursor to a larger study also conducted by him in the Lower East Side of New York City (Labov, 1966a:4). Chambers (2002:16) states that while Labov's NYC study was not the first sociolinguistic investigation, it greatly influenced the field. Labov (1963:308) states that the VS methods for Martha's Vinyard (discussed in section 2.2.2) were refined for his more complex NYC project. Labov's (1966a:4) aim was to investigate the vowel system of NYC and specifically the structure of the sound system. According to Labov (1966a:5), this study shifted linguistics away from mainly focusing on the inconsistent speech of the idiolect or individual and viewed the description of the speech community as more important. Thus, Labov (1966a:5) states that "the language of individuals cannot be understood without knowledge of the community of which they are members". This study developed the concept of the linguistic variable further and added the element of codifying and measuring it "on a linear scale" (i.e. statistically). Labov (1966a:30) set out to discover if linguistic variables such as (r)

were the result of free variation[2] (which was the explanation for this variable in previous studies of NYC (Labov, 1966a:24)). According to Labov (1966a:31), rhotic /r/ is a prestige variant. Overt prestige variants are defined as variants that are considered by the public as "correct" and attribute high social status to their users (Trudgill, 2003:30). Labov (1966a:31) states that not much was known about rhotic /r/ other than the fact that it was preferred among college students and radio announcers. Thus, it was necessary to discover how this variant was distributed among other middle-class groups as well as among the working classes.

In the NYC study, Labov (1966a:135) devised a social class index for grouping speakers into five finely divided social classes: Upper class, upper middle class, lower middle class, working class, and lower working class. The index ranked speakers according to occupation, education and income (Labov, 1966a:135–138). Chambers (2002:22) states that the social class index for (r) was an important method which proved that the existence of linguistic variables was not a result of free variation but consisted of a structured pattern.

2.2.4 The social differentiation of English in Norwich (Trudgill, 1974)

Trudgill's (1974:6) study investigated speech patterns in the urban areas of Norwich, England. An important development from this study was that he expanded the social class index to six components: occupation, father's occupation, income, education, locality of residence and housing type[3] (Chambers, 2002:50). Trudgill (1974:93) not only highlighted morpho-phonological variables (such as the variable (-ing)) in his study, but also included a brief investigation of a non-phonological feature, the present tense marker *-s*, as in *he loves* (*Ibid*.:55). He successfully illustrated how non-phonological linguistic variables could also be correlated to social factors. Trudgill (1974:55) explains that in the East Anglia dialect, it is usual for forms with the present tense marker to vary with forms such as: *he love* or *she have* (also used in the present tense). Trudgill (1974:56) hypothesised that this variable would show a direct relationship with social class. His results indicated that the variants correlated with social class: the frequency of marker-less form rose as one moved down from the highest to the lowest classes (e.g. lower working-class speakers obtained an almost 100% frequency, while middle-class (MM) speakers do not use any marker-less forms (*Ibid*.:61). For Trudgill (1974:61), this result illustrated that the greatest class division in Britain is between the working vs middle classes and that variable (-s) is a "significant marker of social class" (e.g. the marker-less form is a stigmatised variant in the speech community).

[2] Trudgill (2003:52) defines "free variation" as variation in language that is not dependent on any constraints (social or internal. He states that most linguists think that this type of variation is rare.
[3] The current study follows a version of this method for indexing social class and is discussed in more detail in Chapter 3, section 3.4.

The (-ing) variable is frequently cited as an example of a stable sociolinguistic variable (see for example Labov, 1972:241; Chambers, 2002:121; Coates, 2004:49). Chambers (2002:121) defines a stable variable as:

> Those [variables] that are well-established as indicators in a community and are not undergoing change. In the prototypical pattern for a s*table* variable, each age cohort of the same class, sex, ethnic background and other social characteristics, will be similar to older and younger groups in the use of the variant and the amount of style-shifting.

Trudgill (1974:84) describes the variable (-ing) as relating to the suffix in words such as *walking* or *running* (*Ibid*.:84). Chambers (2002:122) explains that this variable is pronounced with a velar nasal consonant (a nasal *ng* sound – [ŋ]) in careful style and in the prestigious Received Pronunciation (RP) (Trudgill, 1974:84), whereas in casual style, the non-RP variant [n] (as in *walkin'* and *runnin'*) (Chambers, 2002:122) is found all over Norwich (Trudgill, 1942:84). Trudgill's study investigated the frequencies of [ŋ] vs [n] in the Norwich speech community. The first finding of the (-ing) variable was that, similar to (-s), it differentiated the working classes sharply from the middle classes: the working classes had a much lower frequency of the RP [ŋ] variant compared to the middle classes, even as frequencies rose for all classes from casual to careful speech styles (Trudgill, 1974:92). The second finding of this variable relates to sex differentiation. Citing a previous study (Fisher, 1958; in Trudgill, 1974:93), Trudgill notes that the [n] variant was more frequent with male rather than female speakers and states that this is also true in the Norwich speech community. He concludes that "a high (-ing) index is typical of male speakers as well as of WC [working-class] speakers" (*Ibid*.:93). Chambers (2002:123) calls this a "sex-graded" variable and it is thus comparable to age-grading, indicating that no change in progress is taking place.

2.2.5 Belfast: Change and variation in an urban vernacular (Milroy & Milroy, 1987)

Milroy and Milroy (1978, in Milroy, 1987:225) studied the vernacular speech of three working-class communities in Belfast, Northern Ireland: Clonard, Ballymacarrett, and the Hammer (an area in Shankill). Milroy (1987:225) states that Belfast working-class society is almost completely segregated with mostly Protestants living in Ballymacarett and the Hammar, while Clonard is 98% Catholic. She also states that these areas have a low status and suffer from "social malaise": unemployment, sickness, juvenile crime, illegitimacy and premature death due to sickness (*Ibid*.:72). Because of tensions and violence between the Catholic and Protestant communities in Belfast, Milroy (1987:43) states that there were certain requirements for gaining access to the speech community that needed to be met. One of these requirements was that the fieldworker had to be a woman because locals were suspicious of strange men and a male fieldworker would be in danger, while a woman was at less risk of being assaulted. Milroy (1987:44) states that the

only way she could gain entry into the community and be allowed to conduct extensive tape recordings was to enter as a “friend of a friend”. This method is elaborated on Chapter 3, section 3.2.

Two of the phonological variants Milroy (1987:118) describes from the Belfast study is an [ʉ] sound, such as in the phrase *good food* (gʉd fʉd) (the “vernacular norm” (*Ibid*.:114) of Belfast, standing in for the prestige RP norm [ʊ] as in [gʊd]), which alternates with the unrounded [ʌ] sound in words such as *foot, took* and *look*. Chambers (2002:135) elucidates that the [ʌ] variant is a non-standard form that is extremely stigmatised to the extent that speakers are embarrassed when they are made aware of their usage of it and the results of the study showed that it is a “vernacular prestige” variant among males (Milroy, 1978:25–26; cited in Chambers, 2002:135).

2.2.6 Above and beyond phonology (Sankoff, 1973)

Rickford (1975:98) states that even though VS was a revolutionary departure from studying language as a homogeneous entity, it had weaknesses such as the fact that the methods of VS focused on morpho-phonemic and phonological variables to the extent of excluding every other area of linguistics such as syntax and semantics. He further states that it was in response to this problem that Sankoff (1973) presented her paper entitled “Above and beyond phonology in variable rules” at the first meeting of *New Ways of Analyzing Variation in English* (NWAVE).

In her paper (1973:45), Sankoff set out to show that “variability occurs and can be dealt with at levels of grammar above (or beyond) the phonological”. One of the studies that Sankoff (1973:55) uses to illustrate this point is Laberge’s (1972; cited in Sankoff, 1973:55) study on pronoun semantics in Montreal French. According to Sankoff (1973:55), Laberge discovered that there existed variability in the way that French speakers used pronouns to represent indefiniteness. The prototypical indefinite pronoun in French is *on* and it “contains no features indicating inclusion or exclusion of the speaker […] or hearer, nor the singular/plural distinction” (*Ibid*.:55), for example:

> Tu as beau parler de l’éléphant, du serpent, mais si **on** peut pas le décrire, hein ?
> [It’s all very well to talk about elephants, about snakes, but if you can’t describe them, eh?]
> (Sankoff, 1973:55)

Additional examples illustrate the use of other pronouns in French used with an indefinite meaning, such as *vous* (formal you), *je* (I), *tu* (informal you), or *une personne* (a person):

> Quand **une personne** élevait une famille à ce temps-là, **vous** étiez pas capable d’avoir de luxe. [When a person was bringing up a family at that time, you couldn’t afford luxuries]
> (Sankoff, 1973:55)

For the above example, Sankoff (1973:56) explains that:

> the speaker does not imply that the 22-year-old addressee has herself raised a family during the depression; this is the indefinite *vous* which is distinguished clearly from personal *vous* in its underlying representation.

Sankoff (1973:56) explains that Laberge's study found that older speakers (between the ages of 82–109) used *vous* as a politeness form when addressing young interviewers but used *tu* variably with *on* to express indefiniteness. Distributional results indicated that there was a difference between how older and younger speakers use pronouns to express indefiniteness: speakers over 50 were found to use *on* frequently as an indefinite marker, alternating with either *tu* or *vous*. However, speakers under 25 used *on* infrequently as an indefinite marker, and "more than half of them use over 80 per cent *tu*" (*Ibid*.:56). It was also the case that most younger speakers used *on* "almost categorically" (*Ibid*.:56) in the same sense as *nous* ("we"), as the subject of a verb. Sankoff (1973:58) states that "this example shows a dramatic and rapid usage change" and that the alternation between the pronouns used to mark indefiniteness showed "increasing input probabilities for younger speakers" (*Ibid*.:58). Sankoff (1973:58) explicates that Laberge's study indicated that the use of *on* as an indefinite marker in Montreal French has become archaic and "gives every sign of disappearing" (*Ibid*.:58). Sankoff (1973:58) concludes that:

> The extension of probabilistic considerations from phonology to syntax is not a conceptually difficult jump. Whenever there are options open to a speaker, we can infer from his or her behavior an underlying set of probabilities [and] underlying probabilities are consistently and systematically patterned according to internal (linguistic) and external (social and stylistic) constraints. There is no reason not to expect similar patterns elsewhere in grammar, aside from the phonological rules and syntactic transformations we have been discussing.

In an example where syntactic variables were studied using VS methods[4], Sankoff and Laberge (1972; in Sankoff, 1973:45) investigated the future marker *bai* in the language of Tok Pisin in New Guinea. In this study, they specifically investigated the variable syntactic placement of the marker before or after a subject noun phrase (NP) (Sankoff, 1973:46). According to Sankoff (*Ibid*.:46), this variation was new because about 50 to 60 years previously, fluent speakers in casual speech placed the marker only (categorically) before the subject NP. Sankoff and Laberge analysed this syntactic variable by counting and comparing the frequencies of the two syntactic positions (before an NP and before a verb phrase (VP)) in first-, second-, and third-person singular, and first-, second-, and third-person plural, other NPs, and the zero subject. This method enabled the researchers to discover that there were grammatical constraints that increased the probability of the marker being used in either the one or the other position. Sankoff (1973:48) states that "[the marker *bai*] is found to be quite stable in some syntactic environments, operating always (in the case of subject NP's [*sic*] containing embeddings) or almost never (in the case of most NP's [*sic*]

[4] An example illustrating how internal (linguistic) factors are correlated to the variable.

consisting of a single pronoun other than *em*)". However, Sankoff (1973:47) states that apart from the grammatical constraints, no social or contextual differences were found. This and other studies, such as Sankoff and others' investigations of variation in Montreal French (e.g. Sankoff *et al.*, 1971; Cedergren & Laberge, 1972; in Sankoff, 1973:48; Sankoff & Thibault, 1980; in Chambers, 2002:189) confirmed that variationist methods can be used to reliably discover the systematic nature of variables beyond phonetics and morpho-phonology. In addition, Labov (1966a:383) mentions that the sample collected in Montreal (Sankoff & Sankoff, 1973; in Labov 1966a:383) was "the first computerized sociolinguistic corpus" (*Ibid*.:383) and the first sample on which in-depth logistic regression analyses were conducted.

The previous sections introduced the field of variationist sociolinguistics and discussed the main assumptions that guide VS analysis, for instance, that language is a heterogeneous system which contains structured variability and because of this, the object of study in VS is the linguistic variable. Previous phonological studies have confirmed that language change in progress is directly observable through the apparent-time construct and have also confirmed that linguistic variables can be fruitfully associated with social factors. Subsequent studies expanded VS beyond phonology and illustrated how the methods of VS could be successfully applied to other areas of the grammar such as the analysis of syntax since variation also exists in these other areas; variation that is systematic and linked to internal and external constraints. Gender, social class and age are the social factors commonly linked to linguistic variables in VS studies in order to identify differences in the speech of young vs older people, men vs women and the differences in the speech of the various social classes; most importantly, social factors have been used to identify linguistic variables that are undergoing changes in a specific speech community. The current study investigates how quotative *be like* and DM-*like* are correlated with the social factors of gender, social class and age. The following sections introduce and discuss the concepts of social class and gender within the context of VS.

2.3 Definition and Background to the Concept of Social Class

Labov (1966a:133) gives a general definition of social class: that it can be seen as a hierarchy based on the social rank of a person, a rank which is determined by the local community. Within sociological theory, the concept of social class is viewed as a lens through which to identify group divisions in society – referred to as "social stratification" (Stewart & Greenstein, 2020:231). Stewart and Greenstein (2020:231) state that the concept of class emerged from "Marx and the conflict perspective" and that "class has been presented as a factor that not only stratifies society into social groups, but shapes almost every aspect of our lives". The Marxist perspective views social class as a struggle between different groups in a society; a struggle between "oppressor and oppressed" (Marx & Engels, 1848:8). Marx and Engels (*Ibid*.:9) state that societies all over the world have been organised into a hierarchical structure of "social rank":

> In ancient Rome we have patricians, knights, plebeians, slaves; in the Middle Ages, feudal lords, vassals, guildmasters, journeymen, apprentices, serfs; in almost all of these classes, again, subordinate gradations. (Marx & Engels, 1848:9)

Marx and Engels (1848:9) believe that society at large is divided into two major classes that are hostile to each other: the "bourgeoisie"[5] and the "proletariat" (*Ibid*.:9)[6]. In the Middle Ages in Europe, a "serf" was a land worker who was a servant to the owner of the land. The landowner was not allowed to remove the worker from the land, and in cases where a new owner took over, the serf would be transferred to the new owner (OED, 2023). At the time, only two social classes existed: the landowners, who were typically nobility or aristocrats, and institutions such as the church as well as state entities. The other social class comprised the peasants or serfs (Kahan, 1973:92, 95). According to Kahan (1973:97), some Western cities developed charters that guaranteed the freedom of their citizens, which caused some serfs to escape to the cities. However, the decline of serfdom was mainly caused by economic depression, which forced landowners to become more lenient in their contracts with their land workers (for who they had to provide compensation), transforming "serf rent-payers into tenants and free-holders" (*Ibid*.:98).

According to Marx and Engels (1848), the bourgeoisie developed from the free citizens who were former serfs and became known as "chartered burghers" (Marx & Engels, 1848:9). The free burghers (later termed the "bourgeoisie") then developed into an intermediate class between the upper (noble) and lower (land working) classes (Wallerstein, 1988:92). According to Marx and Engels (1848:10), the bourgeoisie (or former serfs) struggled to rid themselves of the rule of "feudal nobility" (*Ibid*.:10) and later achieved self-government (an example of this being the result of the French revolution). As society industrialised, the bourgeoisie/middle class became the manufacturers of products and the proletariat/working class became the factory workers (Marx & Engels, 1848:17). Marx and Engels (1848:19) include the lower middle class (also known as the petite/petty bourgeoisie (Scott & Marshall, 2014:785)) in the proletariat group, defined as "the small tradespeople, shopkeepers, retired tradesmen, […] handicraftsmen and peasants" (Marx & Engels, 1848:19), since their capital is too small to compete with the production of the large capitalists and their skills become worthless in the face of "new methods of production" (*Ibid*.:19).

Stewart and Greenstein (2020:233) explain that Marx believed people's awareness of their social rank in society is shaped by their role in production and by the system of the division of labour and resources. He believed that the class system was a result of the struggle over ownership and control of resources. In an early definition of social class, the working class comprised anyone who worked for a salary; the middle class and some of the lower middle class/professionals were those who did not. According to Stewart and Greenstein (2020:236), Marx's argument was that

[5] or "manufacturing middle class" (*Ibid*.:9–10).
[6] or the "modern working class" (*Ibid*.:17).

"people who share the same objectively definable material conditions will most likely share the same class interests". As society widened into a variety of social groups, Marx's division of classes were obscured and became two main and two subordinate classes, especially as differences within the middle class became larger (Stewart & Greenstein, 2020:236).

2.4 Social Class and Linguistic Variables

The reasons why social-class stratification occurs in language use is suggested by Chambers (2002:56). He explains that different social classes are rarely in close contact with each other and because of this distance, differences in each group's language use develop. He also notes that grammatical variables are often markers of social class and that they can even mark social class more clearly than phonetic variables (*Ibid*.:59). According to Chambers (2002:61), it is hypothesised that variables in the dialects of extremely mobile societies would be less sharply stratified by class than those in the dialects of more restricted societies. In addition, Holmes (2017:149) states that the linguistic similarities of individuals in a specific group are more apparent in societies where the social hierarchy is clear.

The concept of "linguistic diffusion" is characterised by Labov (2007:347) as "change from above" (*Ibid*.:346). Language change from above[7] refers to a change that occurred outside of the linguistic system, for instance, change influenced by social factors (Tagliamonte, 2012:59). Labov (2007:346) defines linguistic diffusion as the result of the contact between speech communities and "the transfer of features from one to the other" (*Ibid*.:346). This transfer mainly happens between adults (*Ibid*.:350). Tagliamonte (2012:59) explains that in the diffusion of a linguistic item from one speech community to the next, "the original patterns weaken and there is loss of structural features". Labov (1972:123) states that changes from above involve social pressures from higher status groups applied to linguistic forms. Such pressures involve social correction in the form of "hypercorrection" which the lower middle class apply to their own speech, especially regarding prestige variants such as rhotic /r/ in NYC (*Ibid*.:124).

Hypercorrection occurs when, typically the lower middle class exhibit patterns in speech behaviour that are closer to a higher social group, such as the middle class. For example, in Labov's (1972:124, 127) study of NYC, he found that the lower middle class exhibited infrequent use of rhotic /r/ in casual speech that was similar to the working class. However, the lower middle class rapidly increased their use of it as the interview style became more formal, even surpassing the frequencies obtained by the upper middle class in more formal styles[8]. In addition, Labov (2001:274) explains that changes from above could involve the diffusion of a new prestige variant

[7] As opposed to below: change that results from internal (linguistic) factors.
[8] The interview styles went from casual speech, careful speech, reading style to word lists – where word lists are the most formal.

from outside the speech community, or changes in the distribution of known prestige variants within the community. Unlike changes from below, changes from above occur at a high degree of social awareness (speakers are conscious of them), they exhibit a higher frequency in formal styles, they tend to undergo hypercorrection, and they sometimes turn into "overt stereotypes similar to stable sociolinguistic variables" (Labov, 2001:274). Labov (2001:274) states that for changes from above, women are usually the leaders, thus, they typically exhibit higher frequencies of prestige variants than men, while exhibiting lower frequencies of stigmatised variants. In addition, it is usually the "second highest status group" (such as the middle class) who use prestige variants the most due to hypercorrection, thus, they normally surpass the frequencies obtained by the highest status group (often the upper middle class).

Contrary to diffusion, the concept of "transmission" refers to linguistic changes that happen during the process in which children acquire language from their parents, where deeper structural changes occur (Labov, 2007:346). For instance, "the ability of children to replicate faithfully the form of the older generation's language, in all its structural detail" (*Ibid*.:346), and then "increment" or advance the language further (Labov, 2018:177). Transmission involves changes from below, where change occurs from within the linguistic system (Labov, 2007:346). The relationship between transmission and diffusion is illustrated by Labov (2018:2) with the example of quotative *be like*. He states that the spread of *be like* was first initiated via "diffusion […] into the community by adults who have acquired it from elsewhere[9]" (*Ibid*.:2). From there it was "followed by transmission and incrementation in the next generation" (*Ibid*.:2).

Labov (1972:293) states that linguistic changes can spread from any of the social classes. Labov (2001:31) notes that early theories of linguistic change determined that change happens either from the highest social class or the lowest social class. However, many studies of linguistic innovations indicated that linguistic change is usually located in the middle class (Tagliamonte, 2012:31) or the "central section of the socioeconomic hierarchy" (Labov, 2001:500). This resulted in the identification of a curvilinear pattern of linguistic diffusion, or the curvilinear hypothesis, stated as follows:

> Linguistic changes do not originate in the highest or lowest social classes, but in groups centrally located in the socioeconomic hierarchy. (Labov, 2001:188)

Thus, Labov's principle 1 or the "Curvilinear Principle" (2001:188) predicts that language change from below[10] (Tagliamonte, 2012:58) normally originates in the middle social classes and then spreads to the lower and upper classes. Labov (2001:32) states that studies that provide evidence

[9] I.e. acquired it through language contact between adults.
[10] Changes that happen without the speakers' awareness of the change – a second definition of "change from below" compared to the definition discussed previously.

for this pattern of linguistic diffusion are: “the raising of (oh), (ay), and (aw) in New York City (Labov, 1966a), the backing of (el) in Norwich (Trudgill, 1974b), and the lenition of (ch) in Panama City (Cedergren 1973)”. However, according to Labov’s findings (1966a:172–173), some linguistic variables correlate more strongly with social class than others, and some variables correlate more strongly with an individual class indicator such as education. There are also some variables that have a weak connection to social class, such as the vocalization of (l) in Australian and New Zealand English, illustrated in Horvath and Horvath’s (2001:37) study.

Labov, (1966a:172) explains how social class affects language as follows: the first influence on a child’s speech is their peer group and a child’s peer group is largely determined by the race, sex, and social status of their parents. A person’s childhood speech pattern is possibly modified later based on education. Labov (1966a:172–173) states that: “‘modified by education’ may be taken as shorthand for ‘modified by influences of associates during the period of his education’, and therefore measured by the highest educational level attained”. The next possible influence on speech patterns would be a person’s occupation after leaving school and later in life. The last possible social influence on a person’s speech is their income, which usually reflects their current social status. Labov (1966a:172) gives an example of this process:

> An individual’s use of (dh) and (th), as part of the pattern determined by his sex, ethnic group, and parental background, will be modified more by education than by occupation, and more by occupation than by income.

For the current study, occupation and education were used as two of the indicators to determine the social class of the participants, along with housing (as formulated by Trudgill (1974)), school type, and number of rooms in a dwelling. The last two indicators were chosen in order to be more specific to the South African context and follows indexes created by two sociological studies discussing the stratification of the middle classes in South Africa (Visagie & Posel, 2013; Seekings, 2003). The methods for applying these indicators to the social class grouping of participants are discussed in detail in Chapter 3, 3.4.

2.5 Correlating Linguistic Variables to Gender

Chambers (2002:117) states that there is a distinction between “sex” and “gender”, where “sex” refers to the biological differences between men and women that are formed from birth, such as the differentiation of genitals. However, the differences between genders are those that relate to the sociological concepts of “masculinity” and “femininity” and occur after birth. Chambers (2002:118) notes that the biological difference between the sexes is usually chosen in VS as the social factor correlated with linguistic variables. Coates (2004:4) defines gender as: “socially constructed categories based on sex” and states that most societies function with the view of two genders: masculine and feminine. However, she states that more recent theories have challenged

this binary division and that gender is now often thought of as "plural, with a range of femininities and masculinities available to speakers at any point in time" (Coates, 2004:4). Thus, Coates (2004:3) poses the question: do women and men talk differently?

Empirical evidence from numerous sociolinguistic studies have illustrated that the speech of men and women do differ. For instance, sex differentiation was found for the variable (-ing) in Norwich; England (Trudgill, 174:93); for the non-standard past tense forms, e.g., *seen* and *done* in Sydney, Australia (Eisikovits, 1981; 2011:40–41); and for the phonological variables (uh) and (ay) in Belten High, Detroit (Eckert, 1999; cited in Coates, 2004:61). Some frequently found differences in the speech patterns of men and women have led to what Labov (2001:262) calls the "gender paradox". The gender paradox contains the following principles of linguistic change:

> Principle 2: For stable sociolinguistic variables, women show a lower rate of stigmatized variants and a higher rate of prestige variants than men. (Labov, 2001:266)

> Principle 3: In linguistic change from above, women adopt prestige forms at a higher rate than men. (*Ibid*.:274)

Coates (2004:47) defines "prestige" as linguistic forms that are usually used by social groups with a high social status, whereas "stigma" is linked to non-standard linguistic forms that are usually heavily criticised by society and viewed as incorrect. Conversely, principle 4 states that:

> Principle 4: In linguistic change from below, women use higher frequencies of innovative forms than men do. (Labov, 2001:292)

Because of such patterns as stated in principle 4, women are viewed as the leaders of linguistic change (Labov, 2001:294). However, Chambers (2002:141) cautions against viewing women as less or more innovative than men or connecting any constant constraint to gender, since other factors such as social and geographic mobility might be involved. Chambers (2002:143) also notes that differences in the mobility norms of men and women exist, which can also be contributing to gender-based variability. For instance, Milroy (1987:83) states that in the study of Belfast, the social network of the women in the Clonard were more similar to male social networks, where they worked together and socialised in close-knit groups outside of work, whereas the women in Ballymacarett were not part of such local groups. In addition, Tagliamonte (2012:34) states that it is important to remember that the principles mentioned above are oversimplifications and that differences in the speech of men and women are closely associated with the social context and the social evaluation of the linguistic forms in question. She further states that it is important to consider questions such as "what makes one form prestigious and another form stigmatized or why women favor certain forms while men prefer others" (Tagliamonte, 2012:34).

Some of the reasons that have been proposed for why men and women's speech differs have been elements such as the different roles men and women tend to have in society, especially in societies where men hold more power positions than women. In societies where men and women's roles are less distinct, both men and women use the same linguistic forms, but differ in frequency (Holmes, 2017:171; Mesthrie, 2009:216). Labov (2001:308–9) states that when an innovative linguistic change has been accepted into the community, differences in gendered usage usually disappear.

For the current study, interview participants were selected on the basis of the binary categories of male and female. This is because, according to Labov (2001:263), even though gender influences language as a social factor and not as a biological factor, fieldworkers usually divide participants according to biological sex, thus, they do not consider sexuality. Labov states that:

> Despite the interest shown in recent years in searching for specific features of homosexual speech, reports on gender differences in language continue to report numbers of males and females, rather than 'straight males,' 'gay males,' etc. (Labov, 2001:263)

2.6 Race and Ethnicity in Sociolinguistic Theory

Race has been investigated in sociolinguistic theory as a social variable that has an impact on language use. There are certain social factors connected to race and ethnicity[11] that influence the speech patterns of different racial groups who speak the same language variety; these are inequalities, residential segregation, a common history (Milroy & Gordon, 2003:108–109), a display of ethnic identity (Fought, 2013:388), as well as solidarity and covert prestige (Saville-Troike, 2003:72). Because of these factors, groups such as black and white speakers who speak the same language can be similar in some areas of their language use but differ in others. For instance, in Labov's (1972:118) study of New York City speech, he found that black and white New Yorkers show a similar structure of social and stylistic variation but show differentiation in other variables, such as in semantic variation, in which the African American NYC community used, for example, *mother-wit* to mean *common sense*, which was not found in white New Yorker speech (Labov, 1972:118–119). Labov (1972:118) attributed such differences between black and

[11] Milroy & Gordon (2003:109) define race as referring to physical differences in people which are identified by members of a community and could refer to characteristics such as different skin colours in people. Race is also closely connected to ethnicity. They define ethnicity as referring to a group of people who identify with each other through "common cultural characteristics" which are learned from a young age (*Ibid*.:108). Ethnic groups also share a sense of belonging to a specific place, a common history, social, and religious identity, as well as a language or "communicative conventions" (*Ibid*.:108). Race is a broad category, usually distinguishing groups of people based on genetic similarities, for instance, white people are usually classified as Caucasian based on the genetics of early Europeans, while other racial groups are classified as, for example, the African race, the Asian race and so on (Boyd, 1963:1063–1064). On the other hand, ethnicity refers to narrowly defined groups, such as the Catholics and Protestants in Belfast who are viewed as different ethnic rather than religious groups because of distinct cultural characteristics and histories (Milroy, 1987:71).

white speech to the social (and resulting geographical) segregation of the city and also notes that ethnic identity can have a large impact on linguistic differentiation, sometimes more so than class (Labov, 1972:296). Thus, Labov states that:

> linguistic change may also be differentiated by its association with a particular ethnic or caste group, and that various ethnic groups may treat the same variable in different ways. …contracting one's focus to homogeneous groups—is now somewhat more difficult. A group of working-class men will no longer do; we must be sure to insist on working-class Italian men. (Labov, 1972:298)

There are cases in which racial identity can even mark the speech of standard English speakers. For instance, in one example, an African American judge used rhetorical devices to indicate his racial identity (Fought, 2013:388). Fought (2013:396) explains that:

> Even where, on the surface, extensive interethnic contact and integration might seem to be the norm, the study of linguistic variation reveals the underlying preservation and expression of identities divided along the lines of ethnicity.

In addition, Tagliamonte (2012:38) states that even where the same variable is used between two different ethnic communities, the use of the variable is predicted to differ from one ethnic group to the next, either in the frequency of specific variants or in the pattern of structural constraints influencing the variation.

However, the consensus among sociolinguists is that a language variety is a learned behaviour that is not connected biologically to race or ethnicity (e.g. Labov, 2001:245; Trudgill, 2000:43; Wolfram, 2007:77). For this reason, there are cases in which, for example, white Americans living within an African-American community acquire the variety of the dominant group and vice versa; but these cases are rare (Trudgill, 2000:43). As Trudgill (2000:44–45) states: “Ethnic-group differentiation in a mixed community, then, is a particular type of social differentiation and, as such, will often have linguistic differentiation associated with it”.

The previous sections discussed the concepts of social class and gender and how these social factors are correlated to linguistic variables. First, social class was defined, and an overview was provided of the social hierarchy of European communities in the Middle Ages; how Western societies came to be divided into lower, middle and upper classes, and who belong to these classes. After defining social class, a review was provided of how linguistic variables are correlated to social class. Some of the important concepts which are linked to social class were discussed, such as diffusion (relating to change from above) and transmission (relating to change from below). Quotative *be like* and DM-*like* are thought to be variants that are examples of changes from above. Both variants are thought to have originated in America and then spread to other English dialects via diffusion (Tagliamonte & Hudson, 1999:149, 152), after which they were structurally developed further through transmission. Section 2.7 discuss the historical

development of quotative *be like* and DM-*like*, as well as the findings of previous studies regarding the internal and external constraints on these forms.

2.7 Pragmatic *Like*

One of the early examinations into discourse-marker *like* was Sankoff *et al.*'s (1997) study of the discourse markers used by 17 speakers of Anglophone Montreal French (AMF), including pragmatic *like* (*comme* in Québécois; Sankoff *et al.*,1997:191). In general, pragmatic uses of *like* are placed in the category of discourse markers (except quotative *be like*, which is a fixed grammatical string and therefore placed in a category of its own). Sankoff *et al.* (1997:195) define discourse markers as elements of the grammar that do not enter into a sentence syntactically – this excludes linking adverbs such as "unfortunately" (*Ibid*.:197) and conjunctions with the original semantics intact. The meaning of a sentence does not rely on the presence of discourse markers, and they usually lose the semantic meaning in their original use (this process is known as "semantic bleaching" (Sankoff, 1997:196–197)). In selecting her variable context, Sankoff did not, for example, include uses of *comme* as a conjunction in the analysis, since it falls outside of the definition of discourse markers. Her study followed the Labovian sociolinguistic interview method[12] (Sankoff *et al.*, 1997:200). Thus, with this study, Sankoff illustrated that, similarly to phonetic studies, it is important to circumscribe the variable context of the linguistic variable.

The next section provides an overview of the historical development of *like* and its transition from a lexical item with concrete semantics to its almost complete semantic bleaching, where it became part of a semi-grammatic fixed string in the case of quotative *be like*, as well as a fully pragmatic discourse-marker (in the case of *like* on its own). Secondly, an overview of previous studies of quotative *be like* and DM-*like* is presented.

2.7.1 The grammaticalisation of *like*

According to Millar (2023:130), "grammaticalisation" refers to a type of syntactic change. For instance, Gelderen (2014:59) states that the word *to* was restricted to the very specific meaning of location in Old English but was later used as an indirect object marker in sentences such as: "I gave it to Marta" (*Ibid*.:59). Thus, Gelderen (2014:59) states that through the process of grammaticalisation, the semantic meaning of a word slowly disappears and the grammatical meaning is left. Millar (2023:130) states that this kind of grammaticalisation has been termed "bleaching" and that grammatical items commonly originate from the bleaching of lexical meaning. Romaine and Lange (1990:257) state that when a grammatical item becomes more grammatical, this change involves a process of grammaticalisation. They also state that because change happens slowly, there is a period in which old forms exist simultaneously with new forms, which

[12] See Chapter 3, section 3.6 for a full description of this method.

leads to ambiguity. In Andersen's example (1996:42), the following use of *like* is ambiguous: "What, so she was **like** your friend?". He states that the speaker is either using *like* with the meaning of "similar to" a friend who is not the person referred to initially, or they are using it as a discourse marker in which *she* and *your friend* are the same person.

Romaine and Lange (1990:261) propose the following grammaticalisation path for *like* in Figure 2.1 below. They explain that in the first stage, it was used as a preposition, taking a nominal or pronominal complement, e.g., "She looks **like** her father" (*Ibid*.:244). It was then reanalysed, taking a clause complement, at which time it was used as a conjunction: "Winston tastes good **like** a cigarette should" (Romaine & Lange, 1990:244). They then state that because *like* could appear as a prefix or suffix (e.g. *princess-like; like-minded*), it could be reanalysed as a discourse marker "which shows syntactic detachability and positional mobility" (*Ibid*.:261), such as: "**Like**, this is a crisis" (*Ibid*.:244).

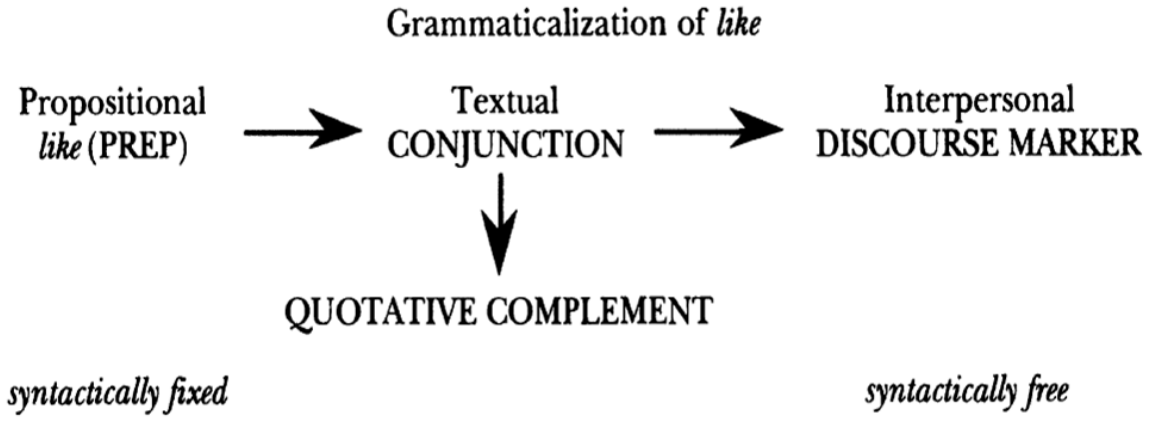


Figure 2.1: Grammaticalisation of like

Romaine and Lange, (1990:261) state that when *like* appears before a clause or sentence that is a quote, it functions as part of the "quotation frame". Thus, because English does not license sentences without verbs, the dummy verb *be* is needed to complete the quotation frame syntactically. They state that during this stage, reanalysis and "syntactic fixing" forms part of the grammaticalisation process (*Ibid*.:262). Thus, in the last stage of *like's* grammaticalisation path, it forms the quotative *be like*, as in:

(30) I **was like**, "If it doesn't work, can I come back to the show?" And they said yes. (Romaine & Lange, 1990:253).

Sections 2.7.4–2.7.5 review previous studies of quotative *be like* and DM-*like* and their main findings. However, quotative *be like* is first placed in the context of the English quotative system in sections 2.7.2–2.7.3.

2.7.2 The quotative system in English

In Tannen's (1986:311, 313) study of the quotative system in Greek and American conversational narrative, she explains that the term "reported speech", referring to dialogue in storytelling and conversation, is not accurate. This is because most of the lines represented in direct quotation were not actually reproduced verbatim due to the failings of human memory. Thus, Tannen uses the term "constructed dialogue" to refer to dialogue that was reconstructed based on limited memory; hypothetical speech that was never actually spoken, such as: "You can't say, 'Well Daddy I didn't HEAR you'" (*Ibid*.:313); representations of the speaker or someone else's thoughts, such as when a Greek woman reported what someone else thought of her: "He says to himself, 'She's up to no good'" (*Ibid*.:314); dialogue represented as the speech of a group of people: "And then all the Americans said 'Oh in that case, go ahead'" (*Ibid*.:314), and instances where an indirect quote fades into a direct one: "It was like he was telling everybody to have your wisdom teeth taken out and I didn't see any point as long as they weren't bothering me". Tannen (1986:314) highlights the process by which the indirect quote blends with a direct quotation by explaining that "the addressee metamorphoses from 'everybody' to 'you' to 'me'".

Tannen (1986:312) views constructed dialogue as a poetic process, one of the "poetic linguistic devices" that is "spontaneously used in ordinary conversation" to create drama since first-person narrative is "more vivid" compared to indirect speech (*Ibid*.:311) and "creates interpersonal involvement among speaker or writer and audience" (*Ibid*.:312). Tannen's definition of constructed dialogue is an important one that has been adopted in VS studies of the English quotative system (e.g. Blyth *et al.*, 1990:216; Romaine & Lange, 1991:231; D'Arcy, 2012:343). It has been used as the basis from which the quotative system is approached. For instance, in Tagliamonte and Hudson's (1999:154) study of quotatives in British (BrE) and Canadian English (CanE), they specify that "each and every instance of a verb introducing constructed dialogue" is included in their analysis. The current study takes a similar approach and uses Tannen's definition (as illustrated with examples above) as the outline from which to analyse *be like* within the context of the quotative system in L1 White South African English (WSAE).

2.7.3 The historical perspective of quotative *be like*

Since older speakers reflect earlier patterns of the quotative system (Buchstaller & D'Arcy, 2009:304), it can be assumed that the pattern in the older group (40+) in Ferrara and Bell's (1995:275) 1990 Corpus of American English (AmE) reflects an earlier state in which *be like* did not appear yet and in which *say* was the main quotative: *say*: 91%, *go*: 2%, *be like*: 0%. For lack of studies including earlier data of AmE, this pattern is compared to longitudinal data of other English dialects to illustrate the development of the quotative system through the years. D'Arcy (2012:346) conducted a diachronic analysis of New Zealand English (NZE) via three

sociolinguistic corpora spanning 1946–2006. The earliest corpus is *The Mobile Unit* (MU), recorded between 1946–1948 (speaker dates of birth: 1851–1910). The MU corpus consisted of fewer women than men (*Ibid*.:348). D'Arcy (2012:349) states that:

> Contemporary perspectives of direct quotation reveal a system that is robustly variable, both lexically and internally, via the operation of a number of constraints that function in tandem as a "choice mechanism" (Poplack, 2011:213). The MU data provide no such perspective. The system can best be described as "contained": What variability exists is highly circumscribed (D'Arcy, 2012:349–350).

The results for the 19th century (1860–1894) showed that there were only three quotatives being used, with a few in the "other" category. The only quotatives at this stage were: *say* (89%), zero (8%), *think* (1%) and other (3%), with *say* dominating (*Ibid*.:350). One can only assume that earlier versions of AmE looked similar in pattern, with perhaps the exclusion of the zero quotative, since Ferrara and Bell (1995:275) note that in 1990, "the older speakers in this corpus have basically two choices: *say* and *said*. Use of *go* was infrequent".

D'Arcy (2012:350) observes that in the MU data, quotatives were restricted mainly to direct-speech reporting and appeared rarely with thought reporting. For instance, *think*, occurring just 1% of the time, was the only quotative used for thought/inner dialogue reporting. Furthermore, D'Arcy (2012:350) reported that the quotatives were used the most with the past tense (69%), followed by the historical present (HP; 25%), where the HP is only used with *say.* The simple present was used minimally with 5% of the quotatives, the past progressive occurred only once, and there were no other aspects and no modal usage (*Ibid*.:351). Ferrara and Bell's (1995:275) data of AmE similarly illustrated mainly past tense usage for *say*: 66% vs 25% for the present tense in the 40+ age group (they included the HP in the present tense category). For grammatical person, D'Arcy (2012:351) observes that the MU data showed that the quotative system mainly utilised third-person reports (80%) and that "speakers are less likely to quote themselves (20%), and second persons are virtually unattested (N = 2)". Ferrara and Bell (1995) did not provide an analysis of grammatical person for *say*, however, Blyth *et al.*'s (1990:219) data of AmE, also recorded in 1990, illustrated that for the older and younger groups combined (20–72), *say* obtained the highest probability in the past tense (0.670) vs minimally for the present tense (0.242). Thus, in English, the earlier state of the quotative system was that *say* was the dominant quotative that was (and still is) mainly used to report actual speech (vs thought) and was also mainly used to report the speech of third persons.

In a comparative cross-variety analysis of AmE, England English (EngE[13]) and NZE, Buchstaller and D'Arcy (2009:305) found that in the Switchboard corpus (AmE, recorded in 1988–1992), *be like* started to emerge for the older speakers (3.6%, N = 16), while it was rare (0.5%; N = 305) in

[13] EngE refers to the region of England only, as opposed to BrE which includes Scotland, Wales and Ireland.

the EngE corpus (Derby & New Castle Corpus, recorded in 1994–1995) and did not appear in the Canterbury corpus (recorded in 1994–1996) for NZE. Cukor-Avila (2012:624) traced the rise of *be like* over six age groups ranging in speaker birth dates of 1901–2002 for African American Vernacular English (AAVE) in Springville (recorded between 1988–2010). She found that *be like* started to rise dramatically with speakers born in the 1970s and that these speakers were a decade behind the Canadian English (CanE) speakers in the adoption of the form as the main quotative, which occurred from the 1980s in CanE, whereas the AAVE speakers adopted *be like* fully from the 1990s. She also observes that the rise of *be like* was accompanied by the steady decline of *say* (Cukor-Avila, 2012:624). It is interesting to note that in Tagliamonte *et al.*'s (2016:831) study, in which they similarly trace the rise of *be like* in CanE, NZE and Australian English (AusE), their results reflect Cukor-Avila's in that it starts to rise rapidly from speakers born in the 1970s for the three mentioned dialects. However, the fact that the AAVE speakers did not use *be like* more than *say* for a decade is likely due to the fact that the Springville corpus mainly consists of rural speakers (working class). Ferrara and Bell (1995:285) state that urban speakers (middle class) were the most frequent users of the form in their 1992 corpus (98%), with few rural speakers 1992 (1%) (*Ibid*.:277, 285). In addition, Singler (2001:273, n16) observed that the rates for *be like* in the Springville corpus were much lower compared to Sanchez and Charity's (1999; in Singler, 2001:273) results for their urban AAVE speakers. Thus, it is hypothesised that the use of *be like* originated in the urban/middle-class regions of North America and Ferrara and Bell (1995:286) hypothesised that it only then spread to rural/working-class regions.

2.7.4 Previous studies: Quotative *be like*

Tannen (1986:15) analysed AmE narratives that were recorded in 1983, and at that time, *be like* was emergent, consisting of 8% of the total quotatives (N = 84/100%), while *say* was the most frequent quotative (43%), followed by the zero quotative (26%). The zero quotative is defined by Tagliamonte and Hudson (1999:156) as cases where direct speech is introduced with no quotative verb or attributed speaker. These are instances such as:

(31) This bloke trailed her in the front door. "Josephine, who's this?" (Tagliamonte & Hudson, 1999:148)

Go was the fourth most frequent (13%). Other quotatives that were infrequent were *ask*, *tell*, and *think*. Another group of quotatives that appear infrequently are what Tannen (1986:322) calls "graphic introducers". These are verbs such as: *call*, *scream*, *shout*, *explain,* and any verb that is not *say*, *think*, *be like*, *tell* and the others listed above. For example:

(32) Ina ducked away, **screaming**, "Get away from me!" (Tannen, 1986:322)

(33) And so one of the neighbors saw it and **called out**, "Claire." (Tagliamonte & Hudson, 1999:155)

(34) Leaf fans are constantly **complaining**, “Oh we haven’t won the Cup in thirty-seven years.” (Tagliamonte & D’Arcy, 2004:500)

In subsequent studies of AmE, *say* remained the main quotative while *be like* rose in frequency. For example, in Barbieri’s (2007:33) study of AmE, she analysed a selection of the American conversation subsection of the *Longman Spoken and Written English Corpus* (LSWE, Biber *et al.*, 1999), recorded in 1995–1996, consisting of 149 speakers ranging in age from 16–55+. The result was that *say* was the most frequent quotative (1251/100,000 words)[14], while *be like* was the second most frequent (472/100,000 words) and *go* was third (239/100,00 words) (*Ibid*.:35). Similar results were found in Ferrara and Bell (1995:274), who analysed narratives of personal experience by 405 speakers native to Texas, recorded in 1980, 1992, and 1994 (*Ibid*.:270). In studies expanding to other dialects of English, such as in Tagliamonte and Hudson (1999:158; EngE, York, UK; recorded in 1996, and CanE, Ottowa; recorded in 1995), *say* was first in both EngE (31%) and CanE (36%), while *go* was second in CanE (22%) but in BrE, *go*, *be like* and *think* were second with the same frequency (18%), and zero was third. Zero was also third in CanE (20%) and *be like* was fourth (13%).

Tagliamonte and Hudson’s (1999:158) results for York, UK had two instances of the quotative *be* without *like*. This is illustrated in the following examples:

(35) I was. "O.K. that's enough!" you know. "Ha! Ha! Ha! (Tagliamonte & Hudson, 1999:155)

(36) Even in class he’d be, “Come here girlfriend.” (Tagliamonte & D’Arcy, 2004:500)

Macauley (2001:16) suggests that *be* on its own is an abbreviated form of *be like* and this is supported by the fact that the verb functions in the same way as *be like* in that it can also introduce both internal monologue and direct speech.

In a study of EngE from the *Bergen Corpus of London Teenage Language* (COLT) by Stenström *et al.* (2002:118), recorded in 1993, it was found that *go* overtook *say* as the most frequent quotative in the data (*go*: 27%; *say*: 25%), thus competing with *say*. In their data, *be like* was still emergent, having only 34 instances (*Ibid*.:117–118). Similarly, Winter (2002:8) found that in her data of 30 speakers aged 15–16 of Australian English (AusE, recorded in 1997–1999), *go* (45%) was more frequent than *say* (24%) and *be like* was emergent as well, consisting of only 8% of the data.

In a comparison of older and more recent AmE data, Rickford *et al.* (2007:6) analysed the speech of California high school and college students (aged 15–25), recorded in 2005, and compared the

[14] Barbieri (2007:34) normalised her results by, for instance, taking the number of quotatives for a group, e.g., 416 for women out of their total number of words in the corpus and multiplying the result by 100,000 words and reported the quotative results out of 100,000 words: ((416/58,454) × 100,000)

results for the overall frequency of quotatives to that of data collected from the same speech community (California high school and college students) by Winner in 1990 and Fought in 1994. The 1990, 1994 results showed that quotative *all* was the most frequent (45.9%) of the total quotatives (N = 246). An example of this quotative is:

(37) The dog just—she was all "Bark! Bark! Bark!" (Rickford *et al*., 2007:3)

Be like was the second most frequent (17.5%), while the zero quotative was third (15.9%) and *say* fourth (10.6%). Interestingly, in this data, *go* was infrequent (2.4%). The results of the 2005 data showed that *be like* rose significantly in the speech of California youth, so that it overtook quotative *all* as the most frequent (69.3% out of the total quotatives; N = 541), while surprisingly *say* rose somewhat in frequency from the older dataset (12%) and zero remained third (10.7%). It also shows that *go* decreased in frequency (1.1%), supporting Vandelanotte and Davidse's (2009:779) observation that *be like* replaced *go* (see Chapter 1, section 1.1.2). Quotative *all*, which was the most frequent in the older dataset, took fourth place in 2005 (4.3%) and decreased significantly in the time span from 1990 and 1994 to 2005. Rickford *et al.* (2007:21) state that the use of *all* is in decline. Furthermore, they state that "only eight tokens of quotative *all* by itself in our 2005 tape recordings come from college students. All of the high school students used *all like*" (*Ibid*.:21), and that *all like* was counted as part of the *all* group, "treating *like* as an approximative or second quotative." Their conclusion is that "the quotative system is unstable and subject to change" (*Ibid*.:22). Their main observations are that new quotative variants are picked up to a certain extent by a speech community but then rejected in favour of a more popular variant such as *be like*. In addition, they state that the rise and decline of *all* parallels that of *go* (*Ibid*.:22).

In a real time study, Barbieri (2009) compared the results to her 2007 study of the *Longman Spoken and Written English Corpus* (LSWE) subcorpus (see section 2.7.4 above for the 2007 study). In this study, she analysed more recent AmE data that was similarly compiled to the LSWE: the *Cambridge Corpus of Spoken North American English* (CAMSNAE), collected from different regions of the US, recorded in 2004–2005 and spanning the ages of 16–55+ (Barbieri, 2009:76–77). The results revealed that *be like* became the most frequent quotative (232/10,000 words), while *say* declined in frequency (136/10,000 words). *Go* declined the most in frequency compared to the LSWE results (44/10,000 words) and Barbieri (2009:83) states that:

> …findings for *go* here square with Buchstaller's (2006a), who suggests that the wave-like, seesaw frequency patterning of this quotative indicates that '*go* is a more sporadic variant which recedes but lingers on and is finally picked up again by the speech community.

This finding lends support to Rickford *et al.*'s (2007:22) statement above that *go* is being abandoned in favour of *be like.* In an updated study by Tagliamonte *et al.* (2016:826), three

dialects of English are compared in their use of the quotatives: Canada, New Zealand and Australia. They analysed the following corpora. Table 2.1 illustrates the details of their corpora:

Place	Variety	Recording date	Speaker birth dates
Toronto; Victoria	CanE	2002–2004; 2011–2012	1916–1996
Christchurch	NZE	2000–2005	1939–1983
Perth	AusE	2011–2013	1922–2002

Table 2.1: Corpora details of Tagliamonte (2016:826–828)

Tagliamonte *et al.* (2016:831) did not provide exact numbers, only graphs of the rise of *be like* vs other quotatives (*Ibid*.:831). They state that "*be like* overtook a large proportion of the quotative system in a relatively short period—under three decades" (*Ibid*.:831). In their comparison of the age of speakers, they found that all three dialects share similar trajectories for *be like*, for instance, in all three, it increased in frequency among speakers born in the 1970s, therefore it increased simultaneously in three separate geographical locations: "In other words, despite the geographic breadth and the diversity of these urban centres, *be like* increases at the same time, in the same cohort, in all locales" (*Ibid*.:831).

Their results indicate that in more recent data of dialects beyond America, *be like* has also become the most frequent quotative. However, the data for NZE showed that even though *be like* was rising in frequency, *say* was still the most frequent (Tagliamonte, 2016:831). This result is likely due to the fact that the data for NZE are the earliest recordings of the three corpora and consist of speakers born only up to 1983, whereas the other two corpora consist of wider age ranges with a larger number of younger speakers (*Ibid*.:831).

Because of the similar trajectories across the three dialects, Tagliamonte *et al.* (2016:842) hypothesise that *be like* is a "black swan event". This means that it is an "outlier" that could not have been predicted from past events; these events are a surprise and carry immense impact. They state that because language is not entirely predictable, it holds the possibility for such an event to occur. They propose that because *be like* is an outlier, it might not fit into the conventional sociolinguistic model of diffusion in which a linguistic innovation spreads in increments from one community to the next among adults through face-to-face contact (known as the wave model). The wave model refers to the pattern in which the characteristics[15] of a linguistic innovation expands "as a wave of continuous weakening[16]" as it reaches each new level of diffusion outward (Labov, 2007:350; Tagliamonte *et al.*, 2016:841). Tagliamonte *et al.* (2016:242) state that the reason why such a black swan event in language is possible is that modern society is a changing and increasingly complex social environment, where social networks are becoming more complex

[15] Such as structural development.
[16] For instance, the weakening of structural complexity.

and speakers are “interacting both locally and globally, and these interactions reinforce and spread changes as the same individuals move through their own local and global networks” (Tagliamonte *et al.,* 2016:840–841).

Very few studies of quotative *be like* and DM-*like* have been done on South African English (SAE). I am aware of only two, which are both M.A. studies. The first is Lochner’s study (2019) which focussed on DM-*like* in White South African English (WSAE): “We’re **like** marking off everywhere we been in South Africa **like** road trips and stuff...” in relation to the marker *just* (Lochner 2019:34). The second study by Horn (2022) investigated *be like* in young Cape Town speech from 64 sociolinguistic interviews of speakers who have been residents of Cape Town for most of their lives. The corpus included black, white, coloured and Indian participants with older white participants as a control group of older linguistic forms (Horn, 2022:33–34). These interviews were recorded from 2006 to 2016. Among the young group (aged 16–35), Horn found that the WSAE group produced the most quotatives (mean of 46.9) compared to the other racial groups (Indian speakers – 6.3, Black speakers – 16.7 and Coloured speakers – 23.6). Within the young WSAE group, *be like* is by far the most frequent quotative (44.3% out of the total quotatives: N = 1792/100%, with *say* being the second most frequent (30.3%, N = 542) and the zero quotative being the third most frequent (9.9%, N = 177). *Go* is the least frequent in the data (2.8%), with *tell* and *think* above it (*Ibid*.:69). Horn’s results for WSAE so far reveal a similar pattern to the studies analysing newer data in AmE, CanE, AusE and NZE discussed above: that *be like* is currently the main quotative, *say* is declining in frequency and *go* seems to be becoming increasingly rare. Tagliamonte *et al.* (2016:825) stated that *be like* has become “highly productive across all national varieties” of the “Inner Circle” English dialects.

The following sections discuss the linguistic (internal factors) affecting *be like*, i.e. those that have been attested in the literature as being constraints on its usage. These are: content of the quote, grammatical person and tense and temporal reference.

2.7.4.1. Content of the quote

Blyth *et al.* (1990:215) state that “most quotatives are associated with either direct speech or inner monologue but rarely with both.” For example, the quotative *think* mainly introduces thought, while quotatives such as *say* and *go* mainly introduce speech (*Ibid*.:215):

(38) And I was **thinking**, “Well surely they can all get on.” (Tagliamonte & Hudson, 1999:148)

(39) I **thought**, “Right OK.” I told them what I was doing. (*Ibid*.:148)

(40) And she **said**, "Would you like me to phone?" (*Ibid*.:148)

Romaine and Lange (1991:243) observed that *be like* was more likely to be used for reporting thought rather than speech. However, Ferrara and Bell (1995:279) found in their data that *be like* was expanding from mainly thought to direct-speech reporting when they compared their 1990 corpus to their 1992 corpus. Dailey O'Cain's (2000:66) study of AmE in Michigan, recorded in 1995, also found that *be like* was used more frequently with internal thought rather than actual speech.

According to Tagliamonte and Hudson (1999:152, 159), the types of constructed dialogue introduced by *be like* is an indication of how far into a speech community it has diffused, where high frequencies of co-occurrence with direct speech (as opposed to thought) signals a later stage of diffusion. They also state that such expansion, along with semantic bleaching, are "indicators of increasing grammaticalization" (*Ibid*.:159). In their study comparing CanE (Ottowa, 1995) with EngE (York, 1996), they found that *be like* is much more likely to occur with thought and non-lexicalised sounds[17] rather than direct speech in both varieties. They also found that, in both varieties, the zero quotative was the most likely to be used with non-lexicalised sounds, second with direct speech and was the least likely with thought, although the results for zero was not statistically significant. *Say* was only used with direct speech in EngE and appeared rarely with thought reporting in CanE (Tagliamonte & Hudson, 1999:152, 159). In a later study, Tagliamonte and D'Arcy (2004:507) investigated quote content in CanE, Toronto (recorded, 2002–2003) and found that *be like* was strongly correlated with thought in the youngest speaker group (aged 10–14). This effect peaked at high school age (15–16), but by the ages of 17–19, its association with thought weakened greatly. They argue that *be like*'s association with thought has greatly weakened in apparent-time and that this provides evidence that it has expanded into direct-speech reporting (*Ibid*:.507). In addition, they interpret this change as evidence for its increasing grammaticalisation. Crucially, they suggest that:

> …as *be like* diffuses in the Canadian context we are beginning to see a bleaching of its pragmatic function. Where once it held its own functional niche, distinguishing internal thought from other kinds of constructed dialogue, it now appears to be extending across all functions of the quotative system (Tagliamonte & D'Arcy, 2004:510).

In AusE (Perth, recorded in 2011–2013), *be like* has also expanded to direct speech, as indicated by the fact that Tagliamonte *et al.* (2016:835) did not find a significant difference between thought and speech reporting regarding its use. The same was found in a recent study of EngE in York (recorded, 2013) by Gardner *et al.* (2020:295).

[17] Non-lexicalised sounds are instances of speech that are not words but sounds such as onomatopoeia, exclamations and so forth. An example of *be like* used with a non-lexicalised sound is the following: "Every five seconds he's like, '(panting noise)'" (Tagliamonte & D'Arcy, 2004:495).

2.7.4.2. Grammatical person

In an early study of the quotative system in AmE, Blyth *et al.* (1990:219) investigated grammatical person as one of the factors constraining *be like* usage, and considered its distribution between first-person singular, third-person singular and third-person plural. Their data consisted of 30 participants from New York, ranging in age between 20–72, recorded in 1990. The results were that *be like, say* and *go* obtained near equal probabilities in first-person singular and third-person plural, while *be like* was rare in third-person singular (0.264) compared to the other two quotatives that, in contrast, were used the most with third-person singular subjects: *go* – 0.708; *say* – 0.659 (Blyth *et al.*, 1990:221).

An important observation made by Ferrara and Bell (1995:279) is that *be like* was, at the time, expanding in AmE in its discourse function. They compared their Corpus 1 (1990) to a second corpus (recorded, 1992) and a third one (recorded, 1994) and found that, while first-person usage of *be like* was 62% and third person was 29% in 1990, first-person usage decreased in 1992 to 60% and even further in 1994 (51%). Third-person usage increased from Corpus 1 to the subsequent corpora: Corpus 2: 33%; Corpus 3: 47% (Ferrara & Bell, 1995:278). They specify that they separated third person *it's like* from the rest of the group because "this variable is used somewhat differently, either chiefly by males to report the collective thoughts of a group, as in…" (*Ibid*.:278):

(41) And now **it's like**, uh, you know, "Oh man, you know, you guys, you hippies are nothing (Ferrara & Bell, 1995:279)

"Or by both males and females to report a habitual style of thought or speech for one individual, as in…" (*Ibid*.:278):

(42) My dad was constantly down on me. **It's like**: "Get a job" (Ferrara & Bell, 1995:279)

Their main finding was that *be like* was expanding from first- to third-person contexts (Ferrara & Bell, 1995:281). This observation seems to be corroborated by Blyth *et al.* (1990:221), who found in their data that, even though *be like* was still rare in third-person singular (example 43), third-person plural (44) was close to first-person singular (45) in frequency.

(43) **He was like** [accent] "I give you a week of my life. Will you stay." (Tannen, 1986:321)

(44) And **they're like**, "Come on, go and get dressed."

(45) **I was like** -"Nice sister, right?" (Tannen, 1986:321)

Buchstaller and D'Arcy (2009:307), provided further evidence for the expansion of *be like* into third-person subjects, where the result was 11.1% for third person vs. 11.7% for first person in AmE in the Switchboard Corpus (1988–1992); the result was barely significant. At first, studies of

grammatical person in CanE, EngE, AusE and NZE did not find that *be like* was expanding into third-person subjects (Tagliamonte & Hudson, 1999:161; Tagliamonte & D'Arcy, 2004:506; Tagliamonte *et al.*, 2016:836). However, in a later study of EngE, York and CanE, Toronto, (1997–2013), Gardner *et al.*'s (2020:296) data indicate that it has increased with third-person subjects in both EngE and CanE over the years to almost equal the rates alongside first-person subjects. However, they still found in both dialects that the difference between first and third person was significant in favour of first-person reporting. However, other studies investigating Scottish English and AmE found that *be like* was used more with third-person subjects than first person: Singler (2001:276; AmE, 1995–1999; first person: 46%, third person: 52%) and Macauley (2001:11; Scottish English, 1997; first person: 36%, third person: 59%). Tagliamonte *et al.* (2016:838) suggest that such differences could either mean that *be like* has a more advanced development in certain dialects or that as it diffused, it developed locally specific constraints.

2.7.4.3. Tense

In AmE, Blyth *et al.* (1990:219) investigated tense as another factor and looked at its distribution between the present and past tenses. At this time, the historical present tense was not considered. Their results indicated that *be like* had the highest probability for use in the present tense (0.712), while the probability for past tense was significantly lower (0.126). This result contrasted with *say*, which was used overwhelmingly in the past as opposed to the present tense, while *go* was also mainly used in the present and rarely in the past tense, similar to *be like* (Blyth *et al.*, 1990:219). In Ferrara and Bell's (1995:274) 1990 corpus, *be like* was equally distributed between the present and past tenses, while the results for *say* was the same (as stated above) in Blyth *et al.* (1990:219). Blyth *et al.* (1990:218) also found that tense was the most significant factor constraining *be like* out of all the other internal factors and stated that it often appeared in a narrative where a switch to the present tense occurs. This detail was investigated more closely in subsequent studies (discussed below); thus the historical present tense (HP) was added to tense analyses of *be like*. In Singler (2001:271; AmE, New York; recorded, 1995–1999), the present tense was also found to be more probable with *be like* than the past tense. In Macauley's (2001:10) study of Glasgow adolescents for Scottish English (1997), however, she found that *be like* was used more in the past tense (58%) than the present. In the historical present tense, the quotative verb is in the present but has past reference (i.e. referring to an event that happened in the past), such as:

(46) She's **like**, "Put this in your eye." (Tagliamonte & D'Arcy, 2007:204)

Winter (2002:11; AusE, 1997–1999), found that *be like* was the most frequent with the HP (72%) vs the simple past tense (27%). These findings reflect the earlier AmE studies in which *be like* was used mainly with the present tense (which included the HP in their analyses). Tagliamonte

and D'Arcy (2007:205, CanE, Toronto; 2002–2004) also found that tense was the strongest factor constraining the use of *be like*. Tagliamonte and D'Arcy (2007:204) describe the HP as: "A specific and highly circumscribed construction in which surface morphology is present tense but the context of temporal reference is past tense" and found that the HP was the most favoured context for *be like* as well. They propose that the developmental trajectory for *be like* regarding tense is the following: that it is mostly associated with the present tense at stage 1. At stage 2 it specialises for the HP, at which point the correlation with the present tense weakens and it is favoured mostly with the HP (*Ibid*.:209). D'Arcy (2012:364–365) found this developmental direction for *be like* to be true in her longitudinal study of NZE.

In a study of more recent data for CanE and EngE spanning 1997–2013, Gardner *et al.* (2020:299) found that both the present and past tense contexts for *be like* increased over the years, and the historical present is associated with speakers born in the 1960s and earlier, whereas the simple past is associated with speakers born in the 1970s and after. Their analysis suggests that the simple past association with *be like* increased with younger speakers and begins with the speakers born in the 1970s. To compare this developmental trajectory with a study of WSAE, Horn (2022:93) found that for young white Capetonian speakers, *be like* was used the least with the present tense and the most with the HP, however the past tense reached almost equal frequencies with the HP. This result corroborates Gardner *et al.*'s (2020:299) observation that the past tense association with *be like* has become more common among younger speakers.

2.7.4.4. Social factors: Gender

In Blyth *et al.*'s (1990) study, they found that the correlation between gender and the use of *be like* was significant, however, contrary to expectations, men obtained a higher probability (0.534) for the quotative than women (0.466) (*Ibid*.:221). Ferrara and Bell (1995:274), on the other hand, found that in their Corpus 1 (1990), women used *be like* more than men (F:29%; M:15%). This result was confined to the 18–25-year-old group since the older group of 40+ had no instances of *be like*. Ferrara and Bell (1995:274) state that Corpus 1 supports the claim that women lead in the use of the quotative, and that "we would expect that over time males would adopt the variable" (*Ibid*.:276). In their Corpus 2 (recorded, 1992), Ferrara and Bell (1995:274) found that gender differences had disappeared in the age range of 16–39-year-olds: women used it twice as frequently in Corpus 1. They thus conclude that the variable is spreading and *be like* is losing its gender-marking (*Ibid*.:277). Dailey O'Cain's (2000:75) study similarly found that by 1995, the difference between men and women's use of *be like* in AmE was not significant. In a study of newer data of AmE (recorded between 1994–1999), Singler (2001:267–268) also found that the gender difference was not significant, however, he noted that in the same-sex dyads (speaker and interviewer), female pairs used more *be like* than male pairs, while mixed pairs weakly disfavour it (Singler, 2001:272).

In more recent data for AmE, Barbieri (2009:76–77) investigated the *Spoken North American English Corpus* (CAMSNAE; recorded, 2004–2005). Her age groups were divided into: <16, 16–26, 27–40, 41–54 and 55+. These results showed that girls below 16 used far more quotatives than all the other age groups (*be like* being the most frequent) and they used *be like* six times more than the boys in their age group. However, her results for the other age groups indicate that men used slightly more *be like* than women in the 16–26 age group (M:39.7/10,000 words; F:20.6/10,000 words) and women used it more than men in the 27–40 age group (M:4.3/10,000 words; F:42.8/10,000 words). However, she states that the 16–26 age group's rates were based on a small sample and should thus be considered with caution (Barbieri, 2009:79). Barbieri's significance testing (2009:81) revealed that *be like* was used equally as much (50%) of the time for men and women below the age of 40 but she states that this cannot be taken as evidence for gender neutralisation due to the low numbers for participants in some of the cells. In comparing the CAMSNAE corpus to the results of the spoken American subset of the LSWE (1995–1996), she found that from 1995 to 2005, both men and women have significantly increased their use of *be like* below the age of 40, by 20–50% in over ten years and that this result signals a "true case of change in progress" (*Ibid*.:85).

Many of the AmE studies have so far found that there is no significant gender difference in the use of *be like* (Barbieri, 2009; Ferrara & Bell; 1995; Singler, 2001 (see above); Buchstaller & D'Arcy, 2009:209). In other dialects, Tagliamonte and Hudson's (1999:159) gender results for CanE and EngE (recorded, 1995–1996), illustrated that in EngE, *be like* was used more by females (.67 vs .36 for males) and men use *say* more (.57 vs .41 for females). In CanE, however, females used *say* more than males, while the difference between men and women's use of *be like* was statistically not significant (reflecting the majority of the AmE results discussed above). Tagliamonte and Hudson (1999:159) interpret these results as conflicting with the idea that gender differences neutralise at later stages of the diffusion of *be like*, since it is used by more speakers and at higher frequencies in the EngE corpus vs the CanE corpus, yet it is in the EngE corpus where females take the lead. They conclude that: "the more diffused *be like* is, the more likely it is to differentiate male and female speech, at least at an early phase of development" (*Ibid*.:167). In a comparative study of CanE, AusE and NZE (see Table 2.1: Corpora details of Tagliamonte (2016:826–828)), Tagliamonte *et al.* (2016:836), divided their data into two eras, those born in the 1970s and those born in the 1980s, and found that females lead in the use of *be like* across all three dialects. However, for those born in the 1980s, females were still in the lead for CanE and NZE, but in Perth (AusE) men weakly favour and the result is not significant. This seems to suggest that the gender effect is weakening in AusE as time goes by.

In a study of more recent data for CanE and EngE (Toronto and York), Gardner *et al.* (2020:309) found that the sex effect in favour of women was significant, meaning that women are still leading

in the spread of *be like* in CanE. In York, "women are also in the lead but "the sex effect is stronger among younger speakers compared to older speakers…" (*Ibid*.:308). Buchstaller and D'Arcy (2009:309) stated that such mixed results for gender suggest that "the effect of speaker sex is not stable across space" and that "…even within EngE there is evidence that the effect of speaker sex is variable from one locale to the next (*Ibid*.:325, n10). This indicates that there could be differing attitudes in the various speech communities about *be like*. For instance, Tagliamonte and D'Arcy (2004:508) stated that Tagliamonte and Hudson's (1999:167) hypothesis that "the further *be like* diffuses, 'the more likely it is to differentiate male and female speech'" is "dependent on the social evaluation of *be like*." For WSAE, Horn (2022:87) found that female speakers used it more than male speakers.

2.7.4.5. Social class

Previous studies of various English varieties have illustrated that *be like* originated in the middle class and found that the middle class use it the most (e.g. Tannen, 1986:324, Ferrara & Bell, 1995:285; Macauley, 2001:11). Baird (2001:10), on the other hand, did not find a significant result between the social groups "professional" and "non-professional" in NZE, Canterbury, for *be like* and therefore social class did not influence its use in her data. In Buchstaller and D'Arcy's study (2009:309), comparing AmE (recorded, 1988–1992), EngE (1994–1995) and NZE (1994–1996) to each other, they found no significant result for social class across all three dialects. For young SAE speakers from Cape Town, Horn (2022:80), considering all racial groups together in terms of social class, found that the highest *be like* users in SAE are those from the middle class (speakers from former model-C schools in her analysis). However, the WSAE speakers all belonged to Ex-model C schools, thus, social class for *be like* in WSAE could not be tested (Horn, 2022:81). Cukor-Avila (2012:622; see also Ferrara & Bell, 1995:285) suggested that "*be like* is diffusing hierarchically from urban to rural" areas.

2.7.4.6. Age

Studies investigating the age distribution of *be like* have been more or less unanimous in defining it as being most frequent from late teens (16+) to about 29 on average in various English dialects (e.g. Blyth *et al.*, 1990:219; Ferrara & Bell, 1995:274; Daily O'Cain, 2000:66; Tagliamonte & D'Arcy, 2004:502; Durham *et al.*, 2012:331; Denis *et al.*, 2019:57), and that it is characteristic of young people's speech. Labov's study (2018:6) of his Philadelphia Neighborhood Corpus (PNC) (recorded, 1973–1990 & 1992–2012), showed that it is young adults (20–39) who presented the first low frequency use of *be like* in the late 1970s, followed by adolescents (13–19) who increased its frequency in the PNC. He also reports that no other age group shows a significant increase in the use of *be like* throughout the years of the corpus except for five people from the 40–59 group who reached the 50% level in the last few years of the recordings. Other studies have indicated

increased frequencies of *be like* with people in their 30s over the years, suggesting real change in progress (e.g. Singler, 2001; Tagliamonte & D'Arcy, 2004; Buchstaller, 2015).

2.7.5 Previous studies: DM-*like*

The following reviews previous studies of DM-*like* and their main findings. DM-*like* has been correlated with the social factors of age, gender and social class, as well as the internal factor of clause position.

2.7.5.1. *Like* in British English (BrE & EngE)

In one of the early sociolinguistic studies, Andersen (1996) investigated DM-*like* in the Bergen Corpus of London Teenage Speech (COLT; recorded, 1993). Thus, the variety investigated was England English (EngE). Andersen's (1996:6) sample consisted of 13–17-year-old adolescents from different socioeconomic backgrounds. His (1996:62) results indicated that DM-*like* is more frequently used among speakers from the higher social classes. For instance, Andersen (1996:62) reported that groups 1–3 (middle to upper class) obtained 50% for *like* while groups 4–5 (working class) obtained 25%. Andersen (1996:62) concludes that in EngE, DM-*like* is used mostly by teenage speakers from the middle to upper middle classes. Furthermore, his results indicated that the adolescent EngE speakers used clause-medial *like* 45% of the time while clause-initial *like* was used 46% of the time (*Ibid*.:63, n.2).

In a follow-up study by Andersen (2001:224–225), he compared the COLT data as a sample of adolescent EngE speech to a sample of the British National Corpus (BNC), representing adult speech, and found that the COLT contained a higher frequency of DM-*like* (2.65/1,000 words) than the BNC corpus (0.43/1,000 words). This result indicates that in EngE, it is predominantly a marker of adolescent speech. In addition, Andersen (2001:225) observes that 83% of the discourse marker in the BNC were produced by speakers who were 41 and younger. According to Andersen (2001:225), this result indicates that the discourse marker has been adopted to some extent by older speakers, but very minimally by speakers over 45. In addition, the BNC data only had 16 instances of clause-final *like*, all spoken by speakers who were 47 and older, indicating that this is an infrequent variant of *like* in EngE (Andersen, 2001:225). Regarding the structural position of *like*, Andersen (2001:274) observed that clause-medial position was the most frequent in EngE (42.8%), followed by clause-initial position (39.7%) and clause-final *like* was the least frequent (17.4%).

In terms of the social factors, Andersen (2001:287) tested the distribution of *like* across gender and states that the adolescent girls use it the most vs the boys. Andersen (2001:288) found that in terms of age, the use of *like* significantly decreases after the age of 20 and it is the most frequent among young people in their late teens (14–16-year-olds: 64.6%). Andersen (2001:289) also

tested for social class in the COLT data and found that, where his (1996:62) previous study only investigated a selection of the COLT, the results for the entire COLT data indicated that upper-middle class speakers obtained the highest frequency of DM-*like* use (3.35/1,000 words) while the middle and working classes obtained similar frequencies (2.80 & 2.89/1,000 words). However, the result for statistical testing was not significant, indicating that social class is not a significant factor for *like* use in EngE. One of Andersen's (2001:294) conclusions is that *like* is grammaticalising in London teenage speech and is in an early stage of grammaticalisation. The evidence he gives for this is that quotative *be like* is still rare at the time of the COLT recordings and that *be like* is a later development of pragmatic *like*.

Levey (2006:416) conducted a study of DM-*like* on more recent recordings of adolescent speech in EngE (recorded, 2000–2004). These speakers were between 7–8 and 10–11-year-olds. Levey's (2006:427) distributional results indicated that *like* was the most frequent among the 10–11-year-old girls. He states that "an increment in the frequency of an expression is considered to be one of the hallmarks of a grammaticalising change". Thus, he states that the higher use in the 10–11-year-old group vs the 7–8-year-old group is evidence of the ongoing grammaticalisation of *like* in the speech of the youngest speakers. In terms of gender, his results indicated that, while the 10–11-year-old girls used a higher frequency of *like* than the boys, the 7–8-year-old boys used it more than their female counterparts. However, there was no gender distinction in the overall distribution of *like* (*Ibid*.:429).

In terms of the syntactic pattern, Levey's (2006:431) results indicated that clause-medial position was used the most (60%), while clause-initial position was used 40% of the time. There was only one instance of clause-final *like* in his data. Levey (2006:431) states that in teenage EngE speech, clause-final *like* is becoming rare, and preadolescents and adolescents show a convergence in the syntactic pattern of *like* where it is similar for both age groups. Regarding the social distribution of DM-*like*'s syntactic position, Levey (2006:432) found that the girls used clause-initial *lik*e more than the boys (45% vs 34.5%), while the boys used clause-medial *like* more than the girls (65.4% vs 54.5%)[18]. According to Levey (2006:432), his data indicates that clause-initial *like* increases with age and that the female lead is more pronounced in the 10–11 age group compared to the 7–8 age group. Levey (2006:435) attributes the male preference for clause-medial *like* and the female preference for clause-initial *like* to differences in discourse strategies between the two genders. Citing a study by Erman (1992; in Levey, 2006:435), he states that women tend to use pragmatic expressions to connect arguments while men tend to use such expressions as "attention-drawing" devices or to "accomplish repair work" (*Ibid*.:435).

[18] Levey (2006:431) separated clause-initial *like* in full sentences from *like* preceding aborted sentences and false starts. His results presented here are percentages calculated from the totals of these two categories combined, since the present study coded *like* before false starts etc. as clause-initial *like*.

Beeching (2016:28–31), who investigated the BNC and the *UWE Role-play Corpus*[19] (EngE; recorded, 2010–2014; consisting of the speech of undergraduate students), found that the BNC contained much fewer instances of DM-*like* than the more recent UWE corpus. Beeching (2016:138) found that the BNC showed no statistically significant difference between the social classes regarding the use of DM-*like*, but that women use it the most (*Ibid*.:139–140). Her results for age indicated that *like* usage rose from the 0–14 age group up to the 15–24 age group, where it obtained the highest frequency (74.66%), after which it declined incrementally in each subsequent age group until it reached 39.2% in the oldest age group (60+). Within the 15–24 age group, Beeching (2016:143) found that men used clause-final *like* more than women. She states that it seems to be a "marker of masculinity" (*Ibid*.:143). Furthermore, her results indicated that young, female, middle-class speakers are leading in the use of clause-medial *like*, while men prefer the clause-final variant (*Ibid*.:144).

Schweinberger (2015:125, BrE)[20] investigated the conversational sub-corpus of the *International Corpus of English*, British variety (ICE-GB, recorded 1990–1993; The International Corpus of English, 2021) and states that this corpus is biased towards educated speakers, but that it does consist of speakers from other social classes. Schweinberger (2015:237) notes that, compared to Andersen's (2001) study of EngE (discussed above), DM-*like* is infrequent in the ICE-GB corpus (0.490/1,000 words). He suggests that the corpus might represent an earlier stage in the diffusion of *like*. In addition, he states that this corroborates Andersen's (2001:289) hypothesis that *like* is infrequent among adults. Schweinberger's (2015:239) results also showed that clause-medial *like* is used the most in EngE, while clause-final *like* is rare. His results indicated "a moderate but distinct decrease in Like use with increasing age" (Ibid.:2015:239), stating that this corroborates Andersen's (2001:289) finding that *like* use decreases with age. It also indicated that there is a small gender difference in which women used more DM-*like* than men in the 45 years and younger group, while men were in the lead among the older speakers (46 and older). Schweinberger's (2015:242) statistical results indicated that the gender difference for the entire speech community was weak but significant, and that female speakers were in the lead. However, there were no significant age differences in the use of *like*. Schweinberger (2015:242) concludes that at the time of data collection, *like* had not yet fully entered the speech of EngE for speakers over the age of 30.

Herat (2018:96) investigated *like* use in Liverpool. All of the speakers were taxi drivers. Herat (2018:96) found a similar result to Beeching (2016) and Schweinberger (2015) for gender distribution, where women used *like* more than men. Younger women (aged 16–25) obtained the

[19] This corpus does not comprise sociolinguistic interviews; it consists of enactments where students were asked to "role-play on a discursive/argumentative topic which lasted for 3 minutes" (Beeching, 2016:30).
[20] Although Schweinberger refers to EngE in this study, the ICE-GB not only includes speakers from England, but also from Scotland and Wales (International Corpus of English, 2021).

highest frequency. However, clause-final *like* was used the most by men aged 16–25 compared with their female counterparts (*Ibid*.:100). The structural distribution of *like* indicated that clause-medial *like* was used the most (N = 241), while clause-final *like* (N = 25) was behind quotative *be like* (N = 30) and clause-initial *like* was used the least (N = 7) (*Ibid*.:97).

2.7.5.2. *Like* in American English (AmE)

Where American English is concerned, Daily O'Cain (2000:64) conducted an investigation of DM-*like* from 30 sociolinguistic interviews of speakers from Michigan, aged 14–69 (recorded in 1995). Her results showed that, similar to the EngE studies discussed above, younger speakers used it more than older speakers and *like* use decreased with increasing age. Another result was that the distribution indicated a weak lead for men. However, the gender difference was statistically insignificant (*Ibid*.:66) while the age difference was significant. In a study by Siegel (2002:35), interviews of high school students in Philadelphia were analysed (no recording date was specified) and the results indicated that the girls used DM-*like* more than the boys (*Ibid*.:44).

Schweinberger (2015:125) also studied DM-*like* in AmE and analysed the *Santa Barbara Corpus of Spoken American English*[21] (recorded, 2000) which represents a "wide variety of people of different regional origins, ages, occupations, and ethnic and social backgrounds" (Department of Linguistics, UC, 2024). Schweinberger found that the higher rate of *like* use compared to other ICE-corpora (such as the EngE sample discussed above) lends support to Andersen's (2001:221–224, 286; cited in Schweinberger, 2015:195) hypothesis that it is an American borrowing which then spread to other varieties of English. For the overall structural distribution of *like*, Schweinberger (2015:196) found that the most frequent position is clause-medial and that it is substantially more frequent than the clause-initial and clause-final variants. Furthermore, Schweinberger (2015:197) found that the distributional results indicated that women use it more than men in every age group except the oldest one. It also indicated a decrease in the use of *like* with increasing age in female speech. In addition, his results indicated that the younger speakers used it much more than the older speakers. Schweinberger's (2015:198) results for social class distribution indicated that working-class speakers used it the most (2.88/1,000 words) while the middle and upper middle classes obtained similar frequencies (2.23/1,000 words & 2.16/1,000 words). However, in statistical testing, Schweinberger (2015:200) found that the gender and social class differences were not significant, while age was the only significant factor.

For the structural analysis of *like* in AmE, Schweinberger (2015:202) states that the results show age-grading for clause-initial *like*, where it decreases with increasing age and females in their 20s use it the most. The statistical results for clause-initial *like* indicated that none of the social factors

[21] This is the main data for the ICE-AmE corpus (Schweinberger, 2015:125).

(age, gender or social class) were significant (Schweinberger, 2015:202). Furthermore, the results indicated that age was only significant among the female speakers and there was a weak significant gender difference among speakers in their 20s, indicating that women in their 20s use it more than their male peers (*Ibid*.:204).

The results for clause-medial *like* revealed that it is highly stratified by age, where it is used the most in the 21–30 age group (2.11 mean value). It decreases in the 31–40 age group (1.63) and decreases further from 41+, where it obtained the lowest mean value (0.51) (Schweinberger, 2015:205). Schweinberger (2015:205) states that no gender difference in the use of clause-medial *like* was observed. Statistical results for clause-medial *like* indicated that there were no significant gender or social class difference and that age was the only significant factor, which indicated that younger speakers use it more than older speakers (*Ibid*.:207). Furthermore, Schweinberger (2015:207) noted that there was one instance of clause-final *like* in the AmE data.

2.7.5.3. *Like* in Canadian English (CanE)

For Canadian English (CanE), Tagliamonte (2005:1896) investigated a corpus of conversational data consisting of speakers aged 10–19 (recorded, 2000–2003). The results for DM-*like* indicated that it was highly frequent in CanE teenage speech, occurring 9,739/200,000 words (Tagliamonte, 2005:1899, 1900). Tagliamonte (2005:1900) noted that for some speakers, *like* was more frequent than the word *and*. The results for structural position indicated that in CanE, it is also clause-medial *like* that is the most frequent, followed by clause-initial position (*Ibid*.:1902). With regards to the distribution of social factors, Tagliamonte's (2005:1903) results indicated that female speakers use it the most in each age group (10–12 yrs; 13–14 yrs; 15–16 yrs; and 17–19 yrs). In addition, her results showed that the 15–16-year-olds using *like* the most, while the frequency for the other age groups were much lower. The 17–19-year-olds used it the least (*Ibid*.:1903–1904). Tagliamonte (2005:1904) states that the drop in age indicates "the classic pattern for age-grading". In addition, she states that, because gender differences are the strongest among the older teens and weakest in the youngest age group, the results suggest that gender differentiation for DM-*like* is developmental and is thus created in the speech community (*Ibid*.:1912–1913).

In another study for CanE, D'Arcy (2005:25) investigated a sample of the Toronto English Archive (recorded, 2002–2004). D'Arcy's results indicated that clause-medial *like* is used more frequently than clause-initial *like* (2005:209). D'Arcy's (2005:85–87, 106, 110, 113, 212) results also indicated that clause-initial and clause-medial position have increased in frequency in apparent-time. D'Arcy (2005:212) states that this rise in frequency across apparent-time for clause-initial *like* strongly suggests that it is undergoing a generational change. The results illustrate that both clause-initial and clause-medial *like* are used most frequently in the 15–16 age group, but with

high frequencies among the 19–24 age group as well (D'Arcy, 2005:212). D'Arcy (2005:212) states that the patterns for clause-initial and clause-medial *like* for the older vs younger speakers are the same: both age groups use it in clause-initial and clause-medial position. It is merely that the younger speakers use it more frequently in these contexts.

For the gender distribution across structural position, D'Arcy (2005:222–223) observed that, across apparent-time, women use clause-initial *like* more than men (although the margin is narrow). On the other hand, men use clause-medial *like* more than women. Levey (2006:435) obtained a similar gender distribution for EngE. D'Arcy's (20015:222–223) results corroborate his statement that there is a functional distinction in the way men and women use pragmatic expressions, (discussed previously in this section) which is also evident in CanE.

2.7.5.4. *Like* in New Zealand English (NZE)

Schweinberger (2015:310) investigated *like* in New Zealand English (NZE) with data from the ICE-NZE component (recorded, 1999; ICE, 2023). The overall results indicated that DM-*like* is frequent in NZE, obtaining an average frequency of 2.175/1,000 words. In terms of structural distribution, Schweinberger (2015:312) found that, contrary to the other varieties discussed so far, it is not clause-medial *lik*e that is the most frequent but clause-initial *like*. However, the difference is minimal. Gabrys (2017:58) also investigated *like* in the ICE-NZE corpus and found that clause-medial *like* was more frequent (55%) than clause-initial *like* (31%). This difference in results is likely due to the fact that Gabrys (2017:58) did not separate instances of *like* before restarts[22] from clause-initial *like* in complete sentences). Schweinberger (2015:324) also separates clause-initial instances from what he terms "non-clausal *like*".

Schweinberger (2015:312, 313) found that DM-*like* in NZE decreases significantly with age, which suggests that it is age-graded. For gender, the distributional results indicated that young adult women (below age 30) lead in its use, while men lead in the older group. However, Schweinberger (2015:312) states that the gender differences are small.

For social class distribution, Schweinberger (2015:314) found that working-class speakers used it much more (3.26/1,000 words) than the middle (1.97/1,000 words) and upper middle (1.51/1,000 words) classes. Schweinberger (2015:314) found that speakers with professional and academic careers use it the least. Schweinberger (2015:315) found that no other factor than age was significant for *like* use in NZE: once again, *like* use decreases as age increases. For social factors correlated with the structural position of DM-*like*, Schweinberger (2015:316) found that clause-initial *like* is age-graded and there is a gender effect in which females lead, but only for post-adolescents younger than 20. The results for clause-medial *like* were similar: clause-medial

[22] Which Schourup (1982:39) considers clause-initial examples.

like is also age-graded since it decreases with increasing age. In addition, a gender effect was observed for speakers between 30–39 years of age in which, contrary to the result for clause-initial position, men obtained the higher frequency. These results obtained significance for age and gender in the contexts described above (*Ibid*.320–321). Social class had a significant effect on the use of clause-medial *like*, where it was most frequent among the middle class. Schweinberger (2015:323) states that clause-final *like* was infrequent in the NZE data, but that it also indicated an age-graded effect where younger speakers used it more than older speakers. However, no other social factor was significant (*Ibid*.:324). Schweinberger (2015:326) concludes that the results for NZE suggest that any changes *like* may have gone through are nearing the end, such as the fact that clause-initial *like* did not obtain a social class difference. However, he states that because clause-medial *like* obtained a gender effect towards males and a social class effect towards middle-class speakers, changes in DM-*like* has not spread to all areas of the community.

2.7.5.5. *Like* in Irish English (IrE)

To investigate DM-*like* in Irish English, Scwheinberger (2013:5) analysed the *Northern Irish Transcribed Corpus of Speech* (NITCS) (recorded, early 1970s). The results indicated that in IrE, speakers used *like* 1.768/1,000 words. The results showed that, at the time of the recording, men obtained a higher frequency of *like* compared to women. Statistical analysis revealed that neither age nor gender had a significant effect on the use of *like* (*Ibid*.:9, 12). Although he did not look at the various structural positions of *like*, he did analyse clause-final *like* and found that it is mostly used by a few speakers in the corpus. It was thus not homogeneously spread across the community (*Ibid*.:24).

Schweinberger (2013:23) states that clause-final *like* is perceived as a traditional variant of *like* in IrE and is associated mostly with older male speakers. Because of this, it has been hypothesised that it is in decline in IrE. Gender was significant: men used more clause-final *like* than women. Age was a significant factor but indicated that it was not the older speakers (ages 65–75) but the younger speakers who used clause-final *like* the most (ages 35–45). Schweinberger (2013:25) concludes that the result for age indicates that clause-final *like* was not in decline in the 1970s. Thus, it is not the process of being replaced by the American variants clause-initial and clause-medial *like* (*Ibid*.:25). In a following study, Schweinberger (2015:125) studied the ICE-Ireland corpus, which consists of speakers from Northern Ireland and the Republic of Ireland (recorded, 1990–1994 and 2002–2003; Kallen & Kirk, 2008:4–5). Compared to the NITCS (Schweinberger, 2013:5; see the frequency reported above), DM-*like*'s frequency in ICE-Ireland rose substantially 4.102/1,000 words). Schweinberger (2015:247) observes that, similar to the NITCS, DM-*like* is not distributed homogeneously but is concentrated among a small group of speakers.

Schweinberger's (2015:248) structural analysis indicated that clause-final *like* obtained the highest frequency as compared to clause-medial and clause-initial *like*. In addition, *like* overall was used more frequently by younger rather than older speakers and this result was confirmed to be significant – speakers in their mid-20s to mid-30s used it the most.

Schweinberger (2015:248–249) states that gender differences in IrE are age specific. Men obtained a higher frequency for DM-*like* in the age range of 34 years and older, however, for speakers in their late twenties and early thirties, it is women who use it more (*Ibid*.:253). Schweinberger's (2015:254) detailed analysis of the structural positions of DM-*like* revealed that, in IrE, younger speakers used clause-initial *like* more than older speakers. Gender and social class were not significant factors constraining the choice of clausal position (*Ibid*.:256).

The results for clause-medial *like* indicated that the age effect is significant and suggests that clause-medial *like* in IrE is age-graded and decreases from age 35 and upwards. However, similar to clause-initial *like*, gender and social class did not obtain a significant effect (*Ibid*.:258), except among older speakers. Women aged 50+ used it more than the men in the same age group, who almost never use it. According to Schweinberger (2015:259), this result indicates that men adopted clause-medial *like* at a later stage than women.

The results for clause-final *like* indicated that it is age-graded, but only among women for whom it decreased as age increases. However, for men, the rate of clause-final *like* remains similar in all age groups. Social class did not have a significant effect on this variant, while a gender effect was observed only for older speakers of 50+ in which men used it more than women (*Ibid*.:260–262).

In a comparison of the 1990–1994 vs the 2002–2005 data, Schweinberger (2015:265) found that while clause-initial, clause-medial and clause-final *like* increased in frequency in IrE over the years in real-time, the highest frequency was observed for clause-medial and clause-final *like* among 26–33-year-old speakers. He states that this indicates ongoing change for these variants in IrE. The results also indicate that clause-final *like* is not disappearing in IrE as previously thought. However, its frequency did decline with speakers from age 26 and younger. The results indicated that the use of DM-*like* increased among younger speakers (Schweinberger, 2015:266–267). Schweinberger (2015:269) states that gender differences in the use of DM-*like* in IrE have decreased and clause-final *like* is losing its status as a stigmatised older male variant in IrE. In addition, he states that once clause-medial *like* became more frequent among female speakers, it decreased among male speakers (*Ibid*.:270). In addition, because clause-initial and clause-medial *like* are more frequent among women in the age range of 36–33, it indicates that DM-*like* is not mainly an adolescent feature in IrE, compared to the findings for age in the other dialects discussed above (*Ibid*.:270).

In contrast to Schweinberger's (2015) results for clause-final *like* (discussed above), Corrigan (2015:57; data recorded, 2008) found that it was young women who used this variant the most in Northern Ireland English. Murphy (2015:71–72) investigated the distribution of DM-*like* and another pragmatic marker, *sure*, in the Republic of Ireland (data recorded, 2003–2012). She found that DM-*like* was the most frequent out of all the pragmatic markers – it occurred almost twice as frequently as *sure* (*Ibid*.:74). Murphy (2015:75) also found that DM-*like* was used more by speakers in their 20s and that its frequency recedes by the age of 40 for female speakers but not for males. Her results also indicated that, among the speakers in their 20s, no significant gender difference was observed (*Ibid*.:75). Murphy (2015:84) concludes that *like* in IrE seems to be a marker of young speakers and its use is mostly influenced by age.

2.7.5.6. *Like* in White South African English (WSAE)

The only study that I am aware of that investigated DM-*like* in White South African English is an M.A. study by Lochner (2019:26), who interviewed 17 participants consisting of speakers ranging in age from 12–21, residing in the Western Cape. Her results indicated that clause-medial *like* occurred most frequently in the data (before noun phrases (44%) and before verb phrases (14%)), while clause-initial *like* occurred 11% of the time (*Ibid*.:37). The results of the social distribution of DM-*like* in Lochner's (2019:38) study indicated that the most frequent users were females aged 15–18 (over 25%) as compared to the 19–21 age group (about 14%, no specific percentage was given). This result indicated an adolescent peak for WSAE, similar to Tagliamonte's (2005:1902) results for CanE. In addition, *like* was used more by women (63%) than men (37%) and the gender difference was more pronounced in the 19–21 age group. The most frequent users were the 15–18 age group, who obtained near equal frequencies between women and men (*Ibid*.:38). Lochner (2019:45) concluded that the drop in frequency of DM-*like* in the 19–21 age group (compared to the 15–18 age group) suggests age-grading and that it is not undergoing linguistic change in WSAE.

2.8 Conclusion

In conclusion, while Chapter 1, section 1.1 discussed the historical development of pragmatic *like* from traditional, dictionary type uses to its development as a hedge, discourse marker and finally quotative in the form of *be* + *like*. Section 2.7.1 in this chapter discussed the grammaticalisation path of *like*. In section 2.7.3, it was noted that older speakers tend to reflect older patterns of the quotative system in English and thus tend to use *say* more than *be like*. In earlier studies of the quotative system, it was found that *be like* mainly functioned as a quotative of thought reporting, while *say* was used to report speech. However, in later studies (2.7.4) it was found that *be like* expanded to direct-speech reporting and thus started to compete with the functional territory of *say* to such an extent that it overtook *say* as the main quotative in young people's speech. The

expansion of *be like* to direct-speech reporting is considered a later stage of its diffusion, as well as an advanced stage of its grammaticalisation, where its specific pragmatic function as an introducer of thought has been bleached[23]. Regarding grammatical person, *be like* was first found mainly in first-person contexts, while *say* was found mainly in third-person contexts. However, Ferrara and Bell (1995:279) found that in AmE, *be like* was expanding to third-person contexts over time, thus, it once again started competing with a context found frequently for *say*. However, in other dialects such as CanE, EngE, AusE and NZE, even though *be like* increased with third-person contexts, its association with first-person contexts remained constant. Thus, the conclusion was that it was not expanding to third-person contexts in the above-mentioned dialects. However, in other dialects such as Scottish English, it was found to be more frequent in third-person rather than first-person contexts, leading to the conclusion that it either has a later stage of grammaticalisation in such dialects or that it developed locally specific constraints. Regarding *be like*'s association with tense, it was first found mainly in the present tense but later expanded into use in the the past tense. Its expansion into past tense is an indication of a later stage of diffusion and grammatical development.

In sections 2.7.5 and 2.7.4.4, the correlation of *be like* and DM-*like* with gender was discussed and showed that results have been mixed. Two viewpoints have emerged regarding *be like*'s association with gender: for AmE, it has been hypothesised that *be like* correlates with women at an early stage of diffusion and loses its gender-marking at a later stage (Ferrara & Bell, 1995:274). On the other hand, for CanE and EngE, Tagliamonte and Hudson (1999:167) viewed the female lead as a later stage in *be like*'s diffusion and grammaticalisation, since the gender effect in favour of females increased over time in these dialects (e.g. Tagliamonte & D'Arcy, 2004:508; Gardner *et al.*, 2020:309). DM-*like* has been correlated with female speakers in EngE, CanE and NZE, similar to *be like* in these dialects. However, gender differences have not been found to be significant in AmE, suggesting a later stage of diffusion. In IrE, DM-*like* has been correlated with men in one study and with women in another study, largely due to the men's frequent use of clause-final *like.* However, other studies of IrE found no significant correlation with DM-*like* and gender. In the few studies of WSAE, female speakers were found to be the most frequent users of DM-*like* and quotative *be like*.

Regarding the structural position of DM-*like*, it was correlated with females in clause-initial position and with males in clause-medial position in some studies (e.g. D'Arcy, 2005:25, CanE; Levey, 2006:432, EngE), while in other studies there were no differences between the genders (e.g. Schweinberger, 2015:202, AmE).

[23] I.e. it no longer only introduces thought but direct speech as well.

Results for *be like*'s correlation with social class have largely been insignificant, indicating that social class is not a strong factor influencing its use. In earlier studies, it was often found to correlate with the middle and higher social classes and Ferrara and Bell (1995:285) suggested that it was expanding to lower-class groups. Thus, an insignificant result for social class could mean that *be like* has reached a later stage of diffusion. In previous M.A. studies for WSAE, DM-*like* was correlated with the middle class, while quotative *be like* was not tested for social class.

The correlation of *be like* with age has largely been that it is found most frequently with younger speakers up to their mid-30s and similar results have been found for DM-*like*. However, for *be like*, it has been suggested that it is undergoing a real change in progress since it is becoming more frequent with older speakers, while studies for DM-*like* have largely found that it is an age-graded phenomenon.

CHAPTER 3: METHODOLOGY

This chapter discusses the methodological reasoning behind the selected participant criteria, including race, gender, age, class, and the specific variety of South African English (WSAE). It then describes the participant-selection process, interview method, the issue of speech style, and ethical considerations. Finally, the chapter discusses the methods used in collecting, analysing and interpreting the speech data.

3.1 Sampling Method

In early sociolinguistic studies such as Labov's (1966) investigation of New York City and Philadelphia (1973–1977; in Labov, 2001:35); Trudgill's (1974) study of Norwich, Cedergren's study of Panama City (1973; in Labov, 2001:225) and Sankoff and Sankoff's study of Montreal French (1973; in Labov, 2001:225), the method of random sampling was used to create their corpora in order to have a representative sample of a speech community (Labov, 2001:50). Schilling (2013:31) defines random sampling as the method in which a researcher selects study participants in such a manner that every member of a population has an equal chance to be selected. She states that a large enough random sample would be statistically representative of the community and it would then be possible to generalise results to the entire speech community. An example of random sampling is the situation in which a computer picks participants randomly from a list such as a telephone or electoral list (Schilling, 2013:31). For instance, in Labov's (1966a:101) study of the Lower East Side of NYC, he randomly selected working-class participants from the list of a sociological survey of New York, the *Mobilization for Youth Survey* (MFY).

Milroy and Gordon (2003:24) state that the main requirement in selecting a representative sample of a community is to avoid bias. For instance, Schilling (2013:31–31) states that telephone directories are biased towards those who have landlines and do not include those who do not have one or those who did not want to be listed. Milroy and Gordon (2003:26) state that linguists today rarely ever adopt the method of true random sampling since the cost and effort of this method is high. In addition, they mention that it is not clear whether strict representativeness of a sample would necessarily provide better insights into sociolinguistic structure. Schilling (2013:33) states that sociolinguists have discovered over the years that people's speech behaviour is more homogeneous than other types of behaviour usually studied in sociology, such as people's purchasing preferences. For instance, Sankoff (1980:51–52; in Milroy & Gordon, 2003:28) notes that linguistic behaviour is more homogenous since people "understand each other with a high degree of efficiency". The need for effective communication results in a regularity not found in other types of social behaviour. Thus, Milroy and Gordon (2003:26) note that at present,

sociolinguists rarely employ random sampling methods but rather opt for the method of "judgement sampling".

Milroy and Gordon (2003:28) state that sociolinguistic surveys are usually much smaller than other types, such as sociological surveys. For example, they explain that Neuman[24] (1997:222; in Milroy & Gordon, 2003:28) suggests sample sizes of 300 for a population of 1,000 and a sample size of 1,500 for a population of more than 150,000. The sample size of linguistic investigations is much smaller than this, which is why Milroy and Gordon (2003:26) view the sample achieved through sociolinguistic methods as a weak type of representativeness. In fact, Sankoff and Laberge (1980:51–52; in Milroy & Gordon, 2003:28) state that the data of more than 150 participants in a linguistic sample tend to become redundant with reduced analytical benefit.

According to Schilling (2013:35), another type of sampling method is more suitable to sociolinguistic study, especially when the investigator needs to have participants from a wide range of social categories, such as different social classes, ages and genders. A true random sample might not produce enough speakers from each group under consideration. The method that is most used by linguists is that of "quota" or "judgement" sampling. Schilling (2013:35) elucidates that a judgement sample involves deciding in advance which types of speakers will be needed for the study. For example, "older, middle-aged, and younger speakers; males and females; African Americans and Whites" (*Ibid*.:35). Next, the researcher uses their judgement of participants' suitability[25] to fill the required "cells". Milroy and Gordon (2003:32) also mention the "snowball" technique, which is described as a technique that "utilizes the social networks of participants in the study to recruit potential new participants" (also known as "network sampling"; *Ibid*.:32). They clarify that this technique involves asking an interviewee to mention people from their own social network who would possibly be willing to participate. They also state that the main advantage of this technique is that it reduces the rate at which possible participants refuse to be interviewed (*Ibid*.:32).

3.2 Recruitment of Participants

The method chosen to build the speech sample for this study was judgement sampling, since it is more suitable for the investigation of sociolinguistic variables, as well as being more fitting for the small scope of the current project. In achieving a judgement sample, Milroy and Milroy's (1978; in Milroy, 1987:43) "friend of a friend" method for gaining access to a speech community was employed. Participants were initially recruited by contacting some family members and friends of mine living in Port Elizabeth and asking them to find willing participants among their friends and acquaintances. After the first few interviewees were recruited through the friend-of-a-friend

[24] A source used as a general guideline in the social sciences.
[25] Or the extent to which they are representative of a pre-determined social group.

method, the snowball technique was employed with which it was possible to extend the reach to the family and friends and/or acquaintances of the interviewees. This was done by asking interviewees for possible participants. Another technique that was used to recruit participants was to post a flyer (created by myself) on Facebook in all relevant Port Elizabeth community groups. The flyer was successfully posted in each group where permission was granted by the page admins. Public Port Elizabeth Facebook groups did not need permission to post the flyer. This flyer can be found in Appendix B.

The method of posting a recruitment flyer to Facebook had moderate success but was not as successful as the friend-of-a-friend and snowball method. However, it did aid with diversifying the social profile of participants, since a few of the lower middle-class speakers, all the older speakers, as well as some of the men, were obtained through the Facebook posts. Even though recruiting participants through Facebook contains the bias that only Facebook users are reached, the combination of all the techniques described above resulted in a relatively diverse judgement sample.

3.3 Sample Design

The initial sample design is displayed in Table 3.1 below.

	Working class		**Middle class**	
Age	Male	Female	Male	Female
18–29	3	3	3	3
30–40	3	3	3	3
Subtotal	6	6	6	6
Grand total	**24**			

Table 3.1: Initial sample design

The planned design was to stratify the sample according to social class, comprising working and middle class; gender, comprising male and female; and two age groups comprising age ranges of 18–29 and 30–40. The aim was to have three participants for each cell, as indicated in Table 3.1. However, at the end of the interview process, only 19 participants could be recruited. In addition, after each speaker was assigned a social class, there were no working-class participants, thus, the social class distribution was redefined in order to stratify the different levels of middle-class speakers. The cut off age for participating was increased from 40 to 59 in order to potentially find more older speakers as well. However, the oldest speaker that could be obtained was 48. The final sample design shows the age range of the younger and older participants that were interviewed. Table 3.2 presents the redefined stratified sample and the number of participants that could be obtained for each cell. As illustrated in Table 3.2, there are no lower-middle class (LM) female speakers in the 33–48 age range and no upper-middle class

(UM) male speakers in the same age range. Furthermore, only one participant each was obtained for the 18–20 male, lower-middle class and middle-class (MM) group. There was only one participant each for the MM female group in the 33–48 age range as well. As is apparent from the table, the most well represented social groups in the sample are the MM female group in the 18–20 age range, followed by the UM, 18–20 group.

	Lower middle class		**Middle class**		**Upper middle class**	
Age	Male	Female	Male	Female	Male	Female
18–20	1	2	1	5	3	2
33–48	2	0	1	1	0	1
Subtotal	3	2	2	6	3	3
Grand total	**19**					

Table 3.2: Final sample design

3.4 The Social Class Index

For discovering the social class of participants, Trudgill's (1974:35) social class index from his study of Norwich was used. According to Trudgill (1974:32), the stratification of social classes is mainly based on the differences in people's wealth and income. In addition, the social classes are not sharply differentiated groups but rather collections of people with "similar economic characteristics" (*Ibid*.:32) that form a continuum (1974:33). Furthermore, Trudgill (1974:32) states that such social differentiation is as effective in creating linguistic boundaries between groups of people as that of geographical boundaries, which decelerate the diffusion of speech forms originating in a specific social group from one community to the next. The definition of social class and its function within sociolinguistic studies is discussed more extensively in Chapter 2, sectionsDefinition and Background to the Concept of Social Class2.3 and 2.4.

The social class indicators used by Trudgill (1974:36) for the Norwich study were "occupation; income; education; housing; locality; and father's occupation". Trudgill (1974:36) states that such a multi-indicator index is a more reliable and refined measure of social class compared to a one-indicator index such as using only occupation or only education. However, Trudgill (1974:36–37) explains that in any modern industrialised society, occupation is usually the strongest indicator of social class, and the two indicators of income and education are closely related to it. In addition, according to Trudgill (1974:37), housing and locality have some influence on social status, though not as strongly as the other three. Labov (1966a:133) states that no single socio-economic index will be applicable to every society and that a social index needs to be curated for the specific social structure of the society under investigation. Thus, for this study, Trudgill's (1974:38) social indices were adapted according to sociological information about the social structure of South African society. These indices were also simplified according to availability of information about

the interview participants. Trudgill's indices and how they were adapted are described in sections 3.4.1–3.4.4.

3.4.1 Occupational indicator

To establish a six-point scale of occupational ranking, Trudgill (1974:38) used the "Registrar General's Classification of Occupations" (1966; in Trudgill, 1974:38), based on the 1966 Sample Census in the UK. Trudgill (1974:38) chose to rank women according to their father's occupation because, in British society at the time, women still had limited occupational opportunities. In addition, he states that the breadwinner of the family often determines the social status of the entire family. Thus, women with a "higher" occupation than both their husband and father were rated according to their own occupation (*Ibid*.:38–39). Trudgill (1974:39) also rated school children according to their father's occupation. Visagie and Posel (2013:150–152), who conducted a sociological study regarding the middle class in South Africa, state that it is better to base an individual's social class on household affluence rather than on an individual's occupation, since those who do not have an occupation, such as the unemployed, housewives, and students, would not be classifiable. In addition, those who "share the same lifestyle, resources and social status" (*Ibid*.:151) in a household would be classified into different categories if the ranking were based on the individual's occupation. Gradín (2021:102) states that the gender stratification of South Africa's labour market has changed over time – there has been an increase in female employment from 38% in 1994 to 45% in 2015. Gradín's (2021:102) study indicated that women are overrepresented in middle-class occupations, such as clerks, tellers, office workers, and in higher occupations, such as professionals and technicians (e.g. teachers and nurses). Furthermore, Gradín (2021:102) states that white South African women have a higher representation as professionals compared to white men, which "helped to reduce the gender wage gap" (*Ibid*.:111).

Thus, based on the topics discussed above, the decision was made to award points for the interviewees' occupational index on the basis of the breadwinner of their household, whether male or female. University students and people who were unemployed at the time of the interview were ranked according to the household breadwinner's occupation. Men and women who were employed were ranked according to their own occupation, unless their spouse had a higher occupation. Young adults who moved out of their parents' households were ranked according to their own occupation, unless they were unemployed, in which case they were ranked according to their most recent occupation before unemployment. To create the social class index for this study, the occupational ranking of Visagie and Posel (2013:157) was used as a guide, as illustrated in Table Table 3.3.

In addition to Table 3.3, the *Occupational Classification list of Statistics South Africa* (n.d) was used in order to classify the participants' occupations as "professional", "technical" and so forth. For example, under the "professional" heading, a wide variety of occupations are listed, as illustrated in Table 3.4.

Upper middle class
Legislators, senior officials and managers
Professionals
Lower middle class
Technicians and associate professionals
Clerks
Working class
Service, shop and market sales workers
Skilled agricultural and fishery workers
Craft and related trade workers
Plant and machinery operators and assemblers
Elementary occupations

Table 3.3: Social class ranking in South Africa (Visagie & Posel, 2013:157)

Professionals
physical, mathematical and engineering science
computer professionals
architects
civil engineers
electrical engineers
pharmacologists
medical doctors
nursing
educational professionals
artists
religious professionals

Table 3.4: Occupational classification (STATS S.A., n.d.)

Trudgill's (1974:38) index scores for the six groups from the *Registrar General's Occupational Classification* (1966) were then applied to the occupational descriptions discussed above. This is illustrated in Table 3.5.

	Index score
Group I	5
Group II	4
Group III	3
Group IV	2
Group V	1
Group VI	0

Table 3.5: Index of occupation[26]

3.4.2 Educational indicator

Participants were also ranked according to educational level. Trudgill's (1974:40) educational index was used as the baseline but adapted for the South African context by using Visagie and Posel's (2013:164) educational groups. **Error! Reference source not found.** presents the educational index used in this study.

Index of education	
	Index score
I: Some tertiary education	5
II: Matric	4
III: Incomplete secondary	3
IV: Some primary	2
V: No schooling	1

Table 3.6: Index of education

3.4.3 School type indicator

One of the social class indicators used by Trudgill (1974:39) is income. However, since the participants were not asked for either their own or the household breadwinner's income, a different indicator was devised based on school type, since it is often the case that those with higher salaries tend to send their children to private/semi-private schools vs those who send their children to public schools (government schools). According to Grant (2013), South Africa's schools are categorised according to the quintile system, which ranks schools based on how

[26] The context for the occupational groups is the following: Group I: Professional workers; Group II: Employers and managers; Group III: Other non-manual workers; Group IV: Foremen, skilled manual workers, own account workers (self-employed); Group V: Personal service, semi-skilled and agricultural workers; Group VI: Unskilled workers, e.g. shop assistants and food servers in the fast food establishments (Trudgill, 1974:38; Oxford Dictionary of Public Health, 2018).

much government funding they should receive. He states that “quintile one is the ‘poorest’ quintile, while quintile five is the ‘least poor’” (*Ibid*.:2013). Quintiles 1–3 are schools known as “no-fee” schools, which means that students do not pay any school fees, whereas quintiles 4–5 are fee-paying schools.

In sociolinguistic studies of SAE, one of the methods often used as a social class indicator of participants is school type, by which participants from “former Model-C and private schools” (Mesthrie, 2015:392) are grouped with a higher social class from those who went to public schools (e.g. Mesthrie, 2015:392; Chevalier, 2016:75; Álvarez-Mosquera & Marín-Guitérrez, 2019:138, n.5). However, all participants in this study went to quintile 5 schools, which is the highest rank for SA schools and they are viewed as the most affluent schools (Ogbonnaya & Awuah, 2019:1). Thus, to refine the school type indicator, the ranking of schools for this study was based on the highest to lowest amount of school fees charged by these Q5 schools. The annual school fees were obtained from the information regarding fees published in each school’s website. Table 3.7 presents the school type index.

School type	Fees/annum	Index score
I: Semi-private	R60 000-70 000	5
II: Semi-private	R45 000-R50 000	4
III: Public	R36 000-R44 000	3
III: Public	R30 000-R35 000	2
IV: Public	R24 000-R29 000	1

Table 3.7: Index of school type (Q5 schools)

One caveat regarding the above ranking system is that often students from lower-income households attend quintiles 4–5 schools via bursaries (Parliamentary Monitoring Group). However, the risk of categorising a participant into a higher social class than what they actually are is somewhat mitigated by the other social indicators used in this study, such as occupation, education and house type (discussed below). For instance, two participants who attended the highest ranked schools from the index of school type in Table 3.7, Grey High School (Owen)[27] and Collegiate Girls High (Harriet), were categorised as lower-middle class speakers based on their combined scores for all the social indicators.

3.4.4 Housing

Trudgill (1974:40) states that his housing scale was based on three parameters: house ownership, age of house and house type. According to Trudgill (1974:40), regarding house ownership, a

[27] Throughout the dissertation, pseudonyms are used for participants and not their real names.

house that is owned by its occupant has a higher social status attached to it than a house that is rented.

Table 3.8 provides the index of house ownership adapted for the South African context. In addition, Trudgill (1947:41) states that normally the house type gives some information about social class status. For instance, a detached house has a higher social status than a semi-detached house; also, a semi-detached house, in turn, has a higher status than a terraced house or flat. Table 3.9 illustrates the index of house type used in this study.

House ownership	Index score
Own	2
Rent	1

Table 3.8: Index of house ownership

House type	Index score
Detached	4
Semi-detached	3
Flat	2
Garden flat	1

Table 3.9: Index of house type

As a third indicator of housing, Trudgill (1974:41) included the age of the occupant's house. However, since not all of the interviewees were asked to provide their residential addresses, it was not possible to categorise the age of each participant's residence. Seekings (2003:42) states that the living conditions of households are closely related to social class, and in his study, one of the indicators used in the evaluation of the social stratification of South African society is the number of rooms in a dwelling. Thus, instead of the age of a house, a different index was used in which the number of rooms in a house was considered, as illustrated in Table 3.10.

Rooms	Index score
9+	4
7–8	3
5–6	2
4–3	1

Table 3.10: Index of number of rooms in dwelling

The method for gathering the demographic information from participants necessary for determining their social class was in the form of questions integrated within the sociolinguistic interview schedule. An extract of the interview schedule, including some of the demographic questions, is illustrated in section 3.6.

The social class scores were calculated by adding the index scores from all the indices discussed above. For example, if a participant stated that they had 5–6 rooms in their house (see Table 3.10 above), that participant would get a score of 2 out of 4 for number of rooms in dwelling. The totals of the index scores obtained for each participant were then calculated. Once the totals were obtained, the participants were categorised into lower-middle class if they received a total score from 0–15. A participant was categorised into the middle class if their score total was from 16–18. Lastly, participants were categorised into the upper-middle class if their score total was between 19–25. The highest possible score a participant could get was 25. See **APPENDIX C**: SOCIAL CLASS CALCULATION for a table showing the social class calculations for this study. All names in the table are pseudonyms and do not represent the real names of participants.

Some social indices weighed more than others. For example, the indices with a possible score from 1–5 weighed the most, since the highest score a participant could get in each of these indices were 5 (e.g. occupation, education and school type). Social indices that weighed less were those with a possible score from 1–4 (e.g. number of rooms in dwelling). The index that weighed the least was house ownership, since it had a possible score from 1–2.

3.5 Age Range and Participant Criteria

The reason for choosing the age range of 18–59 for participant recruitment was that any age below 18 and above 59 would have required extra steps vis-à-vis ethical clearance and, since this study's focus is not specifically on how pragmatic *like* patterns across a broader age spectrum, it was not necessary to include such younger or older ages.

The technique for creating the age categories was modelled according to Labov's (2018:6) analysis that *be like* uses have high frequencies in adolescents aged 13–19 and start to reduce from 39, with the ages 40 and over having the lowest frequencies of use. Thus, after all of the interviews were conducted, participants were divided into the age groups of 18–20 and 33–48 wherever the distributional and statistical analysis considered the patterns of younger vs older speakers.

The criteria for the selection of participants were the following. Participants were restricted to those who were born in Port Elizabeth; those who were still native residents of the region at the time when the interviews were conducted; and those between the ages of 18–59. The study also included male and female speakers. Participants were further restricted to white, English first

language speakers, since this study focuses on the speech of white English-speaking South Africans. The reason for restricting participants to white speakers was to eliminate race as a variable, since that would be beyond the scope of this M-study. This is also the reason why the study focuses only on first language (L1) English speakers and does not include second language speakers.

3.5.1 Rationale for focusing on White South African English

Some considerations contributed to why WSAE was chosen as the focus for this study instead of, for example, BSAE (Black South African English), ISAE (Indian South African English) or CSAE (Coloured South African English). One reason is that, the consensus among sociolinguists is that it is ideal to have intimate knowledge of the speech community under investigation through immersing themselves in that community – this is known as "participant-observation" (Labov, 1981:28), a method taken from sociology (e.g. Eckert 1989:125–27; Milroy & Gordon 2003:68; Milroy, 1987:41; Saville-Troike, 2003:3; Tagliamonte, 2006:9; Trudgill, 2000:65). For example, Toefy (2014:35) had personal experience within the coloured community which informed her study. The researcher of this study has the most experience with members of the WSAE speech community via family members. While it is not necessary for a linguist to be a part of the speech community they are studying (see for example Mesthrie's study (2010:3) of black, Indian and coloured middle-class students concerning the GOOSE vowel), it does make it easier and less time-consuming to have previous experience with the speech group under investigation in a master's research project. Another reason is that, according to Rossouw and Van Rooy (2012:1–3), and Kruger and Van Rooy (2019:180), far fewer grammatical studies of WSAE have been done as compared to the other Southern Hemisphere Englishes (e.g. Australian English and New Zealand English). Furthermore, Rossouw and Van Rooy (2012:1–3) note that phonological variables of WSAE have been more comprehensively described compared to its grammar (loosely-defined), since past researchers (e.g. Lanham & Macdonald, 1979; Titlestad, 1996:168) believed there were not much difference between WSAE and other standard Englishes, such as British English, apart from accent.

3.6 The Sociolinguistic Interview

The main aim of a sociolinguistic interview is to elicit casual speech in order to study the vernacular of a language variety instead of the standard variety (Labov, 1972:86; Trudgill, 1974:50–53; Milroy, 1987:23). The reason for aiming to study the vernacular rather than standard speech is because, according to Labov (1981:29), the vernacular is "highly regular" and therefore "the most systematic data for linguistic analysis". Thus, Labov (1966a:87) states that "the interview is constructed around the problem of isolating contextual styles, and almost every detail of the questionnaire can be understood from that point of view". This questionnaire is known as

an “interview schedule” (Tagliamonte, 2006:38) and contains “conversational module[s]” (Labov, 1981:33) that are structured towards first eliciting general information and then moving towards more detailed information such as personal narratives (Labov, 1981:35; Tagliamonte, 2006:38). According to Labov (1966a:409; 1981:32), the purpose of the sociolinguistic interview is to obtain casual speech as well as information about the participant’s demographic details, such as age, occupation, income, family relations, rent/house values, language background, education level, and how long they have lived in the area.

According to Labov (1981:33), the conversational module is “a group of questions focused on a particular topic: i.e. children’s games, premonitions, the danger of death, aspirations, etc.” Labov (1981:34) explains that these modules form a “conversational network”. Modules have a hierarchical structure which begins with general topics and progress to more detailed ones. The interview schedule begins and ends with transitional questions that connect them to other conversational networks. Labov (1981:36) states that the conversational modules are meant to be a guide to the interviewer in order to simulate the casual conversations of everyday life as closely as possible. Since it is just a guideline, the interviewer is encouraged to follow “the subject’s main interests and ideas wherever they go” (*Ibid*.:37). The main structure of the conversational modules and the specific wording of certain questions are necessary to allow for optimal comparability with other sociolinguistic interviews conducted by different people (Labov, 1981:37). One general rule is to keep questions short – less than five seconds, and in some cases less than one second (*Ibid*.:10).

According to Labov (1966a:64), the first measure for determining the successful elicitation of casual speech is when speech in an interview becomes spontaneous. Labov (1966a:64) defines “casual speech” as the everyday language encountered in informal situations where the speaker pays no attention to the language they are using. Labov (*Ibid*.:64) defines “spontaneous speech” as a pattern of speech in which the conversation is “excited” and “emotionally driven”, which causes the limitations of the formal situation of an interview to be breached. He categorises casual speech as occurring in an informal context and spontaneous speech as occurring in an interview/formal context. Because of this distinction, it becomes apparent that spontaneous speech is not entirely similar to casual speech, but a subcategory of it that can appear in formal situations (Labov, 1966:65). Labov (*Ibid*.:65) notes that, even though there is no guarantee that the values of variables in casual speech and spontaneous speech can be comparable, they can be studied alongside each other.

Labov (1966a:67) demonstrates how the language of a participant in a sociolinguistic interview can shift from careful to spontaneous speech. He defines five contexts in which spontaneous speech often emerges: 1. Speech that is not part of the interview, for example, before the interview begins or during a break; 2. Speech that occurs when the interviewee is talking to a third

person in the room, e.g., a family member or friend; 3. Speech that is not related to the question asked (digressions); 4. Speech that emerges when participants are talking about childhood rhymes and customs (although the context of rhymes is only aimed at casual pronunciation rather than other areas of the grammar, since rhymes are creative writing); 5. Speech that occurs when the topic of "the danger of death" is approached (Labov, 1966a:67); in this instance a question is asked, such as: "Have you ever been in a situation where you thought there was a serious danger of your being killed? That you thought to yourself, "This is it"?" (*Ibid*.:67). Speakers frequently become so involved in the telling of near-death experiences that spontaneous speech emerges. An example of the style shift that occurs with context 5 is given from a participant in Labov's NYC study, in which the answer starts with his usual careful style and shifts to a spontaneous style:

Careful speech:

> [What happened to you?] The school I go to is Food and Maritime – that's maritime training – and I was up in the masthead, and the wind started blowing. I had a rope secured around me to keep me from falling – but the rope parted, and I was just hanging there by my fingernails" (Labov, 2006a:71).

Spontaneous speech:

> I never prayed to God so fast and so hard in my life . . . [What happened?] Well, I came out all right . . . Well, the guys came up and they got me. [How long were you up there?] About ten minutes. [I can see you're still sweating, thinking about it.] Yeh, I came down, I couldn't hold a pencil in my hand, I couldn't touch nothin'. I was shakin' like a leaf. Sometimes I get scared thinkin' about it . . . but . . uh . . well, it's training. (Labov, 1966a:71)

The above example illustrates the necessity of a sociolinguistic interview to elicit spontaneous speech, and the methods formulated by Labov, such as the conversation modules, control the interview environment enough to do this.

The length of the sociolinguistic interviews for this study ranged between 30 minutes to 1 hour for most subjects. There were two interviews that reached one and a half hours. As a result of the Covid pandemic, interviews were conducted through Zoom and the recording equipment was the "call recording" function of Zoom. My own interview schedule followed a modified version of Labov's (1981:35) interview topics created by Tagliamonte (2005) and further modified by me to fit the Port Elizabeth speech community. Below is an excerpt of the interview schedule used for conducting the interviews in this study. The full interview schedule can be found in **APPENDIX D**: LABOVIAN INTERVIEW SCHEDULE.

1. Demographics

- What is your address/in which neighbourhood do you live?
- How long have you lived in that area?

- Are you working or studying?

2. Neighbourhood

- Do the people in your neighbourhood talk to each other?

3. Social practices

- Are there people around here you spend a lot of time with outside your family?

4. Work life

- What was your first day at university like?
- Did anything interesting ever happen at work?

5. Family meals/crafts

- Do you like cooking? Baking?
- Do you have a success or failure story while cooking/baking?

A pilot study was first conducted to test the method of conducting sociolinguistic interviews via Zoom. This was done to evaluate the success of this method in obtaining optimal interview conditions before "official" interviews were conducted. Before the interviews, each participant was contacted via WhatsApp in order to arrange a time for the Zoom interview. After the first two Zoom interviews indicated that spontaneous speech had been achieved successfully, the rest of the interviews continued via Zoom as the interview platform for long-distance conversations.

Ethical clearance was obtained from the North-West University's *Ethics Committee for Language Matters* (Ethics clearance no: NWU-00704-21-A7) before the Zoom pilot interviews started. Tagliamonte (2006:33) provides some ethical guidelines for sociolinguistic fieldwork that were adhered to during data collection: "1. Consent for audio-recording; 2. Guaranteed anonymity; 3. Voluntary participation; 4. Access to researcher and research findings." Before interviews were conducted, participants had to sign a consent form (**APPENDIX A: PARTICIPANT CONSENT FORM**) to allow the video and audio of the Zoom interviews to be recorded. They were also asked to give verbal consent at the end of the interview when the linguistic nature of the study was revealed to them. The anonymity of participants was ensured by removing their identity from the transcriptions. For instance, in the case of those who provided their residential addresses, the addresses were removed from the transcriptions but retained for demographic information in a separate, password-protected document. Anywhere in the transcripts where a participant mentioned their name or the names of friends and family, the names were removed and replaced with the word *name* in brackets, for instance:

(47) my friend [name] and I were at a movie night with two of our other friends (<$C><s n=37>)

In addition, to ensure the anonymity of participants, the Zoom interviews were compressed in a password-protected folder. In addition, since all the consent forms contain the real names of participants, these were also compressed in a password-protected folder. The only persons who have access to the Zoom interviews are me and my two co-supervisors.

3.7 Data Transcription and Coding

The interview recordings were automatically transcribed with the open-source program *Whisper AI*. This program has multiple installations such as Python. Thus, the instructional video linked below was followed in order to install and use it to transcribe the interview audio files: https://www.youtube.com/watch?v=ABFqbY_rmEk&ab_channel=KevinStratvert

After the completion of each automatic transcription, I copied the text of the .txt file to a Word document in order to use the "find" and highlight functions for the preliminary analysis of pragmatic *like*. By reading through the automatic transcription while listening to the Zoom video recordings, I corrected any errors made by *Whisper AI*. I also adjusted the text according to the way a participant spoke – for instance, where the automatic transcription read *going to*, I adjusted the word to the spelling that was closer to how the participant was saying it, i.e. *gonna*. However, to save time, I did not do this for other examples such as *trying to*, which was often said as *tryna*, since such details have no direct impact on this study. Future researchers might adjust the transcriptions to those spellings if they wanted to look at such variants in the spoken language of WSAE.

The *International Corpus of English* (ICE) *Markup Manual for Spoken Texts* (Nelson, 2002) was used as a guide to give anonymous IDs to the interviewees in order to differentiate between the interviewee and interviewer speech in the transcriptions, as well as to adhere to ethical protocol, which requires the identities of the interviewees to be anonymous. The ICE manual is available here:

https://www.ice-corpora.uzh.ch/en/manuals.html

The transcription codes used to transcribe the data for this study can be found in **APPENDIX E**: TRANSCRIPTION CODES.

In terms of ensuring the anonymity of the interviewees, the strategy of giving a pseudonym to each participant was followed, as described in Tagliamonte (2012:51). In her study, the initials of a participant were used, but modified with an ethnically consistent name and surname randomly picked from a list of names from the relevant region's telephone directory, for example, "Katy

Webster = Katherine Walters" (*Ibid*.:52). Since I did not have access to a Port Elizabeth telephone directory, I used the participant's initials and picked out a random name and surname (that start with the same initials as the participant) from the membership list of the specific neighbourhood Facebook group in which the participant resided at the time of the interview.

Sections 3.1–3.7discussed the methods undertaken in this study, from the design of the sample, participant selection, and sociolinguistic interview, to the transcription and coding of the data. The next section focuses on circumscribing the variable context of the linguistic variable and the principle of accountability, as well as how these concepts were utilised in setting up the conditions for the subsequent analysis of quotative *be like* and DM-*like*.

3.8 Circumscribing the Variable Context for Pragmatic *Like*

According to Labov (1966b:15), the term "linguistic variable" refers to a category of variants "ordered along a continuous dimension and whose position is determined by an independent linguistic or extra-linguistic variable." On the other hand, the linguistic "variant" refers to a specific item such as "a morph or a phone" (*Ibid*.:15). For instance, Chambers (2002:17) explains that in Labov's New York survey (1966a), he used brackets to symbolise variables. In the case of Labov's department store study (1966a:32), one of the phonological variables under investigation was (r) (see section 2.2.2 for a description of this variable).

3.8.1 The principle of accountability

Tagliamonte (2011:10) explains that the concept developed by Labov (1972:72), known as the "principle of accountability" directs linguistic analysis not only to consider the variant in question, such as the relative pronoun *who*, but also "all the other potential variants within the relative pronoun system" (*Ibid*.:10). According to Weinrich *et al.* (1968:159), linguistic variants are items that:

> …offer alternative means of saying "the same thing": that is, for each utterance in A there is a corresponding utterance in B which provides the same referential information (is synonymous) and cannot be differentiated except in terms of the over-all significance which marks the use of B as against A.

The principal of accountability is closely related to the definition of a variant, as described in the above quote. For instance, to follow an accountable analysis of a variable, one must identify the variants that say "the same thing" (Tagliamonte, 2012:2) and include all of them in the analysis. According to Labov (1972:72), it is necessary to "isolate the largest homogeneous class in which all subclasses vary in the same way".

The principle of accountability was later expanded to fit other areas of the grammar such as "discourse-pragmatic variables" (D'Arcy, 2017:25). According to D'Arcy (2017:25), the problem presented by discourse-pragmatic variables lies in the requirements derived from the definition of a variant in which it has to be semantically or referentially equivalent to the other variants in the same grammatical subsystem (see Tagliamonte, 2012:10). However, D'Arcy (2017:25) states that discourse-pragmatic variables are "multifunctional" (*Ibid*.:25). This means that pragmatic variants often have distinct referential meanings as opposed to the same meaning (Tagliamonte, 2006:73). An example of this is given by Singler (2001:262), who states that when considering *be like* within the quotative system, there are environments in which other quotatives such as *say* are not used interchangeably with *be like*. One of these environments is when the quotation involves "gesture, facial expressions or non-speech sounds" (*Ibid*.:262). In such cases, *be like*, *go*, and *all* are used but not *say*. In addition, Singler (2001:262) states that *be like* is used to report inner monologue/thought, while "it is not possible with *say* or *go* (however, in one instance of the current dataset, *say* was used to report thought:

(48) I **said** to myself: "No this is too difficult." (<$S><s n=54>)).

The problem relating to discourse-pragmatic features having distinct referential meaning was solved by Sankoff and Thibault to some extent (1981:207; in Tagliamonte, 2012:16). They extended the variable context to include elements such as discourse-pragmatic variables through the notion of "weak complementarity" (*Ibid*.:16). Tagliamonte (2012:16) defines weak complementarity as "the idea that linguistic variables can be identified by their distribution across the speech community rather than by the fact that they mean the same thing". For instance, Tagliamonte (2012:16) states that a variable is usually identified within the speech community where one variant is used more and another variant is used less, even though they do not always mean the same thing. An important extension of the definition of a variable (especially for the study of syntactic variables) is the notion of "structural equivalence" (Tagliamonte, 2012:17). Tagliamonte (2012:17) defines "structural equivalence" as the possibility for two variants of a variable to be considered when they occur in the same kind of context in a language. Sankoff and Thibault (1981:207; in Tagliamonte, 2012:16) state that these variants could be "very different constructions whose only underlying property in common is their usage for similar discursive functions" (*Ibid*.:16).

The notions of weak complementarity and structural equivalence provide the possibility of considering *be like* and other quotatives, such as *say*, as variants of each other in the sense that they occur in the same context: they are used to introduce quotations/ constructed dialogue and usually occur before the quote/constructed dialogue.

Regarding the principle of accountability, Tagliamonte (2012:10) states that quotative *be like* is part of the subsystem of quotatives, and the variants of the quotative system need to be counted in the distributional analysis, such as *say*, *go*, and the zero quotative. This process is called "circumscribing the variable context" (Tagliamonte, 2012:10; D'Arcy, 2017:25). In order to determine the variable context, it is necessary to discover exactly which variants should be included in the range for the chosen linguistic variable. Tagliamonte (2006:86) states that there are certain contexts in which one variant may occur, but another variant never occurs. Tagliamonte (2006:86) explains that environments where a variant either occurs 0% of the time or 100% of the time is called a "categorical context". Such contexts must be excluded from the variable analysis because it is invariable. Sections 3.8.2–3.8.3 circumscribe the variable context for *be like* and DM-*like*, which is then used as the baseline for the analysis of the two variants. A short description of the similarities and some differences in the linguistic internal factors

3.8.2 Circumscribing the variable: Quotative *be like*

The method of circumscribing the variable context for quotative *be like* is as follows. Similar to Tagliamonte and Hudson (1999:154), every instance where a verb introduces constructed dialogue was included in the analysis. Thus, instances where a verb introduces an indirect quote were excluded from the analysis. For example, within complement clauses with *that* or where *that* is omitted:

(49) …when I was in grade 11, they, they **said** to us, when we come back to school, um, the following year, [that] we must have our hair cut by the second day. (<$J><s n=33>)[28]

Instances where constructed dialogue was not introduced by a verb were regarded as a zero quotative (Tagliamonte & Hudson, 1999:156) and were thus included in the analysis:

(50) And we got to the airport, checked in everything, we're waiting, "flight delayed for an hour."

In some examples, the word *like* appears with a quotative verb or on its own before quoted material:

(51) when I came back my mom basically just **said like** "How long have you been smoking for?" (<$F><s n=42>)

(52) he got into trouble. […] **like**, "Nene he broke it." (<$E><s n=48>)

In contexts such as the above examples, the method followed is that of Tagliamonte and Hudson (1999:153): when *like* appears with a quotative verb or on its own before quoted material, it is

[28] Citations for sentence examples from the data of the current study refer to the speaker code and line number in the transcription where the sentence appears.

regarded as a discourse marker. In cases where it appears with a quotative verb, such as *said like*, the example was counted as an instance of quotative *said*. In examples where *like* appeared before quoted material, the quotative usage was counted as an example of the zero quotative, as in example (53). The decision was made not to consider *like* before quoted material as a separate quotative because the literature is inconclusive about its status as such at this time. For instance, Romaine and Lange (1991:253) state that in examples such as (53) and (54), *like* is used to express an example of what might have been said and therefore shares the "similar to" meaning of DM-*like*.

(53) Oh because guys gotta have like, a you know, a tough image **like**, "I don't need no one's help or nothing. (Romaine & Lange, 1991:249)

(54) And he **like**, "I ain't saying Ulysses I'm saying Ulilles." (Ibid.:247)

In example (54), according to Romaine and Lange (1991:253), it is possible that *like* is an elided form of *be like*, in which case it could be regarded as a quotative. However, it can also be regarded as a discourse marker with no reporting function, and its status in such instances is thus indeterminate. Singler (2001:262) states that it could be argued that instances of *like* as in (53) and (54) are not always quotative but that some examples show evidence of being quotative, and are thus developing quotative status.

Singler (2001:260) argues that *be like* occurring with dummy *it*, as in (55):

(55) **It's like**, 'Let's go to the town board meeting; it's more exciting than Monday night wrestling." (Ibid.:260)

is not a variant of the quotative system, since no other quotatives occur with this subject. Thus, it should be excluded from a multivariate analysis. However, Tagliamonte and Hudson (1999:499) consider *it's like* as a variant of *be like*; thus some studies include it as part of their general analysis of *be like* (e.g. D'Arcy, 2004:329; Tagliamonte & D'Arcy, 2004:204). Some studies excluded *it's like* only in the context where the quotatives were analysed in terms of grammatical person, for instance first- vs third-person subjects (e.g. Gardner *et al.*, 2020:288; Gardner, 2017:54), while other studies have excluded *it's like* from the quotative analysis entirely (e.g. Barbieri, 2005:234; Tagliamonte *et al.*, 2016:829; Durham *et al.*, 2012:322). Davydova (2020:10) conducted a factor analysis test with and without *it's like* as a quotative and in both cases, grammatical person was significant. She states that this result justifies the inclusion of *it's like* in the analysis, since the effect on grammatical person is not attributable to *it's like* alone (*Ibid.*:10). The current study follows D'Arcy (2004:329) and Tagliamonte and D'Arcy (2004:204) by including *it's like* in the overall analysis of the quotatives. It also follows Davydova (2020:10) by including it in the analysis of grammatical person as well. This is because *it's like* as a quotative is infrequent in the current dataset and forms part of the overall quotative context. For instance, Gardner (2017:65) states

that since *it's like* can be replaced with *he says* or *I say* in examples such as (56), "the difference between when speakers choose any kind of *be like* construction vs. some other kind of quoting strategy can be evaluated" (*Ibid*.:65).

(56) **It was like** "Where you all going?" (*Ibid*.:68)

Following Tagliamonte and Hudson (1999:157), instances where the quoted speech was incomprehensible from the recording, such as in (57), were included in the frequency counts for the quotatives but excluded when considering the analysis of the content of the quote. For example, when considering thought vs direct-speech reporting. Such instances were coded as "other" when categorising the content of the quote as either thought or direct speech.

(57) It was like, "[incomprehensible]." (Ibid.:157)

Instances where a speaker referred to specific words were excluded from the quotative analysis, since it is not constructed dialogue, such as (58):

(58) Like all the streets are, some **say** "road", some **say** "avenue", some **say** "street." (<$B><s n=59>)

In addition, ambiguous cases where it was not clear if the example was a quote or not were excluded, such as:

(59) I was still on the first disc and I **thought** to myself, I think I still got it somewhere. Just maybe look in the garage and maybe start learning. (<$S><s n=54>)

Examples where a quotative verb with an unrealised quote was later completed with a different quotative verb was only counted once as an example of the latter:

(60) He's always like, um, with projects and stuff, he would always check and he'd **say**, "Uh, maybe word it a different way." (<$P><s n=45>)

For example, the sentence in (60) was counted as part of quotative *say* and not *be like*, which appears first but is unrealised. Following Tagliamonte and Hudson (1999:156), in instances where there was a change in the person or number within the quoted segment, the two instances were coded as two separate quotatives, for example in (61), asked is coded as a quotative verb before the first quotation, while the second quote is coded as a zero quotative:

(61) every time she **asked** me, "[Name], what's your name?". **"Poopsie."** (<$I><s n=59>)

However, where no change of person occurs, the quoted contexts were coded for one quotative verb (in this case, zero):

(62) Some lecturers are, are nice, **like**, "Oh, here's a example" or "Here's like this type of calculus. Here's an example of how to do it." (<$A><s n=42>)

The content of the quote in the dataset included thought, direct speech, and written material as in (63):

(63) the school got **notification**, "Okay you're under 14B boys are playing against, uh, Daniel Pienaar and Uitenhage this afternoon." (<$K><s n=54>)

It also included non-lexicalised sounds, as in (64):

(64) So, ja, I wasn't like super focused on school. I was kind of like, "**mmm**". (<$N><s n=47>)

and gestures, as in (65):

(65) it really freaked me out, so I was like <,> <&>**shakes head**<&/> (<$F><s n=105>)

Thought was coded as [I] for internal and direct speech was coded as [S]. Content of the quote that were written material, non-lexicalised sounds and gestures were all coded as [N/A]. Instances where the quotative verb was not completed with a quoted segment (unrealised quotes) were also categorised under the n/a label, such as in (66):

(66) then I looked to him […] he was like **<,>** and he just laughed. (<$N><s n=89>)

Because it is difficult to distinguish thought from direct speech, the coding followed Tagliamonte and Hudson's (1999:156) method where, using the broader narrative context, if the quoted segment reported a "general feeling or attitude", it was considered thought, for instance:

(67) I'm not that typical teenager, like, "Eew mom, eew dad, get away from me." (<$M><s n=64>)

Instances that were part of a sequence of dialogue or was a response to an utterance by someone else were coded as direct speech:

(68) So I always wanted like a ukulele and I wanted to play it […], but my mom was like, "I'm not going to buy you a musical instrument." (<$N><s n=159>)

In order to test for the internal factor of grammatical person, quotative verbs were coded for use with first-person (69), second-person (70) and third-person subjects (71), as well as no person, such as in the case of dummy *it* (72):

(69) And then I looked to him and **I** was like, "But they'd suit you." (<$N><s n=89>)

(70) **You**'re like, "Cool. You can enjoy them. I can't see them, so they're your friends." (<$J><s n=71>)

(71) …**he** was like, "But why don't we get married?" (<$L><#><s n="76">)

(72) …and **it** was like, "What?" How do you even think of these things before you go to bed? (<$C><s n=76>)

The zero quotative, which does not have a specified subject, was coded for grammatical person in terms of an implied subject, such as:

(73) **I** wanna say it so bad, but obviously **I**'m not gonna. But like, "Bruh, you're working for your mom, how can you not get that job and how can you possibly get fired bro." (<$R><s n=71>)

In (73), the zero quotative was coded for first person since the speaker quoting his thoughts is specified earlier as "I".

To test for the internal factor of tense, the method of Tagliamonte and D'Arcy (2007:204) was followed, in which the data was coded for: 1. Quotative verbs in present tense form with present temporal reference; 2. past tense form with past temporal reference; and 3. present tense form with past temporal reference (the historical present tense (HP):

1. She takes my clothes without asking and **I'm like**, "You need to ask. What if I wanted to wear that?" (<$N><s n=77>)

2. We **were like**, um, "If you do this and this, you must try on my skates." (<$N><s n=33>)

3. One of my guy friends **went** up there and **he**'**s like**, um, "Can you guys stop, please stop doing this?" (<$G><s n=89>)

However, for the statistical analysis, the HP and present tenses were grouped together as present tense so that the binary option between present vs past tense could be tested. Quotative verbs in other tense constructions that were rare, such as present/past progressive, present/past perfect and infinitive forms (such as *he'd **be** like*…) were categorised into either the past or present tense. The zero quotative, which does not contain a quotative verb to base a tense analysis on, was categorised as *n/a* for the tense analaysis.

3.8.3 Circumscribing the variable: DM-*like*

In order to analyse the distribution of DM-*like* across all social groups in the dataset, only the pragmatic uses were considered. The older uses that do not have any pragmatic effect were removed from the analysis. These are cases where it is categorised as the following:

(74) Verb: I do **like** to swim. (<$M><s n=70>)

(75) Preposition: My brother just treats me **like** a little baby. (<$H><s n=78>)

(76) Conjunction: I feel **like** we dodged a lot of drama when we were younger. (<$J><s n=74>)

(77) Compound adjective: It's like a gathering of people that are all **like-minded**, you know. (<$R><s n=58>)

Examples that were included as pragmatic instances of *like* were the following:

(78) Approximative adverb: I lost one of my friends once and only like found her **like** two days later. (<$E><s n=86>)

(79) The “for example” meaning: So, there's a lot of different slang, like, you'll get **like** the coloured slang and in Afrikaans, Afrikaans people are talking slang. (<$E><s n=112>)

(80) The focus marker: Most of **like** the places we went to are so safe for **like** girls to just walk around. (<$N><s n=96>)

(81) Hedge: We used to have **like** a little **like** lake. It's more like just **like** a little big pool of water.” (<$B><B><s n=9>).

(82) Hesitation marker: She's like the, the hard truth, you know, she's **like**, I mean, she shows me love. (<$R><s n=89>)

(83) They often have, **like**, um, come to each other's houses, have the drinks, that kind of thing. (<$A><s n=32>)

(84) The dummy subject it + like: **It's like**, I still end up doing it. (<$M><s n=50>)

Ambiguous instances where it was not clear if *like* was functioning as a discourse marker or some other word class were removed from the sample of DM-*like* uses, such as:

(85) That's basically **like** a defence class or a thousand defence classes put into one. (<$R><s n=82>)

In another example, *like* could either be a interpreted as a verb or a hesitation marker:

(86) Languages aren't my strong suit. You know, like, I **like**, I do the Maths stuff. (<$A><A><s n=103>)

In (85), it is not clear if *like* is being used as a preposition, meaning that being coached in boxing is *like* a defence class (“similar to”), or if it is used as a focus marker, emphasising the noun phrase “defence class”.

For the analysis of the clause position of DM-*like*, the three structural positions, as described in Chapter 1, section 1.1.1, were included in the analysis. The following are examples of the structural positions from the current dataset:

Like before a matrix clause:

(87) But then some lecturers are like "Read these two pages and then answer these questions". **Like**, they don't help at all. (<$A><s n=42>)

(88) I mean, **like**, it's pushing you, you're not like giving something to it to go. (<$R><s n=106>)

(89) He can make me tired and **like** that's the only thing with the age gap, it's like, um, I'm not used to that anymore. (<$F><s n=46>)

Like before subordinate clauses:

(90) Um, I remember trying to explain that to my friend when he asked me **like** what I dreamt of. (<$C><s n=76>)

(91) So, it was **like** where we lived. (<$C><s n=14>)

Clause-medial *like*:

(92) before determiner phrases: I used to have a little bit of **like** a temper tantrum. (<$I><s n=55>)

(93) Before adjective phrases: He had cracked his **like** very expensive watch. (<$I><#><s n=57>)

(94) Before noun phrases: I think having that throughout my **like** childhood. (<$I><s n=80>)

(95) Before verb phrases: It's just easier to **like** explain it like that. (<$I><s n=3>)

(96) And everyone was **like** cheering. (<$H><s n=117>)

(97) And then he **like** ran into a tree. (<$H><s n=96>)

3.9 Data Analysis and Interpretation

The analysis of the data for quotative *be like* and DM-*like* consisted of a distributional analysis as the first part and a statistical analysis using the conditional inference tree (c-tree) method. According to Tagliamonte (2012:121), the first step in conducting a quantitative analysis is to obtain the frequency count of the variants of a variable in a specific dataset. The second step is to investigate the distribution of the variants across all the factors that have been found to influence their use: the internal (linguistic) and external (social) factors. Tagliamonte (2012:121) states that after the distributional analysis, statistical modelling can be done. According to Tagliamonte (2012:121), "statistical modelling provides a formal mathematical assessment of the relationship between the dependent variable and the independent variables". Thus, the next section describes the methods for the distributional analysis of pragmatic *like* undertaken in this study.

3.9.1 Methods for distributional analysis

In order to investigate pragmatic *like* in the WSAE dataset, it was first necessary to retrieve all instances of the lemma *like* from the transcribed interviews. This was done using the free concordancing program AntConc (version 4.2.4). Antconc is a "freeware corpus analysis toolkit for concordancing and text analysis" developed by Laurence Anthony (Anthony, 2024). This software is available here:

https://www.laurenceanthony.net/software/antconc/

After the transcription files were uploaded to AntConc, the following settings were applied: For left and right context, the maximum of 25 tokens was selected since many instances of quotative and discourse-marker *like* relied significantly on context to be coded correctly in terms of internal factors for quotative *be like* and word classification for *like*. In the **Search Query** window, only the "Words" box was ticked. In addition, "All hits" were selected and in the **Search Options** ribbon, "sort to right" was selected, as well as the options sort 1:R1; Sort 2:2R; Sort 3; 3R. Lastly, "order by frequency" was selected. Figure 3.1 provides an illustration for the concordance and also shows the settings:

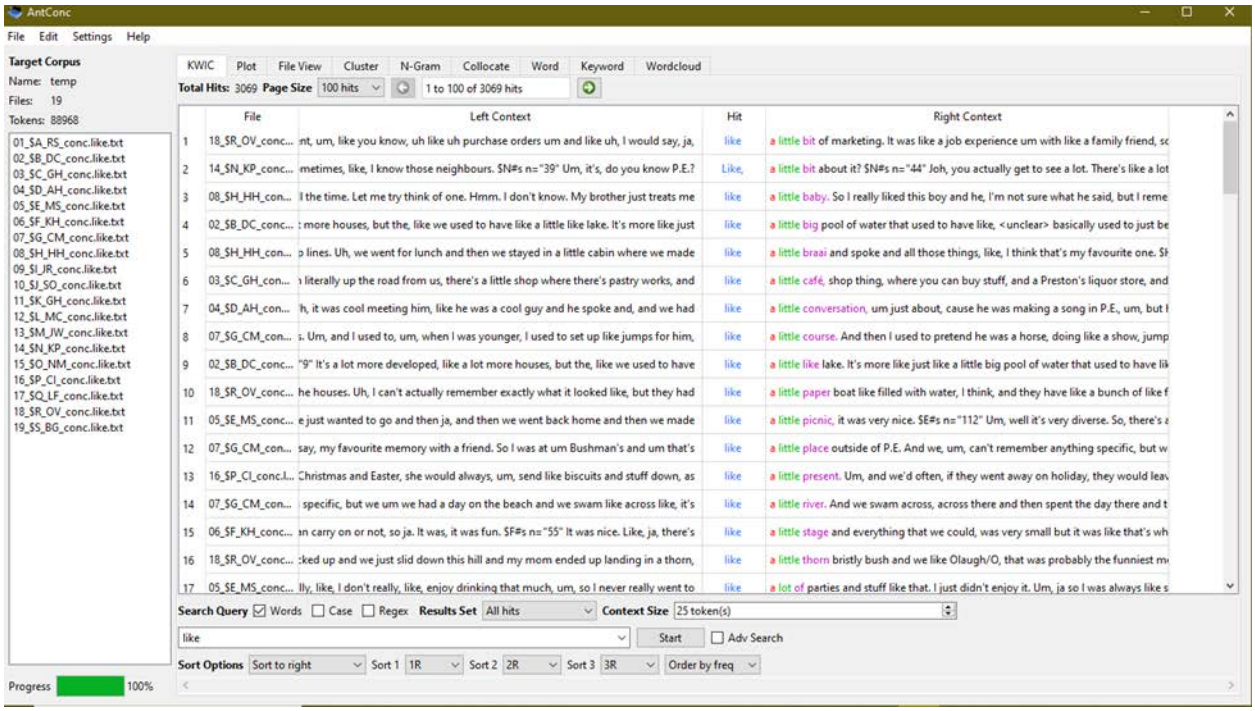


Figure 3.1: AntConc search results for lemma like

After the *like* tokens were extracted, they were copied and pasted to an Excel sheet. The tokens were then analysed, and one-letter codes were used to classify the word categories such as "verb", "preposition", "quotative", and the various pragmatic functions of DM-*like* such as "focus marker", "hesitational marker" and so forth. With this method, all instances of discourse-marker *like* and quotative *be like* were identified. AntConc was also used to identify all the other quotatives in the dataset for comparison with *be like*, using the search terms for each, such as *say*, *go*, *think* and so on. The less frequent quotatives (such as *decided*, *scream* and so forth) were identified by reading through the transcripts and annotating them with the symbol <quote> at the start and end of the quote. The zero quotatives were also identified manually and annotated in the same way since these quotatives do not have a verb to base the concordance search on. The symbol <quote> was then used as the search term in AntConc in order to generate concordance lines for them and copy them to the quotative Excel sheet.

3.9.2 Methods for the distributional analysis of quotative *be like*

All the instances of quotatives in the data with their left and right contexts were copied into an Excel sheet. The quotatives were coded for quote kind, e.g. *say*, *go*, *be like* and so forth; they were also coded for pseudonym, gender, social class, and age of participant. The quotatives were also coded for the internal factors of grammatical person, tense and quote content. After the completion of the coding, the quotative dataset was ready for distributional analysis of the variants.

As a general principle for counting variants, Milroy and Gordon (2003:164) cites Guy's (1980, in Milroy & Gordon, 2003:164) suggestion that 30 tokens per variable "is a reasonable objective" (Milroy & Gordon, 2003:164) since (N = 30) "is an important dividing line in statistics generally between large and small samples" (*Ibid*.:164). Milroy and Gordon (2003:164) explain that the figure of 30 tokens is important since different parametric tests of significance are chosen for samples that are either more or less than this number, "which take account of different relationships to the population from which they are drawn" (*Ibid*.:164). Milroy and Gordon (2003:164) state that the general statistical law is that with a number lower than 10, the likelihood is large that there will be random fluctuation, while a number of more than 10 "moves towards 90 percent conformity with the predicted norm" and rises to 100 percent at 35 tokens (*Ibid*.:164). The predicted norm refers to normal distribution. This is defined as a "symmetric bell-shaped distribution" (Bolstad, 2007:131) and occurred so frequently in the past that it was thought of as normal. This distribution is known as the "normal" or "Gaussian" distribution (*Ibid*.:131). According to Labov (2001:480), one of the areas in VS where a bell-shaped distribution is observed is in stable variables where there is no language change in progress. For instance, he illustrates the bell-shaped pattern with a histogram for the distribution of the (o) variable in the region of Midland:

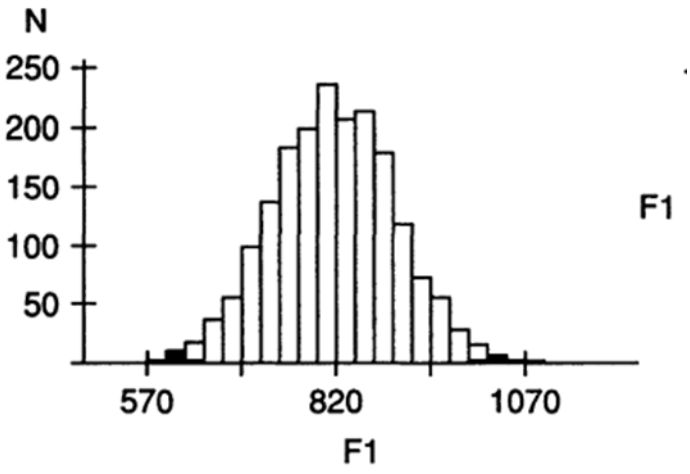


Figure 3.2: Normal distribution pattern for stable linguistic variables (Labov, 2001:480)

This pattern is often observed in the distribution of stable linguistic variables across age where there is an incremental rise in frequency from very young ages up to an adolescent peak (the frequency is the highest among teenagers), after which the frequency declines with subsequent older age groups. An example of the adolescent peak in the stable linguistic variable (dh) from Labov's (2001:269) Philadelphia Neighbourhood Study is illustrated in Figure 3.3:

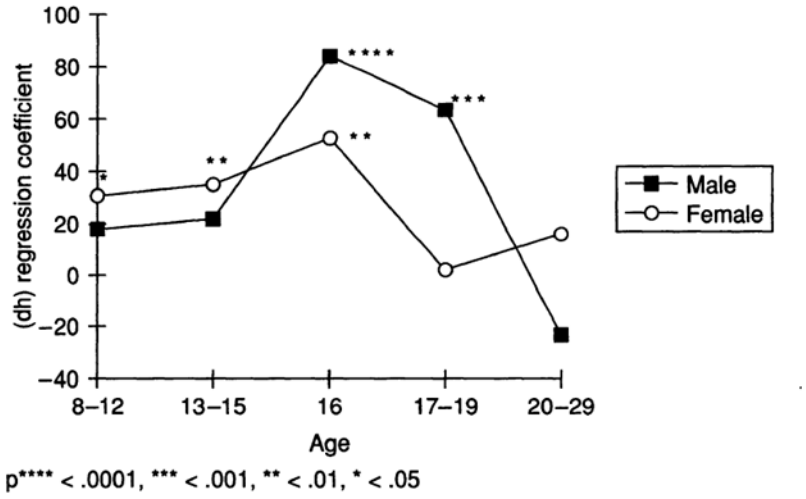


Figure 3.3: Adolescent peak in variable (dh) (Labov, 2001:269)

Thus, having established the minimum of 30 tokens per variable required for reliable distributional analysis, the next section describes the methods undertaken for presenting normalised frequency counts of the quotative variants.

According to Lindquist and Levin (2018:42) the method of normalising data is necessary when corpora or subcorpora differ in sizes. In the case of a sociolinguistic study such as the current one with a small dataset, the frequencies of the different social groups under investigation were normalised to ensure accurate representativity of the data since some groups had more speakers than others, thus, their absolute frequencies for the quotatives, discourse-marker *like* and total words in the dataset tended to be higher than the smaller groups (e.g. the women consisted of 11 speakers while the men consisted of 8 speakers). Firstly, in reporting the overall frequencies of each quotative out of the total number of quotatives, the absolute frequencies or Ns (See Chapter 5, section 5.1) were provided and the percentages calculated from the total number of quotatives (N = 430). These results were not normalised since it was not necessary to do so – the result merely indicated how frequently each quotative occurred out of the total quotatives used by all the speakers in the sample. Overall distributional results were generated in the form of proportional bar graphs and in some instances in pie charts in *R*, a free statistical program available at: https://www.r-project.org/. *R* was used to obtain the raw frequencies illustrated in all the tables reporting results. For working in *R*, the instructions of Gries (2021:50) was followed.

First, *R* was installed from the website mentioned above. After *R* was installed, the instructions for installing RStudio was followed, obtained from:

https://rstudio.com/products/rstudio/download/#download.

The next step was to calculate the raw frequencies for each quotative in R. In order to upload the data from the Excel file, the package *tidyverse* was installed and then applied with the library(tidyverse) code. Then “import” was selected and “From Excel”. The quotative data sheet was then uploaded from the “browse” function. The “Age” category was set as a “numeric” rather than “character” variable. After this, the data was uploaded to *R*.

To report on the quotative distribution across age groups, social class, and gender, the data were normalised by calculating the percentages for each quotative from the N values out of the total quotatives for each group. For instance, Table 5.3 in Chapter 5, indicated that *be like* occurred 52% of the time, calculated from the N value of 178 out of the total quotatives of 341 produced by the 18–20 age group.

In order to report the distribution of the linguistic internal factors of content of the quote, tense and grammatical person, the Ns that a quotative obtained separately for, for example, thought, speech and the n/a category (e.g. *be like*: thought: N = 60; speech: N = 121; n/a: N = 15) were taken and the percentages for each were calculated out of the total number of the relevant quotative (e.g. *be like*: N = 196, see Table 5.15). This method was used since the main concern was to see how frequent a quotative is in the context of reporting either thought or speech.

3.9.3 Method for the distributional analysis of DM-*like*

Before the distributional analysis for discourse-marker *like* could commence, an Excel sheet containing all DM-*like* examples (excluding the quotatives) was prepared and each instance of *like* was coded for structural position (e.g. clause-initial or clause-medial) as well as for the gender, social class and age of the speaker who produced the example. The Excel sheet was then uploaded in *R*, which was used to calculate the raw frequencies of DM-*like* for each social group as well as for the structural positions.

Following the method of Tagliamonte (2005:1904), normalisation of the results for discourse-marker *like* was achieved by calculating the percentages of *like* out of the total number of words in the dataset for each social group. To get the total number of words for each group, the transcript files were prepared in order to count the words in AntConc. First, the files were saved in UTF-8 encoding, after which an R script was used to tag each word in the transcripts with <w> for words and <p> for punctuation, in order to avoid counting punctuation such as commas as words (each

transcript was uploaded separately to R and tagged). The script was provided in Gries (2009:87) for tagging a vector text and adapted to some extent before the best result was achieved.

The transcripts were then edited in order to represent words with hyphens such as *like-minded*<w> as one word; as well as words with punctuation such as apostrophes or full stops where the apostrophe was initially cut off from the word, such as: *I'm*, which were initially tagged as I<w>'<p>m<w> but edited to be one word: *I'm*<w> and P<w>.<p>E<w>.<p> which was corrected to P.E.<w> (Port Elizabeth). The names of places were also corrected to count as one word instead of two, such as: Bluewater Bay<w>. The transcripts already contained annotations such as <unclear>; <O>laugh<O/>; <?>not sure<?>, [address removed]; [name] and <&>transcriber comments<&/> before tagging each word, which meant that all <w> tags had to be manually deleted from words within the annotation tags.

The total number of words for each group was obtained through the concordance programme AntConc, by first opening the "Corpus Manager", selecting "Raw files" and then "Add Files". The total number of words were obtained for each subgroup by, for example, selecting only the transcription files of the 18–20 female speakers and uploading it to AntConc.

The normalised percentages for discourse-marker *like* were then calculated out of the total words for each group, for instance, the 18–20 female group obtained a frequency of N = 1474 *like* tokens (Table 5.33, Chapter 5) and the result of 2.8% was calculated from the young female group's total words in the sample: 52,141. The proportional bar charts for DM-*like* were created in Excel since the code in *R* would have been too complicated to add the total word counts and it also helped to save time.

3.9.4 Statistical analysis: Conditional inference trees

The method chosen for the statistical analysis of the quotatives is the conditional inference tree (c-tree) (Hothorn et al., 2015:1). Levshina (2020:611–612; see also Gries, 2021:13; 461) explains that conditional inference trees or CITs is a method of recursive partitioning which first selects the best predictor to help distinguish between the different values of the dependent/response variable[29]. The next step is to split the dependent variable into several datasets using a p-value as a splitting criterion (Levshina, 2020:612; Gries, 2021:461). Levshina (2020:612) notes that most algorithms split the variable into binary partitions. In the last step, the first two steps are repeated recursively until no further splits can be made. Levshina (2020:612) states that the objective of binary partitioning is to have "maximal purity (or minimal impurity) in the terminal nodes". For instance, her graph illustrated in Figure 3.4 below shows how binary partitioning

[29] For instance, if the dependent or response variable is DM-*like*, it might select either gender, age or social class as the best predictor to help distinguish between the different values it obtained.

groups data together in similar types (such as stars or circles) and splits the data until a dataset with the least impure objects is achieved, such as mostly blue and mostly white circles.

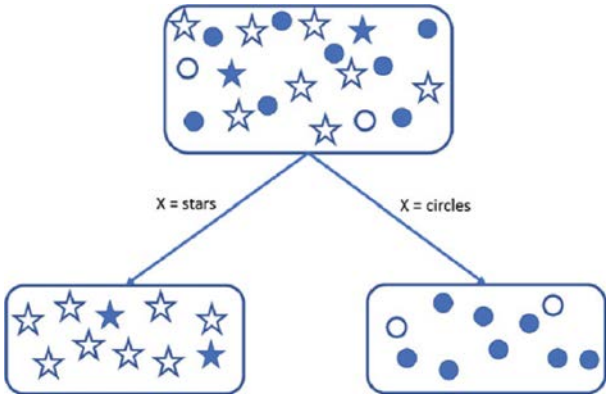


Figure 3.4: Binary Partitioning of Conditional Inference Trees (Levshina, 2020:612)

Levshina (2020:612) states that in the node with stars, 80% are white and in the node with circles, 80% are blue and "this means that splitting the data according to shape allows us to predict the colour correctly in 80% of the cases" (*Ibid*.:612). Levshina (2020:613) states that conditional inference trees are more suitable for small datasets for which logistic regression models are not suitable. Tagliamonte and Baayen (2012:29) state that c-trees are able to include continuous variables such as age.

To illustrate the method of binary partitioning with a linguistic variable, Tagliamonte and Baayen (2012:2) use the variation between was/were in sentences such as in (98), which is the standard English form, while the variant in sentences such as (99) is non-standard.

(98) There **were** badgers in there.
(99) There **was** black clouds.

One of the factors that was tested for *was/were* variation was "proximity", for which two levels (Adjacent and non-adjacent) were coded, which tested all cases of "adjacent verb to plural element sequences", such as in (100) and (101):

(100) There **were** badgers in there. (adjacent)
(101) There **was** [some funny] people. (non-adjacent)

Tagliamonte and Baayen (2012:22) explain that for adjacency, the conditional inference tree calculates whether "splitting the data into adjacent and non-adjacent utterances" (*Ibid*.:2012:22) results in a dataset where *was* is used more frequently and another set where *were* is used more frequently. Furthermore, for each step of the recursive splitting process, each predictor is tested whether it is independent from the response variable. If it is independent, the predictor is meaningless for predicting the use of *was* or *were*. If the "null hypothesis of independence"

(*Ibid*.:22) is rejected, it means that the predictor is meaningful. The algorithm for splitting the data stops when there are no more meaningful predictors. In the case where there is more than one meaningful predictor, the algorithm chooses the predictor with the strongest link to the response variable and the p-value is given, after which a binary split based on that variable is executed. Furthermore, Tagliamonte and Baayen (2012:22) state that conditional inference trees have safeguards which prevent the selection of predictors and variables from being biased in favour of those predictors with many levels or biased in favour of numeric predictors such as the age of speakers.

The p-value involves what is called the "Null Hypothesis Significance Testing" (Gries, 2021:29). Gries (2021:29) explains that quantitative analysis attempts to show that the statistical form of H_0 (the null hypothesis) is wrong, and because the H_1 is the counterpart of H_0, it would simultaneously support the H_1. In order to do this, a level of significance is established, usually 0.05 (or 5%), which is the "threshold value" from which H_0 is either accepted or rejected. Gries (2021:29) states further that linguistic data are analysed via the calculation of some effect, such as a difference in frequencies, distribution or a correlation. Furthermore, the significance test calculates the probability of error p – something that quantifies the probability of finding a specific effect (such as a gender difference in the use of DM-*like*) or something that "deviates from H_0" (*Ibid*.:29) more in the sample than when H_0 is true in the population. Gries (2021:29) states that p is compared to the level of significance and if it is smaller, the finding is significant and H_0 can be rejected and H_1 can be accepted. If, on the other hand, p is greater than the pre-selected level of significance, the finding is considered not significant and H_0 is accepted and H_1 rejected. Gries (2021:29) states that the term "significant" in the context of statistical analysis means that a result is less likely to have occurred under the null hypothesis (which implies a random or chance distribution) than the defined standard threshold which is usually 0.05 or 5%. According to Gries (2021:29), a significant result does not necessarily mean that H_1 has been proven, since there is the possibility of error p that H_0 is correct even though the observed result which accepts H_1 has occurred. Thus, different levels of significance have been established: $p<0.001$ is considered highly significant; $0.001\leq p<0.01$ is very significant (meaning if the value of p is from 0.001 to 0.009); $0.01\leq p<0.05$ is significant (meaning if the value of p is from 0.01–0.049); $0.05\leq p<0.1$ is considered marginally significant (if p is from 0.05 to 0.099). However, Gries (2021:29), states that since the last expression is larger than the 0.05 standard, its use is discouraged and a p-value that is larger than 0.05 is usually referred to as not significant.

To illustrate how a conditional inference tree is used to analyse linguistic data, Tagliamonte and Baayen (2012:4) use the effect of polarity (negative or affirmative contexts, e.g. *was/weren't*) on the use of the *was/were* variable. They state that the most common effect of polarity is where

weren't occurs more often in negative contexts and *was* occurs more often in affirmative contexts, for instance:

(102) There **weren't** always bulls. Sometimes there **was** a few pigs, a few sheep (Tagliamonte & Baayen, 2012:4).

According to Tagliamonte and Baayen (2012:26), the conditional inference tree emphasises complex interactions which characterise a dataset and it clarifies how predictors work together. An example c-tree, which tests the polarity contexts *of was/were* against the external predictors of age and the individual, is illustrated in Figure 3.5 below:

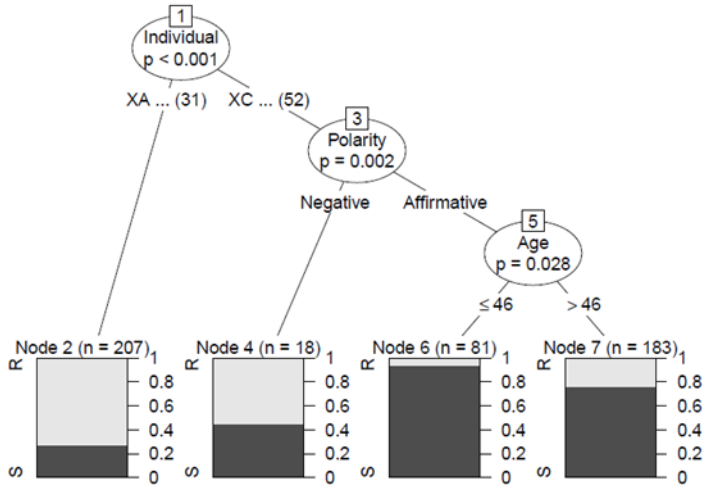


Figure 3.5: C-tree of polarity for the was/were variable (Tagliamonte & Baayen, 2012:26).

The analysis which Tagliamonte and Baayen (2012:26) provide is that "polarity is relevant only for a subset of the individuals" (nodes 4, 6 & 7), and "the effect of Age (nodes 6 and 7) is further restricted to positive polarity items for that subset of individuals". Tagliamonte and Baayen (2012:32) conclude that mixed effects modelling is more suitable to relatively balanced data with one or more effects factors, however, for highly unbalanced data and complex interactions, conditional inference trees are more suitable. Thus, the method of conditional inference trees for the statistical analysis of the current dataset was chosen since it is small and unbalanced (for instance, middle-class, younger females are overrepresented in the data).

3.10 Conclusion

In conclusion, chapter 3 discussed the methodological approaches chosen for this study such as the data collection, sampling, participant criteria, recruitment and interview methods. The study made use of a judgement sample and recruited participants via the friend-of-a-friend and snowball methods, as well as through posting a flyer on Facebook. The chapter also discussed the initial and final sample designs achieved for this study. It also discussed the methods for creating the social class index, index scores and weights, as well as how participants were scored to

categorise them into the three groups of middle class, lower middle class and upper middle class. In addition, the selection of the age-ranges (18–20 and 33–48) for the analysis was discussed and the reasons why these age ranges were chosen. It was also explained why White South African English was the chosen SAE variety in this study. With regards to the interview method, the Labovian interview method and interview schedule were discussed, as well as the concepts of casual and spontaneous speech since a Labovian interview aims to elicit spontaneous speech.

In section 3.7, data transcription and the coding of the interview transcripts were discussed. Section 3.8 discussed how the variable contexts for quotative *be like* and DM-*like* were circumscribed in order to reliably conduct distributional and statistical analyses. Section 3.9 discussed the methods chosen for data analysis and interpretation. Firstly, the methods for the distributional analysis of pragmatic *like* was discussed. Secondly, the method that was chosen for statistical analysis, the conditional inference tree (c-tree) was defined and the reason for choosing this method was discussed. Chapter 4 will discuss the chosen speech community for this study (Port Elizabeth) and elaborates on the historical context and linguistic development of White South African English.

CHAPTER 4: THE SPEECH COMMUNITY

4.1 South Africa's Social Groups and their Languages

Archaeological evidence suggests that the first group of people living in South Africa were hunter-gatherers known as the Khoesan[30] (Barnard, 1992:28; Meyer, 2014:33–34). The Khoesan speak languages that use click sounds and Barnard (1992:22) describes these languages as belonging to a Khoesan "superfamily". According to Mesthrie (2002:11), most of these languages, with a few exceptions, are almost extinct. The second group are the Bantu-speaking communities, who were discovered to be the earliest farmers in South Africa and whose presence in the country dates from about 2000 years ago (Parkington & Hall, 2010:70; Thompson, 2014:32). According to Herbert and Bailey (2002:51), the Bantu languages consist of Northern Sotho, Southern Sotho, Ndebele, Swati, Tsonga, Tswana, Venda, Xhosa and Zulu.

4.2 The Arrival of the Dutch and English at the Cape

Ross (2010:169–170) recounts the sequence of early contact between Europeans and the above-mentioned groups, such as the arrival of the Portuguese ship of Bartholomew Dias at Mossel Bay in 1488. There the Portuguese arrivals had an altercation with a member of the Khoekoen, who was killed. Bartholomew and his crew were looking for a path to travel from Europe to Asia by sea. They achieved their aim with Vasco da Gama, after which Europeans frequently visited the coasts along Africa and traded with the Khoekhoen for cattle and sheep. From around 1590 the English and Dutch joined this trade route and the Dutch became the most frequent visitors to Africa.

According to Ross (2010:15,173–182), the Dutch East India Company (VOC: *Vereenigde Oostindische Compagnie*) made the decision to set up a permanent base in Table Bay; three of their ships arrived there in 1652, one of which was captained by Jan van Riebeeck. Over the years, this base developed into the Dutch Cape colony, governed by the VOC. This colony consisted of Dutch and English and was joined in 1672 by 150 Huguenots and a population of slaves, including Angolans in 1657. From 1779, over half of the VOC central administration workers in the Dutch colony were born in the Cape and all of them were children of VOC officials (Ross, 2010:185).

Thus, a third social group that developed a distinct community in South Africa were the white Afrikaans-speaking people who were mostly descendants of the Dutch settlers, the French

[30] Consisting of various distinct groups that are collectively referred to as the San and the Khoekhoen.

Huguenots and other Europeans (Mesthrie, 2002:4,12). According to Roberge (2002:79), during the years of 1652–1759, the language that developed among the European settlers in South Africa was not standard Dutch but the "popular and regional varieties" (*Ibid*.:79) of the common people. Roberge (2002:79–80) states that Afrikaans either developed from a southern region of South Holland (Kloeke, 1950; in Roberge, 2002:79) or from a seventeenth century "inchoate koine" (*Ibid*.:80) from Amsterdam. According to Roberge (2002:82) the slave community in the Dutch colony developed, among other creoles, a "jargonised form of Dutch" (*Ibid*.:82). One of the Cape Dutch communities that had formed by the 1800s were the Griqua who had a mixture of European, Khoekhoe and slave lineage (*Ibid*.:81). Thus, a fourth social group that formed were the coloured community. These two groups developed Afrikaans and speak two varieties of the language (White Afrikaans and Coloured Afrikaans). Roberge (2002:83) states that in 1925, Afrikaans was recognised as one of the official languages of the Union of South Africa, along with Dutch and English.

Mesthrie (2002:11–12) summarises some of the minority languages in South Africa, such as German, Hindi, Urdu, Gujarati, Tamil, Telugu, Portuguese, Polish, Dutch, Italian, as well as, more recently, the French and Portuguese varieties of Africa. Some languages that were used by slaves in the Cape, such as Malay and Malagasy, are now extinct in S.A. Chinese languages such as Cantonese, Hakka and Mandarin can also be found in South Africa. Some minority Bantu languages that exist in the country are Chopi, Kalanga, Shona, Chewa, which are spoken by migrant workers from neighbouring countries (Mesthrie, 2002:11).

4.3 The White English-speaking South African Group

According to Ross (2010:208–210), the VOC's power in the Cape had declined significantly by the 1790s, and in 1795 a British naval ship arrived at False Bay (see Figure 4.2). After a brief battle, the British took control of the Cape colony until 1803, when it was in the hands of the Batavian Republic[31] for a short time, before being retaken by the British in 1806 and permanently surrendered to them in 1814 (Legassick & Ross, 2010:253). See Figure 4.1 for the extent of the Dutch Cape settlement in 1795.

According to Lambert (2014:588), the British government sent 4,500 settlers in 1820 and established them in the Eastern Cape frontier region. Figure 4.3 shows the Eastern frontier of the Cape of Good Hope colony.

[31]After the French conquered the Netherlands, the Batavian Republic was established in a constitution of 1798. Its government was modelled after the Directory in France and was tied to France through an alliance (Britannica Academic, 2024).

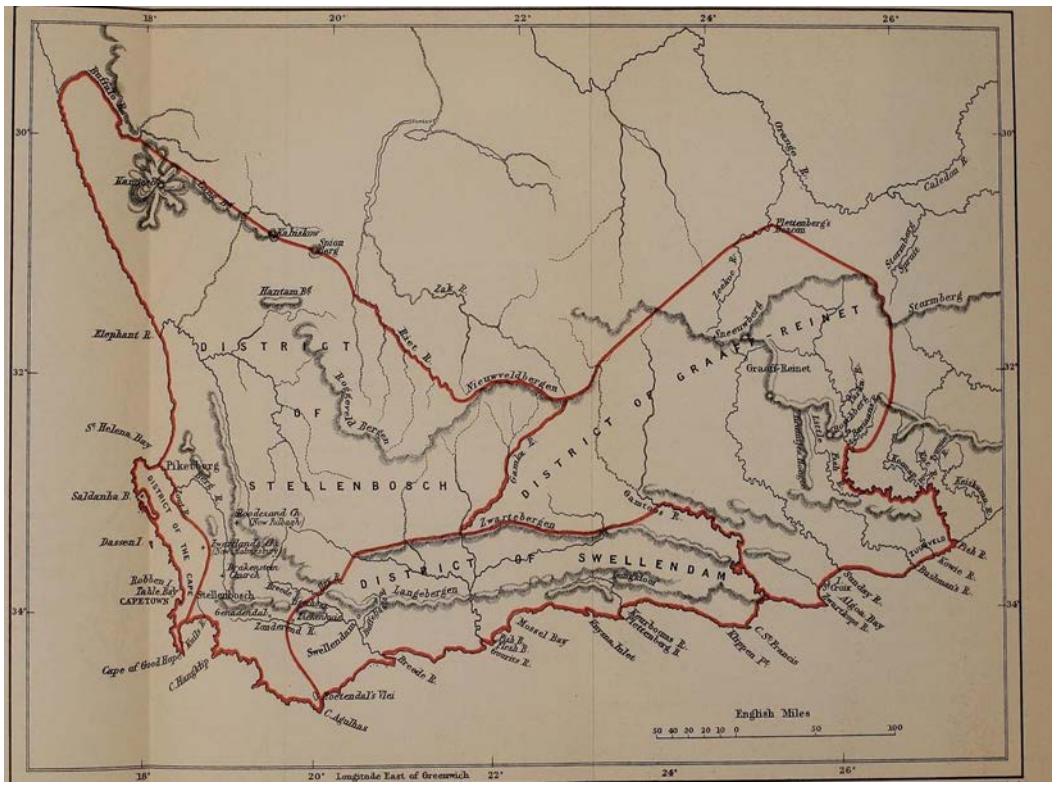


Figure 4.1: Map of the Dutch Cape Colony in 1795 (Theal, 1910:344)

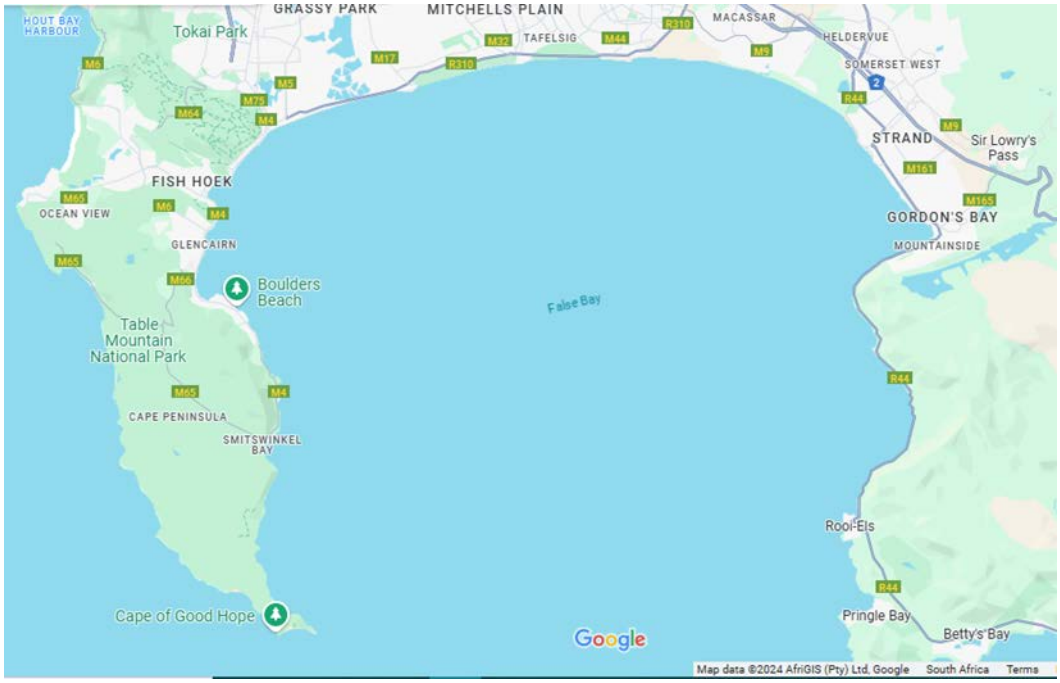


Figure 4.2: Map of False Bay (Google Maps, 2024)

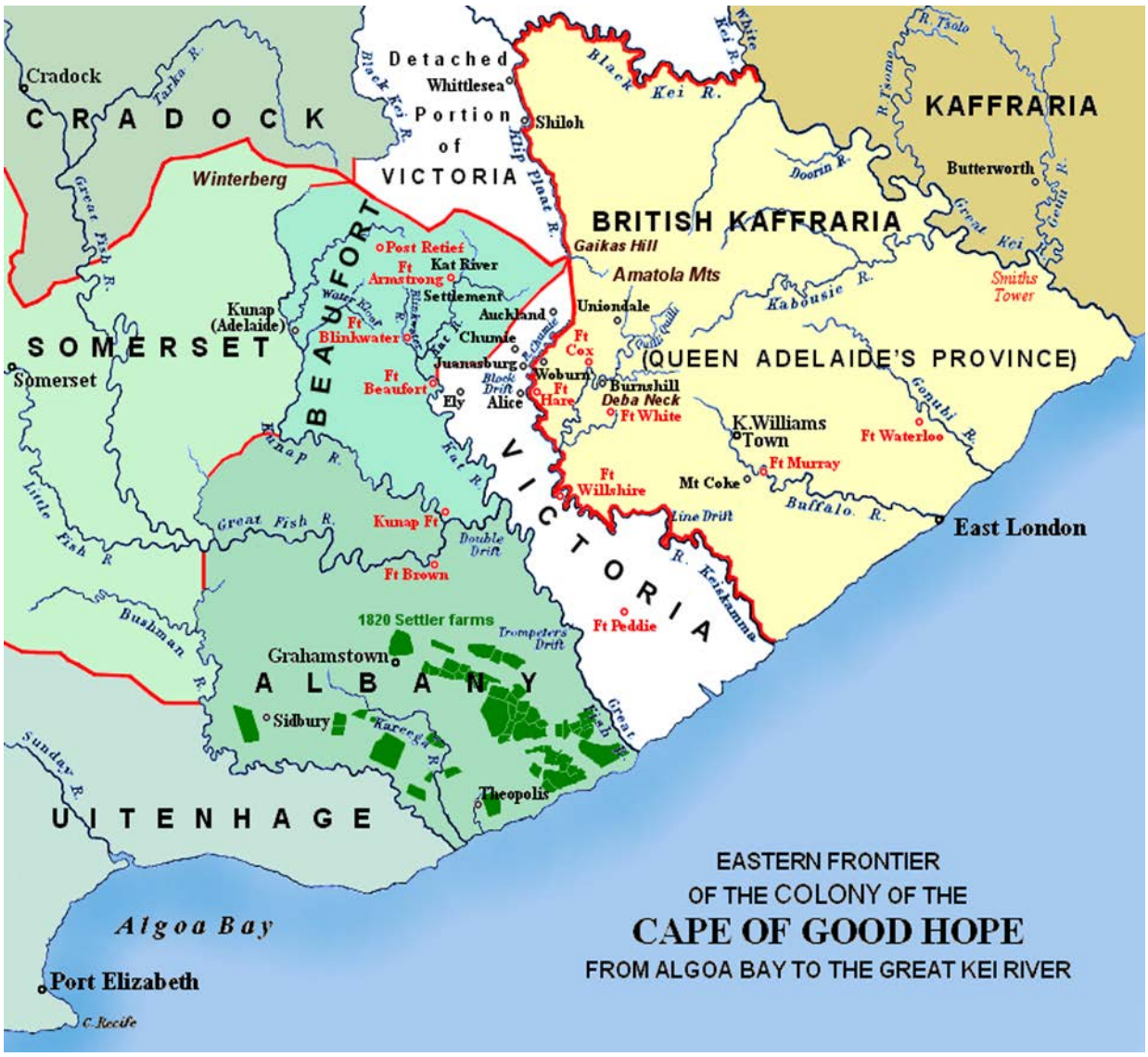


Figure 4.3: Eastern frontier of the Cape of Good Hope colony, ca. 1820–1850 (JMK, 2021)

According to Ross (2017:191), during the time of the British occupation in the Cape, there were many British immigrants who came to settle in the British Cape colony, some of them were from the British military, while others travelled from India. Ross (2017:191) states that these immigrants were initially a separate group in the colony, but as time went by, they married into the Cape Dutch community and those who were bilingual in Dutch and English as a result of family ties often became successful merchants in the Cape. According to Ross (2017:191), the British government organised "the most significant wave of immigration" in the history of S.A. after the Napoleonic wars. The reason for this was that the British government wanted to strengthen their control of the colony. Another reason for the immigration was that it was viewed as a way to improve the poverty of the British people caused by the post-war depression. These settlers came to the Cape during 1820, and from there travelled to the "Eastern Districts" (*Ibid*.:191) – the

Zuurveld in Albany, between the valleys of the Kowie and Fish Rivers south-east of Grahamstown, and Salem further west. Figure 4.4 illustrates the settler route from Algoa Bay.

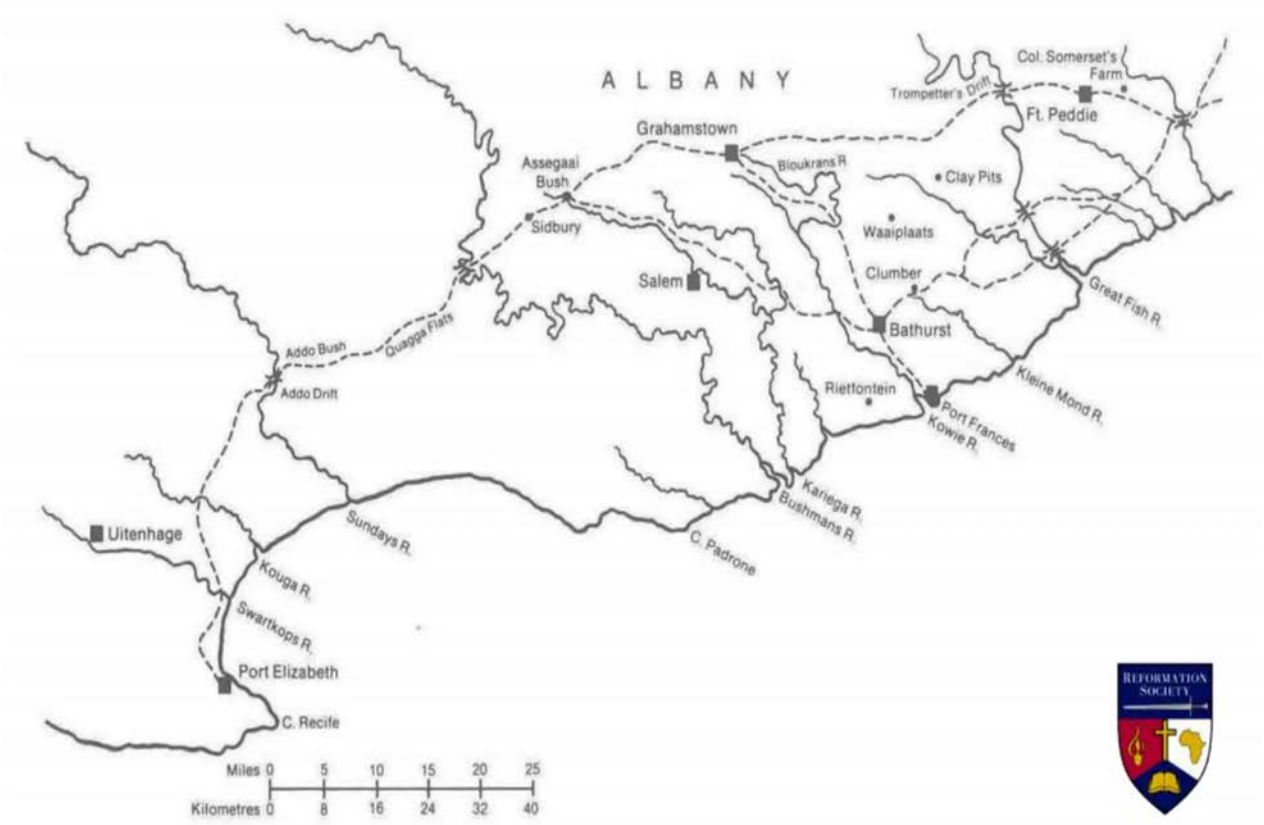


Figure 4.4: The 1820 British settler route from Algoa Bay to Assegaai Bush (Hammond, 2020)

Furthermore, Ross (2017:191–192) states that the 1820 British settlers were divided into 68 parties and each of them received a piece of land in the Colony. Of these settlers, 31 % were skilled labourers or artisans, 11% were traders, 5% were from the army and navy and 4% were people with professions. According to Ross (2017:192), the intention was for the settlers to form agricultural groups along the coast near Grahamstown, however, only 39% of them had agricultural experience and only a few of them became successful farmers. Most of the others abandoned their farms after a few years and became artisans and merchants in the small towns of the Eastern Cape. Initially, the Scottish settlers were sent to the Baviaans River north of Grahamstown, the Welsh settled near Caledon in the Western Cape and the Irish settled in the Oliphants River valley. After some years the Irish moved away from the valley, some back to Ireland while others joined the English in Albany (Ross, 2017:192). Figure 4.5 shows the areas of these initially isolated British settlers.

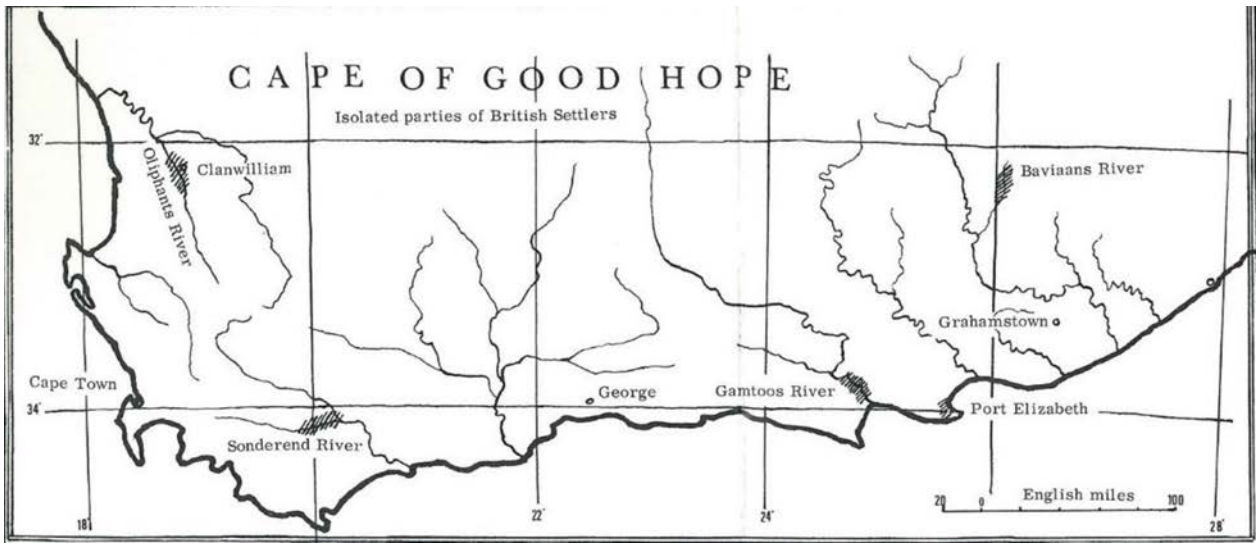


Figure 4.5: Locations of the Scottish, Welsh and Irish settlers in the Cape (Morse, 1971)

Through another immigration scheme of the British government (the Byrne scheme), 4500 British immigrants were sent to Natal (Lambert, 2014:588) in 1849. Ross (2017:193–194) states that because of economic pressure and their aversion to British rule, Dutch-speaking farmers (later known as the Voortrekkers) migrated and settled in the Highveld Grasslands (now Gauteng) and across the Drakensberg to settle in Natal. This migration came to be known as The Great Trek. The Voortrekkers soon set up farms behind Port Natal, which threatened British rule since they did not want any other European power to control a harbour. This resulted in the British government initiating several immigration schemes such as the one described above. Because of these immigrations, the trekkers moved on between the Vaal and Gariep rivers and in 1852 established the South African Republic (Transvaal) through the Sand River Convention.

According to Lambert (2014:588), because of the discovery of minerals in the 1870s (diamonds in Kimberley and gold in Witwatersrand; Ross, 2017:194), more British immigrants entered South Africa and after the Anglo-Boer War, they were persuaded to settle in the Transvaal Colony which had newly been annexed by the British government. Ross (2017:195) states that many of these immigrants were from the mining communities of Cumberland and Cornwall. According to Ross (2017:195) 75,000 British people migrated to Johannesburg after the discovery of gold, along with other Europeans and North Americans. The Jewish community also increased from 4,000 before 1840 to 40,000 after one generation with immigrants mainly from Lithuania.

4.3.1 The identity of the White South African English community

Lambert (2014:588–589) states that because of a lack of roads and postal services between the different British colonies in South Africa, these colonies were isolated from each other; many formed close social networks with Afrikaners and many intermarried. For instance, some Afrikaners are descended from Scottish settlers.

Differences developed between the British of the Cape Colony. For instance, Lambert (2014:589) describes the British who settled in the Eastern Cape as living a "frontier way of life" (i.e. pastoral) and those in Cape Town as living a metropolitan lifestyle with liberal traditions. Another group, the Natal British, called themselves Natalians and identified the most with Britain. All of the groups kept their British identity close and referred to themselves as British or English until the early twentieth century, when the term "English-speaking South Africans" (*Ibid*.:589) started being used more generally. This change was influenced by the defeat of the Zuid-Afrikaansche Republiek (later Transvaal) and the Oranje Vrijstaat. This defeat occurred after the Ango-Boer war, which ended in May 1902, when the two generals Louis Botha and Jan Smuts accepted British rule over South Africa (Lambert, 2014:595).

According to Lambert (2014:600, 607), fighting alongside Afrikaners in WWI strengthened the English community's identity as South Africans, however, even though they began to see South Africa as their own country, they still saw themselves as British subjects. When the National Party, led by Afrikaners, won the general election in 1948, English traditions, culture, the Union Jack and British national anthem lost their status. At this time, there were fewer British-born English speakers than locally born ones and other white communities joined them (such as the Jews), who adopted English as their home language.

4.3.2 Defining English-speaking South Africans

Some sociological researchers on White English-speaking South Africans (WESSAs) have stated that they are an under-researched group who are difficult to define culturally and who tend to avoid defining themselves as a group (Garman, 20012:266; Salusbury, 2003:24; Schlemmer, 1985:1; Pedersen, 2020:iii). Some cultural aspects that sociologists have mentioned of WESSAs are, that they have an individualistic (as opposed to communal), metropolitan, and largely middle- to upper-middle class culture (Salusbury, 2003:22, 24; Schlemmer, 1985:11;13). Salisbury (2003:22–23) states that about two-thirds of WESSAs are of Anglo-Saxon and Celtic descent. Others are of mixed Dutch, German, Portuguese, Greek, Jewish and Afrikaner descent (Schlemmer, 1985:4). According to Schlemmer (1985:4), WESSAs also include people from a mixed English and Afrikaans-speaking background who "preserve a dual identity".

4.4 Social Class in South Africa

The colonial history of South Africa has resulted in its class system being mainly based on race. For instance, Dubow (2011:34) states that from 1910, the British Union of South Africa began to develop an all-white nation, to the exclusion of "Africans, coloureds (those of slave, Khoesan, and mixed-race descent), and Indians as full citizens of the new nation-state" and was in the process of dividing the population into fixed racial hierarchies (Dubow, 2011:34). According to Glaser

(2001:115), during the twentieth century many African landowners lost their land in the market as well as through political circumstances, and African tenants on white farms had their status lowered to labour tenants and wage labourers, while African families in reserves became more and more dependent on wages and moved out of the countryside to cities. In the earlier stages there had also been class inequalities between white English and Afrikaans citizens. For instance, after the mining revolution (1970s), English-speaking South Africans dominated economically in urban areas as well as in the skilled working-class occupations, leaving the Afrikaans communities living in the cities to struggle financially (Glaser, 2001:166). Glaser states that part of the reason why the English community never really formed a common identity was because they consisted of the wealthiest businessmen and the working-class (2001:116). According to Glaser (2001:117), the Afrikaner community started closing the financial gap between themselves and the English from 1948, when they advanced in the industrial and commercial middle-class industries as well as in skilled labour. Glaser (2001:117) states that in the urban hierarchy of occupations (after the white occupations), Indians ranked at the top, tending to fill white-collar occupations, while the coloured community ranked below them, mainly in skilled working positions, and the black African community ranked last, mostly in unskilled labour. From the 1960s, the black African community moved into semi-skilled and skilled occupations, resulting in the development of a small black middle class.

4.4.1 Social class among WESSAs

As previously mentioned, sociological studies have revealed that WESSAs fall mainly into the middle to upper classes of South African society (Salusbury, 2003:70; Pederson 2020:26). Lanham and McDonald (1979:24), state that English-speaking South Africans had minimal representation among the labouring class, and those in the lower classes mixed frequently with the middle classes. This could be because of their early dominant role in politics during British colonial rule and later their mainly financial dominance. The findings of Lanham and McDonald (1979:25–26) reveal that Afrikaans ancestors of English-speaking participants increase the further one goes down occupational status, and the Afrikaans relatives or ancestors have decreased levels of education. This confirms the observation that WESSAs tend to occupy the higher social classes in the general hierarchy of South African society, while Afrikaans-speaking South Africans tend to occupy the middle to lower social classes.

Research has shown that social divisions in the country are much less dependent on someone's financial status and/or type of occupation and much more on shared cultural values and type of lifestyle. For instance, historically, white and black working-class miners formed separate groups and were often hostile to each other (Dubow, 2011:45–46). This is supported by Pederson's (2020:24) interviews in which the WESSA interviewees distinguish themselves from others in their group based more on a certain culture than on their financial status. Pedersen states that in South

Africa, the label of “middle class” seems to be used equally as a cultural and economic label (2020:24).

As an example, in an earlier study of WESSAs by Salusbury (2003:60), one respondent distinguishes his small group from other WESSAs who are equally well-off financially by describing them as different from the “nouveau riche” (*Ibid*.:60). Salusbury (2003:71) states that one respondent who moved to Ireland from South Africa at an early age felt out of place in a working-class Irish community, describing herself as middle-class, despite being cared for financially by charitable organisations after the passing of her father.

4.5 Historical Overview of South African English

According to Lass (2002:104), the origin of South African English (SAE) is the diffusion of Southern British English to the communities in SA. Lass states that the British settlers who arrived in the Eastern Cape (1820) were “urban working-class” (*Ibid*.:109; Branford, 1994:436; Gough, 1996:xvii), and that this group provided the main input for the development of a new local English variety called Cape English (CpE) (Lass, 2002:108; but see Bekker, 2012a:141 and Bekker, 2012b:129, n4, who indicates that the lower-class status of these first settlers has been questioned and that there is a possibility that they consisted of a wider variety of social classes – for instance, see section 4.3 above). The Natalians discussed in section 4.3 are said to have been of middle to upper-class status and developed another variety known as Natal English (NE) (Lass, 2002:109; Branford, 1994:434; Bekker, 2012a:141). Lass (2002:109) states that because of their higher social class, the Natal settlers brought the standard English variety into S.A., unlike the variety of the mostly urban working-class 1820 settlers. Bekker (2012a:141) suggests that a third variety developed during the formation of Johannesburg and included, as input, CpE, NE and other English varieties and could be the basis for the variety currently spoken by white South African L1 English speakers, known as WSAE. However, according to Bekker (2012a:141), the “standard” view is that the main influence on current WSAE were the inputs of the 1820s settlers in the Eastern Cape and the 1940s Natalians (e.g. Lass, 2002:109; Lanham, 1996:20). According to Lanham and McDonald (1979:18) another influence on WSAE throughout the years has been the “Afrikaner as English speaker” who contributed to the social meaning and norms of the accents of WSAE. They have also contributed to the attitudes that the SAE community hold towards these social accents. Lanham and McDonald (1979:26) state that the ethnic groups that WSAE speakers consist of are Afrikaners, Jews and the group of British origin.

Lass (2002:109) defines three “major lectal types” of WSAE which developed because of settlement history, internal evolution and ongoing close ties with Britain among the upper and middle classes. According to Lass (2002:109–110), these three types tend to be hierarchically ranked among speakers. The first type, which is modelled according to the British standard,

Received Pronunciation or RP, is called “Cultivated” or “Conservative SAE”. Examples of this type were news anchors of the South African Broadcasting Corporation (SABC) up to the 1990s. Lass (2002:111) states that this speech is typical of “first families” of the older urban cities such as Cape Town, English teachers and the upper-middle class.

Type 2 is called “Respectable SAE” (Lass, 2002:111) and described by Lass as “A new local standard” (*Ibid*.:110) which shares prestige with type 1 but is a more local variety and stigmatised by older, more extreme type 1 speakers. According to Lass (2002:111), examples of the kinds people who speak type 2 are, for example, (in the past) Democratic Party and English-speaking National Party politicians, university lecturers, doctors, accountants, and lawyers.

Type 3, which also bears the lowest social status, is “Extreme SAE” (Lass, 2002:111). This type is described as a group of local vernaculars which is highly stigmatised by both type 1 and type 2 speakers. This variety is linked to people with low social status, sparse education and non-professional occupations. Lass (2002:111) states that the more extreme an SAE variety is, the harder it becomes to differentiate it from second language Afrikaans English.

South African English is mainly distinguished from other L1 varieties on the grounds of pronunciation. According to Lanham (1996:20), the speech of the 1820 British settlers started to change in their new and unfamiliar environment and as their contact with Britian diminished over time. A detailed account of the changes from British to South African pronunciation is given by Lanham (1965:90–95). Considerations of SAE as a distinct variety on a lexical and grammatical level appeared later than the description on its phonology, for instance in *A Dictionary of South African English* (Branford, 1978); *English in South Africa* (Branford, 1994); and *A Dictionary of South African English on Historical Principles* (Silva, 1996).

4.6 Situating SAE Within the Linguistic Environment of South Africa

According to the 2018 *General Household Survey of Statistics South Africa* (Stats S.A., 2018:14), the distribution of home languages in S.A. is illustrated in

Table 4.1. The results of the Household Survey show that the three most spoken home languages in S.A. are: IsiZulu (25.3%), isiXhosa (14.8%), and Afrikaans (12.8%). English is listed as the sixth most spoken home language with 8.1% out of the total respondents from all racial groups.

South Africa	
Languages (inside home)	**%**
IsiZulu	25,3
IsiXhosa	14,8
Afrikaans	12,2
Sepedi	10,1
Setswana	9,1
English	8,1
Sesotho	7,9
Xitsonga	3,6
SiSwati	2,8
Tshivenda	2,5
Other	1,9
IsiNdebele	1,6
Khoi, Nama and San languages	0,1
Sign Language	0

Table 4.1: Distribution of home languages in South Africa (2018:14)

Table 4.1 shows the distribution of English and Afrikaans as home language among the white South African group:

Demographic	**White**
Language (inside home)	**%**
Afrikaans	61,2
English	36,3
Total respondents:	**4442**

Table 4.2: Main home languages among white South Africans spoken inside the home. (Stats S.A., 2018)

Among the white South African population, the 2018 Household survey shows that Afrikaans is the most spoken home language with 61.2% and English second with 36.3% out of the total respondents.

4.7 The Data Collection Site: A brief history of Port Elizabeth (Gqeberha)

A prominent location in the colonial history of South Africa is Algoa Bay, where the first recorded exploratory landing was made by Bartholomew Diaz in 1486 (see section 4.3 for the settler discussion and Figure 4.4 for the location of Algoa Bay). The city of Port Elizabeth is located in the western side of the Bay (Goschen & Schumann, 2010:6). After the British occupation of the Cape in 1795, several forts were built to protect the Dutch inhabitants of the area from Bantu and Khoekhoe raids. One of these forts was “Star Fort”, built in 1798 at Cradock Place (today it is a suburb of Algoa Park; Inggs, 1986:81, n14; Beavon 1968:160). Star Fort was attacked in early

1799 by the Bantu and because of such raids and the fear of a French invasion of the Eastern Cape, the British established a permanent military fort in Algoa Bay, completed in late 1799 and known as Fort Frederick. Fort Frederick was built to protect against Bantu raids and to prevent possible foreign arrivals to the coast. According to Beavon (1968:161), after the construction of Fort Frederick and before the arrival of the 1820 British settlers, the few buildings at the fort were mainly linked to the military. Beavon (1968:161) states that Fort Frederick quickly developed into a town directly after the arrival of the 1820 settlers, which later became Port Elizabeth. The British government's decision to develop Port Elizabeth as a port instead of Port Alfred aided in its rapid growth. The settlement "Port Elizabeth" was named in memory of Sir Rufane Donkin's wife, who was at that time the governor of the Cape (Beavon, 1968:162).

Figure 4.6 below provides a map of Main Street, from which the rest of Port Elizabeth developed. The Dutch immigrant Frederick Korsten (1772–1839) settled in Fort Frederick in 1812. He established a large trading business and had a monopoly on trade for a time and his contributions influenced the development of the city (Inngs, 1986:80–81).

Table 4.3 displays information from Hill (2020:1114) which contains selected subplaces from the 2011 census data and includes the order of development of these places by historical year.

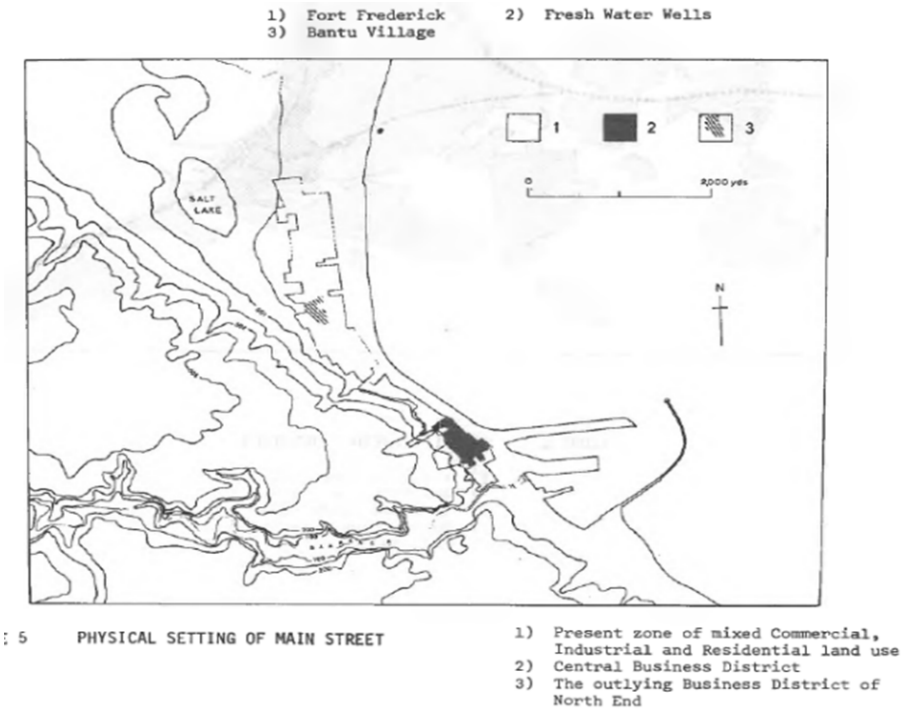


Figure 4.6: Map of Main Street in Port Elizabeth, 1849 (Beavon, 1968:165)

Hill (2020:1114) traces the early development of the British settlement to the streets of Govan Mbeki Avenue (formerly Main Street, see Figure 4.6) and Russel Road (located in Port Elizabeth Central). The town was divided into the areas of South End and North End, the earliest suburbs,

and according to Hill (2020:1114) these two places became the first measurement of social class distinction. For example, Beavon (1968:168) states that South End developed as a non-white area with poor quality houses and shops, while North End had a rise in land values, which stimulated commercial growth between Market Square and Russel Road. The area between North End and Russel Road tended to consist of low-class housing and was inhabited by the Irish and Malay lower classes (Beavon, 1968:168). According to Hill (2020:1114), the street of Cape Road, which runs through Port Elizabeth Central, represents the expansion of the town to the west where, up to the present, it mainly consists of wealthy white inhabitants. Beavon (1968:168) states that The Hill (presently a residential area in Mill Park, Port Elizabeth Central) and St George's Park developed as high-class neighbourhoods (which later became more accessible through public transport and changed to mixed middle- and upper-class residences (Beavon, 1968:167–168). Furthermore, Humewood was a residential area some distance away from the business centre but developed as a coastal resort (Beavon 1968:168). See Figure 4.7 which provides context for the above discussion.

Subplace	Date of origin
Bethelsdorp	1803
Uitenhage	1804
North End	1820
Port Elizabeth Central	1820
South End	1820
Walmer	1855
Gqebera	1896
New Brighton	1903
Newton Park	1907
Fairview	1907
Despatch	1942
Kabega	1951
Malabar	1951
KwaZakele	1956
Algoa Park	1956
Zwide	1968
Motherwell	1982

Table 4.3: Subplaces from the 2011 census data (Hill, 2020:1114)[32]

The area between North End and Russel Road tended to consist of low-class housing and was inhabited by the Irish and Malay lower classes (Beavon 1968:168). According to Hill (2020:1114), the street of Cape Road, which runs through Port Elizabeth Central, represents the expansion of the town to the west where, up to the present, it mainly consists of wealthy white inhabitants.

[32] Note that in the list, Gqebera is a subplace near Walmer and does not refer to the newer name for Port Elizabeth, Gqeberha.

Beavon (1968:168) states that The Hill (presently a residential area in Mill Park, Port Elizabeth Central) and St George's Park developed as high-class neighbourhoods (which later became more accessible through public transport and changed to mixed middle- and upper-class residences (Beavon 1968:167–168). Furthermore, Humewood was a residential area some distance away from the business centre but developed as a coastal resort (Beavon 1968:168). See Figure 4.7 below, which provides context for the above discussion.

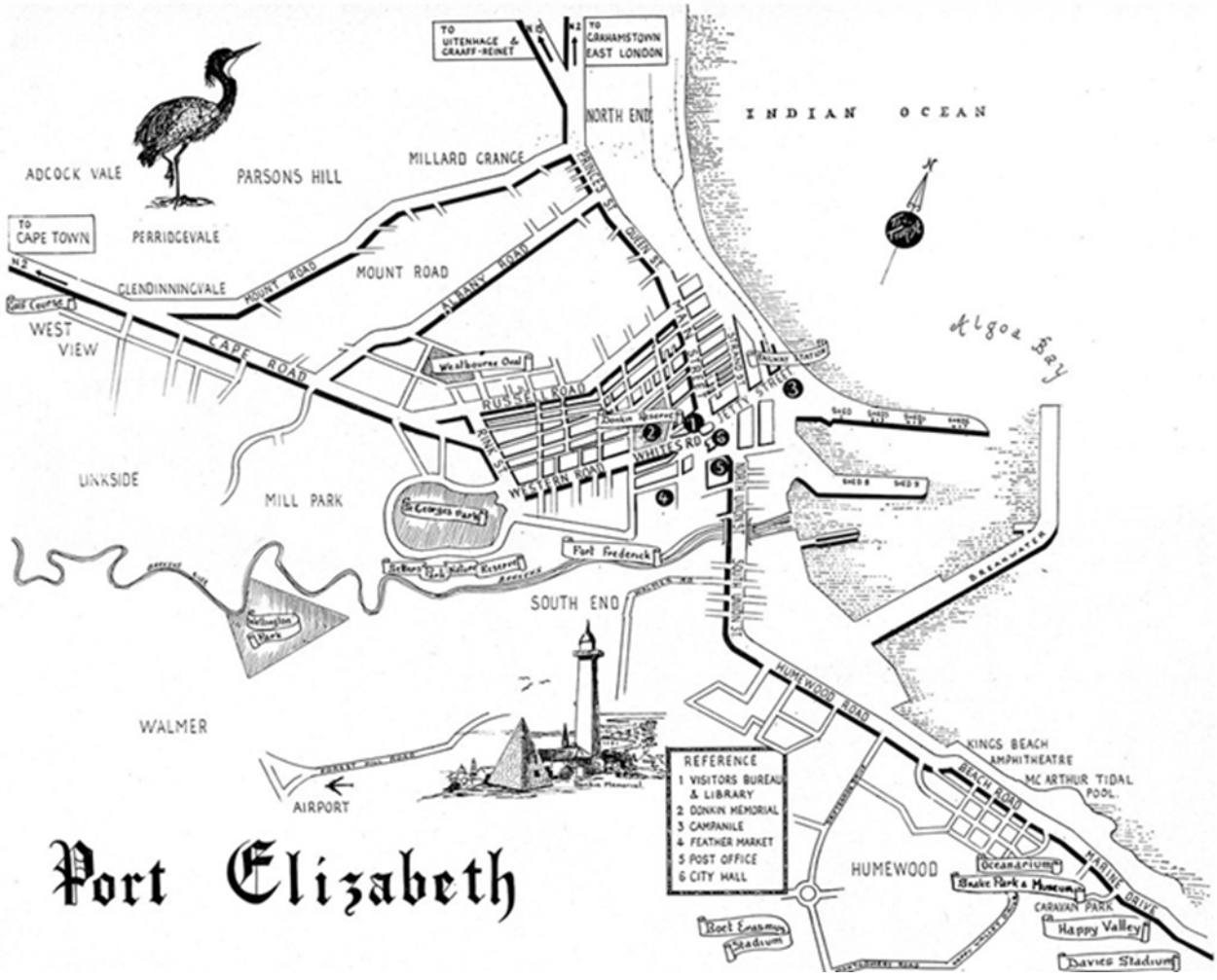


Figure 4.7: Port Elizabeth street map 1965/6 (Hilton, 2011)

Based on the origin of Port Elizabeth as a British settlement, it is apparent that it started out as a town in which the English inhabitants were loyal to England. Terblanche (1993:101) states that by 1921, one out of five white people in Port Elizabeth were born in Great Britain. According to the census of 1921, the white group in Port Elizabeth were 27,236 in total. Out of this number, 5 003 were born in the U.K. Of the number of people born in the U.K., 3 524 were from England, 1 035 from Scotland, and 400 were from Ireland and Wales (*Ibid*.:101). Port Elizabeth retained a strong British identity longer than many others in S.A. For instance, Lambert (2014:608) states that up until the 1990s, Union Jacks could be seen on the municipal building in Port Elizabeth on days that had royal and British South African importance, similar to Durban and Pietermaritzburg

in Natal. However, Lambert (*Ibid*.:608) notes that from the 1990s onward, such British nostalgia was declining in South Africa. Figure 4.8 shows the location of all the subplaces in Port Elizabeth/Gqeberha.

Figure 4.8: Suburbs of Port Elizabeth (RoomsForAfrica, 2024)

The following Google Maps links show the areas of Port Elizabeth as they are at present:

Govan Mbeki Avenue: Govan Mbeki Avenue

Russel Road: Russel Rd

Market Square, now Vuyisile Mini Square: Vuyisile Mini Square

The Hill, now a residential area in Mill Park: The Hill

Cape Road: Cape Rd

Hill (2020:1116) states that Port Elizabeth was a predominantly Anglophone town until the early twentieth century, after which it became more complex in its demographics. For example, the population grew from 13,000 to 41,000 between 1875 and 1911. In 1875, the town consisted of 70% British settlers and 15% consisted of Mfengu- and Xhosa-speaking Africans. According to Hill (2020:1122) until 1994, many of the inhabitants of the Western suburbs of Port Elizabeth

(such as Kabega, Linton Grange, Framesby and Cotswold) were mainly white Afrikaans-speaking people. In contrast, higher numbers (75% or more) of first language, white English speakers are found in the more economically well-off suburbs such as Greenacres and Linkside.

4.8 The Present Geographical Boundaries of Port Elizabeth/Gqeberha

The city of Port Elizabeth (now Gqeberha) is part of the Nelson Mandela Bay Metropolitan Municipality (NMBM), established in 2000). According to the NMBM 2020 profile of the Department of Cooperative Governance and Traditional Affairs (COGTA 2020:6), Port Elizabeth is South Africa's second oldest city. The NMBM spans an area of 1,950km^2. Figure 4.9 shows the approximate distances of the area to other major cities (image from COGTA, 2020:6).

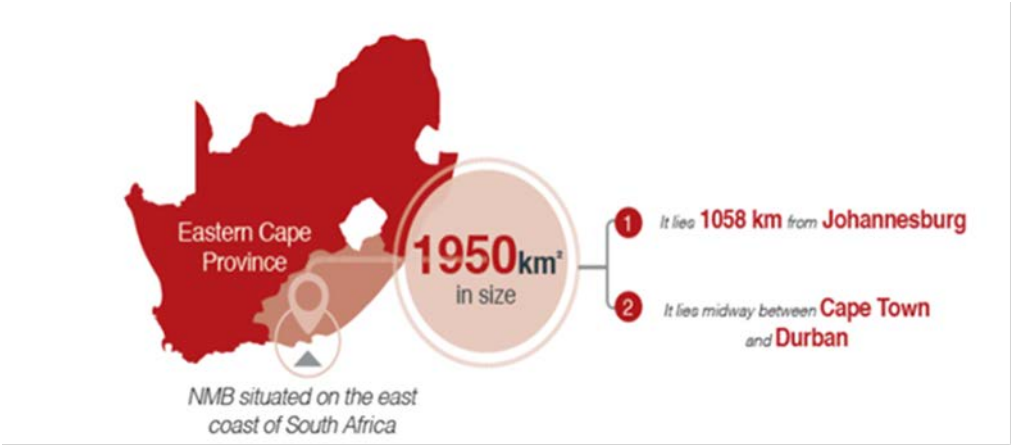


Figure 4.9: General location of NMBM in South Africa

Figure 4.10 indicates the boundaries of the NMBM, along with an indication of the distribution of race and main home languages in each area. The map was created by me in ArcGIS, using the Allotment boundaries shapefiles (2023), available from the Nelson Mandela Bay Metropolitan website: https://www.nelsonmandelabay.gov.za/documentslist?catID=66. The colour coding was done by me according to the area specific demographic and home language data from the Census 2011 Community Profile Databases by Statistics South Africa, made available by Firth, 2013: https://census2011.adrianfrith.com/. The legends for the map were created by me according to the purpose of the study. shows which areas in the NMBM were mostly inhabited by WSAE speakers at the time of the 2011 census and could be an indication of their approximate distribution across the metropolitan at present. Apart from the historical classification of social class for some of the suburbs, referred to in section 4.7, the researcher has not found more recent classification of the NMBM areas according to social class.

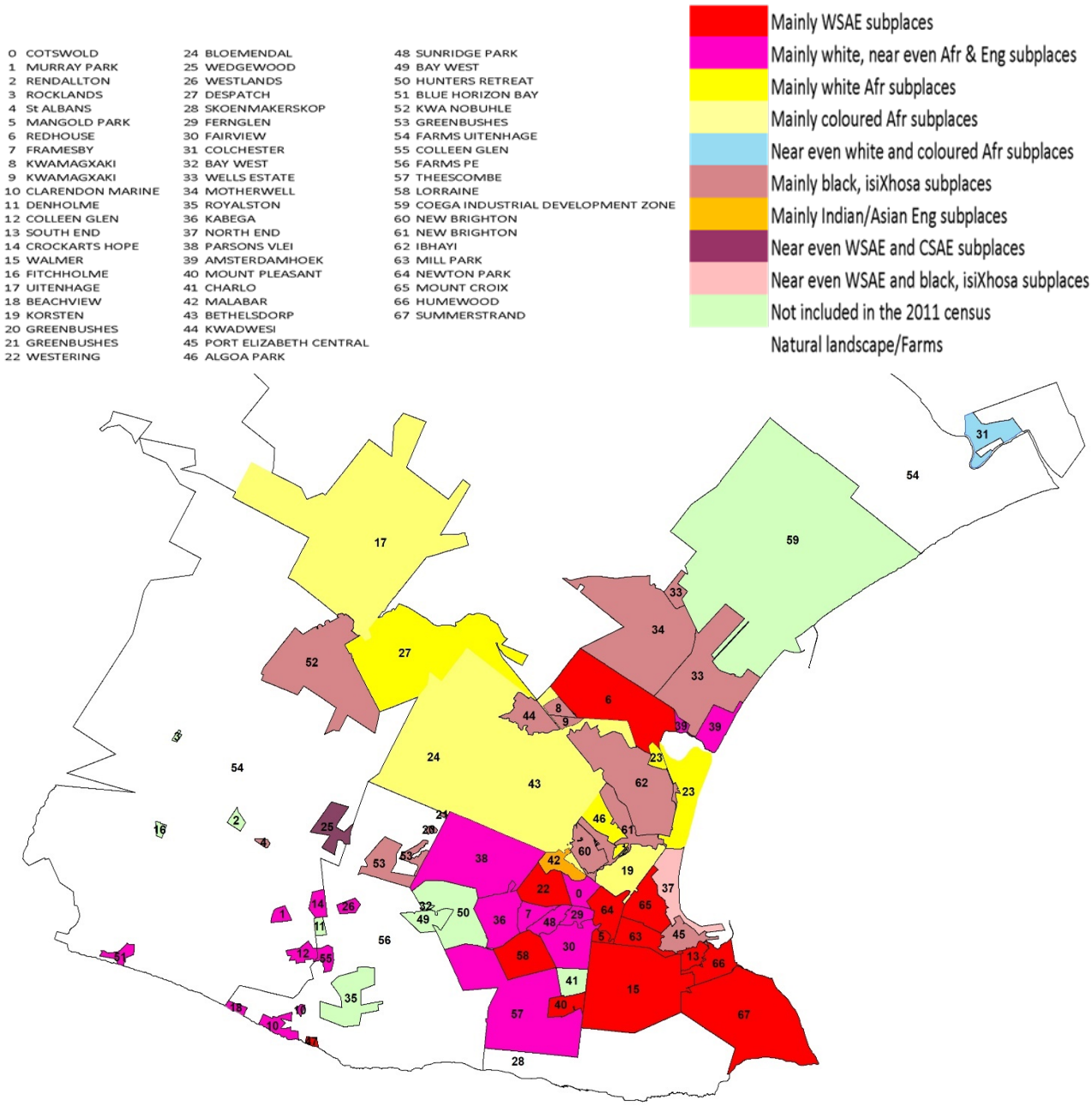


Figure 4.10: Distribution of the main home languages by area in the NMBM

Figure 4.10 indicates the various groups of L1 or first language speakers in Port Elizabeth and what their languages are. Yellow indicates the mainly white Afrikaans speakers, located in Algoa Park (46) and Swartkops (23, not shown in the legend) and Despatch (27). The purple areas indicate the suburbs that have mainly white, near even (where the percentages for specific language and ethnic group combinations indicated similar levels) Afrikaans and English speakers. The areas coloured in red show the places where mainly white, English speakers are found.

The participants in this study resided in the following places at the time of the interviews (refer to Figure 4.10 for the location of their neighbourhoods). All six of the upper middle-class speakers had only English-speaking parents and therefore did not grow up in a bilingual Afrikaans–English household. All of them reported little to no fluency in speaking Afrikaans. These speakers are

from Bluewater Bay (in Amsterdamhoek, area nr. 39), Mount Pleasant (40), Fernglen (29), Summerstrand (67) and Parsons Hill (in the Walmer suburb, nr. 15).

The nine middle-class speakers reported a range in which they can either speak very basic Afrikaans, have it as a second language or are fully bilingual in Afrikaans and English. These speakers live in the areas of Lorraine (58), Walmer, Westering (22) and Kabega (36). Of these speakers, two grew up in bilingual Afrikaans and English households in which one parent is Afrikaans and another one is English. The third bilingual participant stated that he grew up in an English household (both parents are mainly English speakers but are bilingual in Afrikaans and English). His parents are both from Afrikaans families but went to English schools and brought their children up as English speakers. As seen from Figure 4.10, all of the areas in which the middle-class speakers reside are also mainly white English-speaking suburbs.

Of the five lower-middle class speakers, three are from Lorraine, one from Walmer and one from Mount Pleasant. Most of these speakers reported only a little knowledge of Afrikaans while one speaker said that he is fluent. All of these participants reported that both their parents are English speakers. Taken together, the two areas most represented by the participants are Lorraine and Walmer. As indicated in Figure 4.10, all of the speakers live in the mainly white English-speaking areas.

4.9 Conclusion

In conclusion, this chapter provided a brief history of the arrival of the Dutch and English at the Cape and the demographic and linguistic developments of the Cape Colony, from Dutch control to British control. It briefly described the development of Dutch to Afrikaans during the period from 1652 up to 1925 when Afrikaans was recognised as an official language in the Union of South Africa. It then reviewed the arrival of three waves of English settlers: the first arrivals in 1820 to the Eastern Cape, who influenced the development of a Cape English variety; the second wave of arrivals in 1849 to Natal, who introduced standard English to South Africa; the third wave of arrivals in the 1870s with the discovery of gold and diamonds in Johannesburg. An overview of the major lectal types of WSAE was provided, these being Conservative SAE and Respectable SAE which enjoys a higher social prestige among speakers than the Extreme SAE variety which is heavily stigmatised. Next the chapter provided a linguistic overview of all the languages of South Africa and where English fits into this context. The development of Port Elizabeth from a British military fort to a city was discussed, along with important suburbs and streets that played a role in its development. Lastly, the linguistic composition of Port Elizabeth was discussed, as well as the areas in which the participants for this study reside – this illustrated that all of the participants live in mainly white English-speaking suburbs, regardless of their social status.

Chapter 5 presents the analysis and results of the study. First the results of quotative *be like* is presented in terms of its frequency vs other quotatives overall. This comparison of usage frequency aimed to discover which of the quotatives were used the most out of the entire speech community. The next sections compared the frequencies of the quotatives for the 18–20 age group vs the 33-48 group. The aim was to discover if the younger or older speakers use *be like* the most. Frequency comparisons were also done for gender and social class respectively, which were aimed at discovering if men or women in the speech community use *be like* the most and what the patterns of the other quotatives are for each of the social groups compared to *be like*. The results for combined social groups were also discussed, such as the gender distribution of the quotatives within the younger and older age groups distributed across gender. This comparison aimed to discover if the gender distribution differs when the younger and older age groups are isolated as compared to the overall gender distribution. The frequencies of the quotatives were also examined with regards to the gender distribution within the three refined social classes of upper middle, middle and lower middle class. This was done to discover if there were any differences in the quotative usage of men vs women as compared to the overall gender distribution in the speech community as a whole. The frequency analysis of the quotatives was concluded with a c-tree analysis of the results.

The second part of the quotative analysis consisted of the analysis of the linguistic factors influencing the use of *be like*, such as the analysis of the content of the quote (thought vs speech reporting), grammatical person, and tense. This analysis was done to discover if *be like* was used most with speech or thought reporting, with first- or third-person subjects, and within present or past tense contexts. These linguistic factors were also analysed with regards to the social factors of age, gender and social class, for example, to find out if men or women use *be like* more with speech or thought reporting. This section was concluded with a c-tree analysis which combined the internal and social factors in order to discover which factors have the strongest influence on the use of *be like*.

Chapter 5 also presented an analysis of DM-*like*. The first part consisted of a frequency analysis in order to discover how frequent DM-*like* is in the dataset compared to all the words in the corpus. The second section aimed to discover if men or women, younger or older and upper-, lower-, or middle-class speakers used DM-*like* the most. This discussion was concluded with a c-tree analysis of the social factors which possibly influence the use of DM-like in the speech community. The internal factor of the structural position of DM-*like* was examined to discover if it is used the most in either clause-medial or clause-initial position. The analysis was expanded by analysing the age, gender and social class distribution of the two structural positions of DM-*like*. This analysis was done to discover which of the social groups use DM-*like* in clause-medial or clause-initial position the most. The analysis of DM-*like* was concluded by combining the linguistic

(clause-initial vs clause-medial position) and social factors that might have an influence on its use in the community.

CHAPTER 5: RESULTS AND DISCUSSION

This chapter reviews the results for quotative *be like,* as in:

(103) The one guy had like a knife and he **was like**, "You better give me like all your stuff." (<$N><s n=142>)

The sociolinguistic interviews of 19 native inhabitants of Port-Elizabeth/Gqeberha, aged between 18–59 are analysed. First, *be like* is examined within the quotative system by comparing its frequency to the other quotatives (such as *say* and *think*). Second, the social factors of age, gender and social class will be examined to find out which of these factors influence its usage. Third, the language internal factors of quote content, grammatical person and tense will be examined to determine how these factors influence its use in the speech community. The individual speaker will not be examined as a factor but outlier individuals will be discussed where relevant. The analysis will first provide proportional results and then a c-tree analysis in *R*.

5.1 *Be like* vs other quotatives

Error! Reference source not found. provides the specific percentages and Ns for each quotative out of the total quotatives in the dataset. Figure 5.1 represents the same proportions graphically.

Quotatives	%	N
be like	46	196
say	25	107
zero	9	38
think	4	18
go	4	18
tell	4	17
ask	2	8
be	2	7
decide	2	7
get like	0	1
other	3	13
Total		**430**

Table 5.1: Overall distribution of the quotatives in WSAE

While the limited sample of speech collected for this study is not enough to generalise to the entire community, it does suggest that, similar to more recent studies of the quotative system in English, (e.g. Tagliamonte & D'Arcy, 2004:501, CanE; Barbieri, 2009:83, AmE; Rodríguez-Luoro, 2013:59, AusE) *be like* (46% out all the quotatives) has overtaken *say* (25%) as the main choice for introducing constructed dialogue in WSAE. An example of *say* in the data is:

(104) Um, and he **said**, "You must just shake it and pass it on." (<$A><s n=43>)

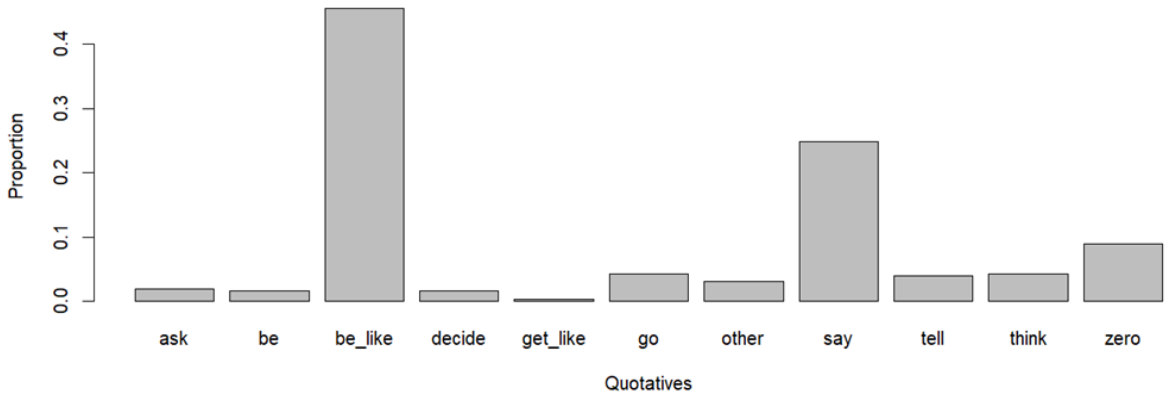


Figure 5.1: Overall distribution of the quotatives[33]

The third most frequent quotative in the sample is the zero quotative, where no verbal introducer is used before a quote, which consists of 9% of the total quotatives, as in:

(105) He's also the one that tends to forget. “Oh, I forgot my blazer at school.” (<$L><s n=50>)

The quotatives *think*, *go* and *tell* have equal frequencies (4%) (see below for examples). Their frequencies are, however, much lower than *be like*, *say* and *zero*.

(106) So we **thought**, "Well he's leaving now so play one last prank" (<$P><s n=38>)

(107) I mean, I can hold their hand in public and **go**, "This is my mom, this is my dad.” (<$M><s n=64>).

(108) Then they **tell** me, “Give it a couple of years and you'll miss it”. (<$H><sn=106>)

However, *go* is an anomaly in the sample because it is used mostly by one speaker, Katie, a 19-year-old middle-class female who uses it 17 times. The one other time it occurs, it is used by another 19-year-old female speaker, this time from the upper class:

(109) She says that I walked into the house and I **went**, "It's so quiet” cause she has one sibling and it's an older sister (<$P><s n=54>)

Thus, *go* seems to be active in WSAE, but to what extent is unclear because it is rare in the speech sample. In a sample of 20 WSAE speakers between the ages of 16–35, Horn (2022:69) found that *go* is relatively infrequent. She does not give exact percentages, but the proportional estimate for *go* is less than 10% (*Ibid*.:65, 68–69). This result seems to suggest that *go* is not a preferred choice of quotative for WSAE speakers. Horn (2022:69) obtained a higher percentage

[33] The “other” category consists of verbs introducing quoted dialogue that are rare in the sample. These are verbs such as: *scream, know, getting, calling and remember*.

for *be like* (over 60% vs 46% in my data), while *say*, *think* and zero have a similar result: over 20% vs 25% for *say*, around 10% or more for zero vs 9% and 5% or less vs 4% in my data. This suggests that WSAE speakers in the Cape Town region have a similar pattern regarding *say*, zero, and *think* but that the Cape Town speakers tend to use *be like* more than those in Port Elizabeth.

The full inventory of quotatives that rarely appear in the sample, from the "other" category, are the verbs *perceive*, *call*, *get*, *know*, *look*, *scream*, *do*, *remember*, *threaten*, *fangirling, notification, will*. The quotative verbs *ask* and *decide* appear somewhat more frequently in the sample than those in the "other" category. Examples of these quotatives are:

(110) I could be **perceived** as like "Bru, I'm not actually listening to what you're saying and like my words are more valuable than yours." (<$R><s n=129>)

(111) I was **screaming** like "Help me." (<$C><s n=72>)

(112) But whenever it gets nighttime, like always I stand there calling her, **calling** her, "Come, where are you?" (<$C><s n=70>)

(113) …they would always wanna **know**, "Where, where are we, what are we, when are we coming home?" (<$L><s n=43>)

(114) we're **looking** like "Who's this man?" <$G><s n=84>)

(115) I just go to school and then, you know, and then **getting**, "Happy birthday" and like maybe, like one or two of my friends will bring me like a present. (<$O><sn=122>)

(116) …so the doctors there had to train and **do** "Okay, this, he needs this, he needs an IV and he needs oxygen." (<$I><s n=68>)

(117) the school got **notification**, "Okay you're under 14B boys are playing against, uh, Daniel Pienaar and Uitenhage this afternoon." (<$K><s n=54>).

(118) It would have been the first time where they would actually **remember**, "Oh okay, that's my great-grandmother." (<$L><s n=65>)

(119) My mom would kind of just **threaten** me, "I'll take my slop off" or whatever, dadada. (<$I><s n=49>)

(120) they still do like shows and stuff so you can meet them easily. But no one like fan, like **fangirling**, "Oh my gosh!" (<$C><#><s n="65">)

The verb *be* without *like* is used 7 times in the sample and functions in the same way as *be like* in that it can introduce both thought and spoken dialogue and has been suggested to be an abbreviated form of *be like*. This use is illustrated in example (121):

(121) but uh, but, ja then, it **was** "I'm not your friend anymore" and, ja. (<$D><#><s n=62>)

However, this analysis follows previous studies in separating *be* from *be like* because it is said to interact with social factors such as gender (e.g. Macauley, 2001:11). There is one instance of *get like,* which was used by an upper middle-class female speaker (Gabrielle, 19):

(122) I, I feel like I'm that person. And I **get like**, "Eee get out of there!" (<$C><sn=73>)

This quotative has been noted in Romaine and Lange (1991:247) and is described by Vandelanotte and Davidse (2009:800) as being related to *be like* since it is one of the copulative verbs combining with *like* to form a quotative introducer (along with *go*). However, other than such brief mentions, it has not been discussed in the literature. It was separated from the other quotative *get* and thus not included in the "other" category, because unlike in example (f) above, where "get" retains its semantic meaning of receiving something, *get like* functions on a more pragmatic level, in the same way as *be like*. For instance, Vandelanotte and Davidse (2009:779) state that quotatives such as the verbs *be* + *like* and *go* are innovative quotatives (IIQs) since they lack any semantic reporting sense and are formally intransitive verbs. *Get like* fits this description as well since in examples such as:

(123) Suddenly it got cloudy and we couldn't **get** our bearings (OED, III.30.b.i)

Get is described as an intransitive verb, with the meaning of "to become" (OED. III.30.b), while in its quotative use it does not contain any meaning related to reporting speech or thought such as *say*/*think* has. This could be a newer pragmatic quotative that is emergent in WSAE.

When comparing the results for quotative distribution in Figure 5.1 to the most recent results for international varieties such as Australian English (Rodríguez-Luoro *et al.*, 2020:336), Trinidadian English (Deuber *et al.*, 2020:8), Canadian and New Zealand English (Tagliamonte *et al.*, 2016:831), British English (Buchstaller, 2015:463), African American Vernacular English (Cukor-Avila, 2012:636), American English (Barbieri, 2009:83), it becomes apparent that the same repertoire of quotatives is used across varieties: *be like, say, zero, go, think*, and the "other" category which usually contains the graphic introducers (see section 2.7.4).

In the current dataset as well as in the international varieties mentioned above, the quotatives preferred the most are: *be like*, *say* and the zero quotative. In the study of Trinidadian English (Deuber *et al.* 2020:8; ICE-Corpus, recorded 2006–2011) and Tyneside, EngE, (Buchstaller 2015:463; recorded 2007–2011), *say* was the most frequent, followed by *be like* and zero was the third most frequent, while in the studies for Australian English (Rodríguez-Luoro, 2013:59; Rodríguez-Luoro *et al.*, 2020:336; recorded, 2011–2015) and American English (Barbieri, 2009:83; recorded, 2004–2005) *be like* is the most frequent, followed by *say* and the zero quotative. The AmE data, being based on the earliest recordings, implies a more advanced stage of diffusion for *be like* in comparison to the other dialects mentioned, since it is the dialect from which it originates. The next section investigates the quotative distribution for the younger age group and older age group in the dataset separately. For the quotative analysis of the social groups, only the results for *be like*, *say*, *zero* and *think* are provided while all other quotatives are placed in the "other" category since they do not provide additional insight into *be like* usage.

5.1.1 Quotative distribution across age

Table 5.2 and Figure 5.2 show the un-normed distribution of the various quotatives in the sample for the 18–20 and 33–48 age groups. Percentages were calculated out of the total quotatives.

Age	be like	%	say	%	zero	%	think	%	other	%
18-20	178	**41**	67	**16**	30	**7**	11	**3**	55	**13**
33-48	18	**4**	40	**9**	8	**2**	7	**2**	16	**4**
Total										**430**

Table 5.2: Un-normed quotative distribution across age

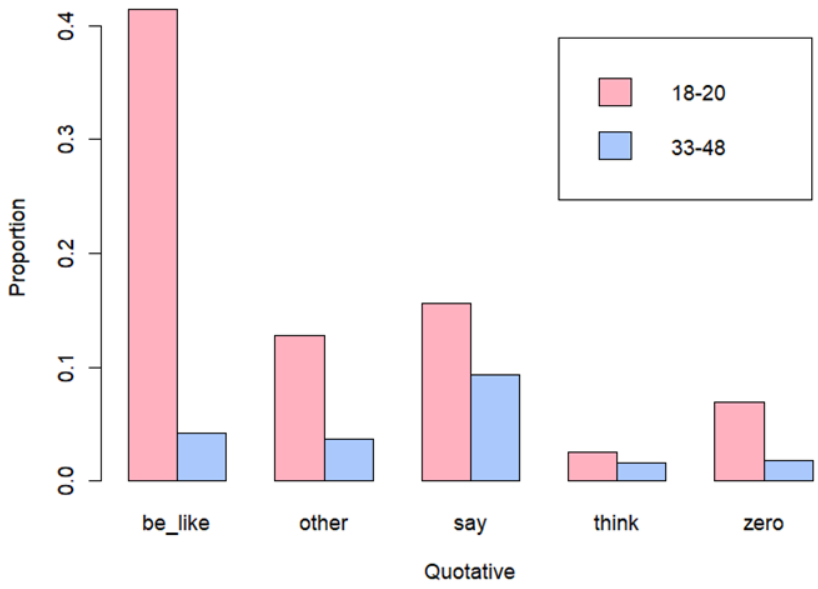


Figure 5.2: Un-normed Quotative distribution across age

The overall results indicate that the younger group uses *be like* much more frequently than all the other quotatives. The frequency of *say* is 25% less than *be like*, indicating that the quotative "territory" in young people's speech is being taken over by *be like* and it is replacing *say*. However, they also use the *zero* quotative and *think* to a lesser degree. The results indicate that the older group prefers *say* over *be like*. The older speakers also use the zero quotative and *think* to a lesser degree.

Table 5.3 shows the normed distribution of the quotatives for the 18–20 and 33–48 age groups. The data was normed by calculating the percentages out of the separate totals for each age group. Figure 5.3 provides the normed results from Table 5.3 for easier comparison.

	18-20			33-48	
	Quotative	%	N	Quotative	%	N
	be like	**52**	178	be like	**20**	18
	say	**20**	67	say	**45**	40
	zero	**9**	30	zero	**9**	8
	think	**3**	11	think	**8**	7
	other	**16**	55	other	**18**	16
Total			**341**	**Total**		**89**

Table 5.3: Normed quotative distribution across age

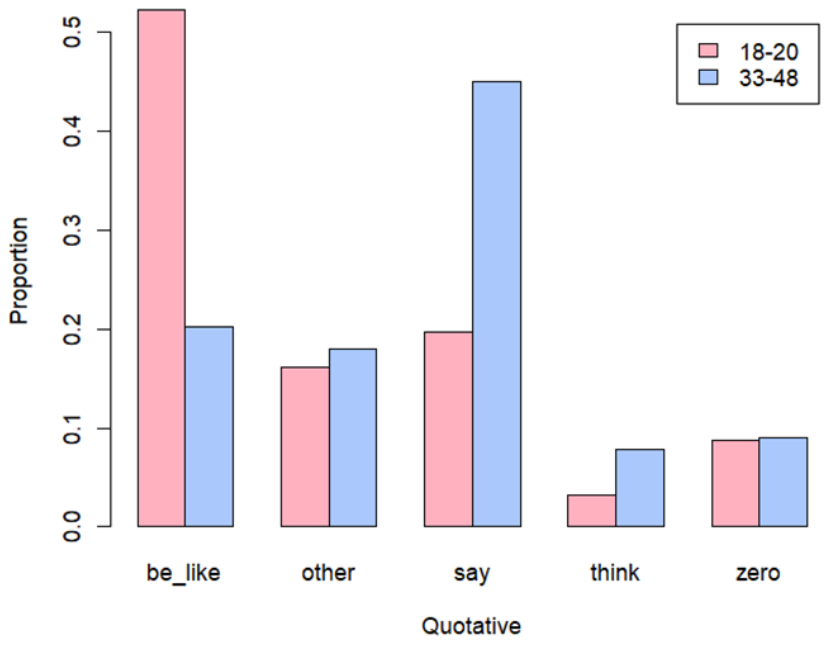


Figure 5.3: Distribution of the quotatives across age

Firstly, compared to D'Arcy's (2012:350) longitudinal investigation of the quotative system in NZE in which the oldest recordings (1946–1948) with speakers born between 1851–1910 mainly used *say* as a quotative (89%), followed by zero (8%) and *think* (1%), it is apparent that the quotative repertoire in WSAE has also expanded and now includes not only those three forms, but also *be like* for older speakers as well as *be* on its own, which makes up 6% of the total quotatives (not

displayed in Table 5.3). In addition, the results show that the younger and older speakers use the zero quotative at an even frequency (percentages were rounded).

Furthermore, Table 5.3 indicates that younger speakers of WSAE use *be like* more than older speakers. This result is in accordance with all previous studies investigating the quotative system: *be like* is associated with young people (e.g. Ferrara & Bell, 1995; Daily O'Cain, 2002; Singler, 2001; Buchstaller and D'Arcy, 2009; Rodríguez-Luoro *et al.* 2020). Tagliamonte and D'Arcy (2004:510) address the question of age grading and state that in their data for the 10–19-year-olds, the highest frequency of *be like* is found with adolescents (mid-teens). However, they state that a high frequency is maintained with first-year university students and that recent research such as Singler (2001:268) suggests that the use of *be like* has expanded to speakers in their 30s. Table 5.3 suggests that *be like* is accelerating in frequency in the speech of older speakers of WSAE since quotative *be like* is not limited to the younger speakers – it is the second most frequent quotative for the older speakers. This result implies a real change in progress regarding the spread of *be like*. For example, of the five older participants, all three in the 33-40 age range produced instances of *be like* (Simon, M, 33: 8; Melissa, F, 37: 4; Greg, M 40: 4, Lily, F, 40: 3). Only the oldest speaker Ben, M, 48, produced no instances of *be like,* but produced 9 instances of quotative *say* and one of *think.* This speaker also produced the least amount of quotatives (N = 10) compared to the other older participants.

Not shown in the displayed results, the frequencies of *think*, *be* and *tell* are higher in the older age group vs the younger age group: *think* and *be* is 5% more frequent; *tell* is 2% more frequent, while zero and *ask* have the same frequency in both groups: 9% and 2% respectively. Thus, the higher frequencies of *think*, *be* and *tell* for older speakers seem to support Tagliamonte & D'Arcy's (2004:510) argument that an increasing quotative repertoire as age increases might explain the decreased rate of *be like* in older speakers. However, the younger speakers use newer innovations such as *get like*, and use more graphic introducers represented by the "other" category (8% vs 4%). Thus, it might be the case that older speakers have a richer repertoire of quotatives to choose from, but the younger speakers are also expanding the repertoire further. In the older age group, the results reflect earlier studies of *be like* in AmE where *say* was the main quotative. However, as has been mentioned previously, *be like* seems to be gaining ground among older speakers in WSAE. Only time will tell if *be like* increases even further with older speakers, becoming just as frequent as with younger speakers to the point of overtaking *say*, or if the increase will reach saturation at some point.

Buchstaller's (2015:467)[34] study of EngE speech recorded in Tyneside between 2007–2009 and 2011–2013 is illustrated in Table 5.4 and sheds further light on the issue of age-grading. In

[34] Buchstaller (2015:461) studied the *Diachronic Electronic Corpus of Tyneside English* (DECTE), "a

Buchstaller's data, the frequency of *be like* increases from 2007–2009 to 2011–2013 by 13% in the 30–34 age group; the total amount of quotatives for this age group also increased. In the 35–40 age group, however, it decreased by 15%. In the last two age groups, its frequency was already low in 2007–2009 and decreased even further by 2011–2013.

2007-2009				2011-2013			
Age	%	N	Total	%	N	Total	Age
30-34	**19**	21	112	**32**	77	241	30-34
35-40	**21**	31	151	**6**	9	160	35-40
41-45	**5**	13	249	**0**	1	382	41-44
46-60	**3**	8	293	**2**	4	214	45-59

Table 5.4: Distribution of be like across older age groups (Buchstaller, 2015:467)

Thus, by 2011–2013, an increase in *be like* in EngE has spread to the 30–34 age group, indicating that it is advancing in the older age groups, compared to the earliest studies, such as Blyth *et al.* (1990:219) who commented that it was never used in their corpus with speakers past the age of 38 and in Ferrara & Bell's (1995:275) 1990 corpus, where it did not appear at all in their 40+ age group.

5.1.2 Quotative distribution across gender

Table 5.5 shows the un-normed results of the distribution of *be like* across gender. Figure 5.4 displays this information graphically.

Gender	be like	**%**	say	**%**	zero	**%**	think	**%**	other	**%**	
Female	160	**37**	54	**13**	15	**3**	10	**2**	48	**11**	
Male	36	**8**	53	**12**	23	**5**	8	**2**	23	**5**	
Total											**430**

Table 5.5: Un-normed quotative distribution across gender

collection of sociolinguistic interviews conducted in Newcastle and its surrounding areas since 2007". Buchstaller states that the thirty to thirty-four age group consists of four speakers, while the older age groups consist of seven to nine speakers (Buchstaller, 2015:467, 487, n8).

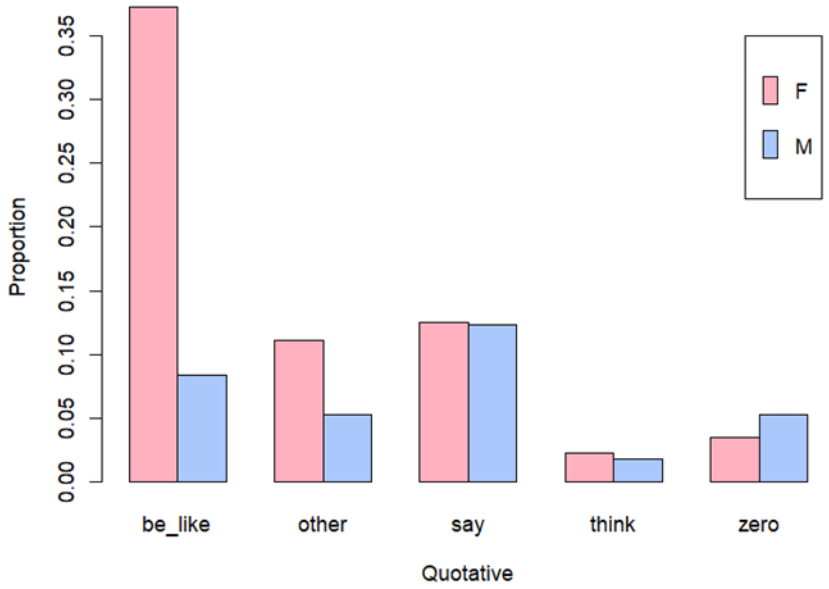


Figure 5.4: Un-normed quotative distribution across gender

In previous studies of various English varieties, gender results have been mixed (see section 2.7.4.4). The data illustrates that the women use quotative *be like* (37%) more than *say* (13%). This is similar to the results for the younger age group. On the other hand, the men use *say* (12%) more than *be like* (8%), which is similar to the results for the older age group. Table 5.5 also indicates that the men use the zero quotative (5%) more than the women (3), which might imply that this quotative is usually preferred by men in WSAE.

The other quotatives are used too infrequently to judge gender distribution, however, in the sample, *be* on its own (not displayed in Table 5.5) seems to be preferred more by the men than the women. Table 5.6 presents the normed results for gender distribution (see also Figure 5.5). It indicates that women use *be like* much more than men (56% vs 37%) in WSAE. This is an expected result since it is a well-established fact that women lead in the use of linguistic innovations (Labov, 1972:301).

Horn's (2022:58, 87) obtained a similar result for old and young WSAE speakers in Cape Town. This means that WSAE speakers are following the trend in dialects such as CanE and NZE and some regions in EngE in which it was found that women were leading in the spread of *be like*. It is, however, not similar to the tendency in AmE for the gender effect to be neutralised (see Chapter 2, section Social factors: Gender). The normed results indicate that the zero quotative is indeed used more by men than women. The men also use *think* (6%) more than the women (3%). For some of the more infrequent quotatives not displayed in Table 5.6, the results show a tendency for male speakers to use *be* on its own more than women as well. When compared to

the older group (see Figure 5.3) who use *say*, *think*, and *be* more (than the younger group), it becomes clear that the men's pattern in this regard is similar. This suggests that the men follow an older quotative pattern compared with the women who tend to pattern more similarly with the younger and thus more innovative group.

Females		
Quotative	%	N
be like	**56**	160
say	**19**	54
zero	**5**	15
think	**3**	10
other	**17**	48
Total		**287**

Males		
Quotative	%	N
be like	**25**	36
say	**37**	53
zero	**16**	23
think	**6**	8
other	**16**	23
Total		**143**

Table 5.6: Normed quotative distribution across gender

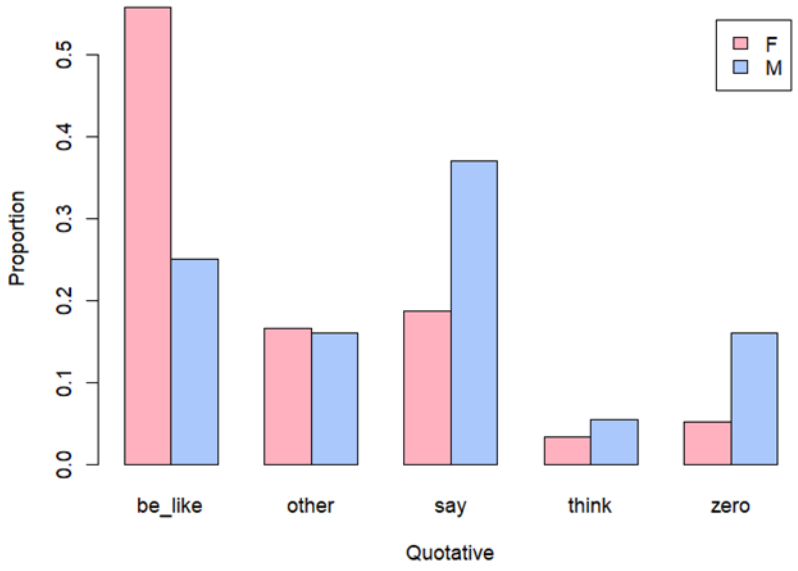


Figure 5.5: Normed[35] quotative distribution across gender

In diachronic studies of the gender distribution of be like, it has been shown that not only women, but men below the age of 40 are also increasing their frequency of use over the years (e.g. Ferrara & Bell, 1995:277; Barbieri, 2009:83; Denis *et al.*, 2019:56). However, as in the current dataset, the proportion of use for men remains lower than for women in many English dialects such as

[35] The un-normed data show percentages calculated out of the total quotatives (N = 430), while the normed data show percentages calculated out of each group's totals. For example, the number *of be like* produced by female speakers out of the total for their group (N = 287)

CanE. Denis *et al.* (2019:57), interpreted the evidence in their data for the increasing frequency in be like usage for men over the years as evidence that not only women but men participate in vernacular reorganisation. They explain that vernacular reorganisation is "a period from about age 5–17 during which children and adolescents shift their grammars in ways that reflect the diachronic trajectories of changes in progress" (*Ibid*.:44). They pose the question: "…if men undergo vernacular reorganization, how do we explain the male lag in the use of innovations […] and with many other changes in progress, a lag sometimes as long as a generation?" (Denis *et al.*, 2019:61). They suggest that this lag could be due to the social environment in which men and women "acquire different baseline variable grammars" because the speech that is presented to children by their parents is different for each gender (Denis *et al.*, 2019:62).

5.1.3 Quotative distribution across social class

The stratification of the speech sample according to social class and gender is illustrated in Table 5.7.

		SEC	UM	MM	LM
Gender	F		3	6	**2**
	M		3	2	**3**
Total			**6**	**8**	**5**

Table 5.7: Stratification according to social class and gender

Table 5.8 and Figure 5.6 provide the un-normed distribution of the quotatives across social class. These results suggest that the middle class use quotative *be like* significantly more than the other two groups. The quotatives *say* and *zero* show an even distribution across all three social classes. Even though *think* is relatively infrequent in the data, it might also suggest differences between the social classes in the normalised results.

SEC	be like	%	say	%	zero	%	think	%	other	%
UM	33	**8**	40	**9**	11	**3**	5	**1**	8	**2**
MM	119	**28**	33	**8**	16	**4**	6	**1**	48	**11**
LM	44	**10**	34	**8**	11	**3**	7	**2**	15	**3**
Total										**430**

Table 5.8: Un-normed quotative distribution across social class

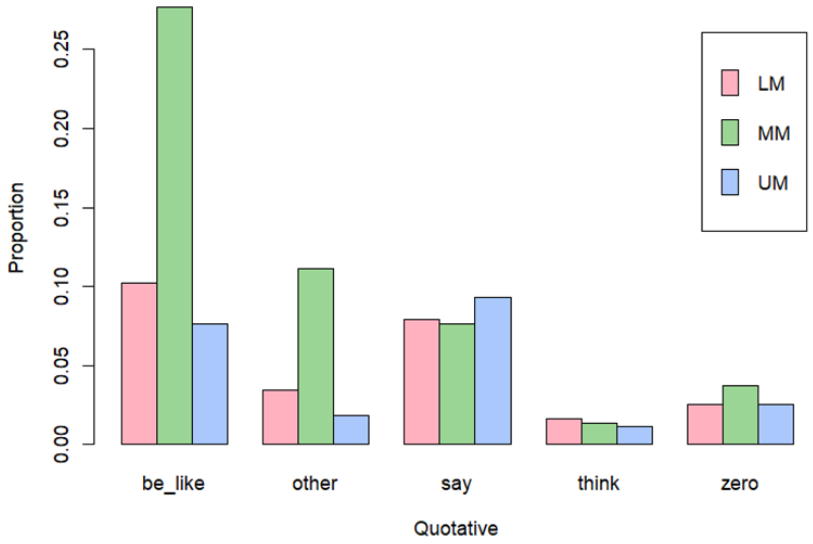


Figure 5.6: Un-normed quotative distribution across social class

The normalised frequencies across social class in Table 5.9 (see also Figure 5.7) thus provide a slightly different picture. Table 5.9 confirms that the middle class uses *be like* the most among the three social classes (54%), which is a similar result to Horn's (2022:80) study for young Cape Town speakers and also similar to studies of international English varieties such as Tannen (1986:324) Ferrara and Bell (1995:285) and Macauley (2001:11) (see discussion of previous quotative studies regarding social class distribution in Chapter 2, section 2.7.4.5). In the current data, even though the middle class is leading with the highest proportion, the differences between the three social classes are not as pronounced for *be like* as with gender. For instance, the lower middle class (40%) does not trail far behind and although the upper middle class uses *be like* the least, this group's proportion is not very low (34%). For quotative *say*, the results indicate that the upper middle class uses it the most (41%), followed by the lower middle class (31%). The middle class trails behind significantly (15%). This result suggests that the middle class might be replacing *say* with *be like.*

UM		
Quotative	%	N
say	**41**	40
be like	**34**	33
zero	**11**	11
think	**5**	5
other	**8**	8
Total		**97**

MM		
Quotative	%	N
be like	**54**	119
say	**15**	33
zero	**7**	16
think	**3**	6
other	**22**	48
Total		**222**

LM		
Quotative	%	N
be like	**40**	44
say	**31**	34
zero	**10**	11
think	**6**	7
other	**7**	15
Total		**111**

Table 5.9: Normed quotative distribution across social class

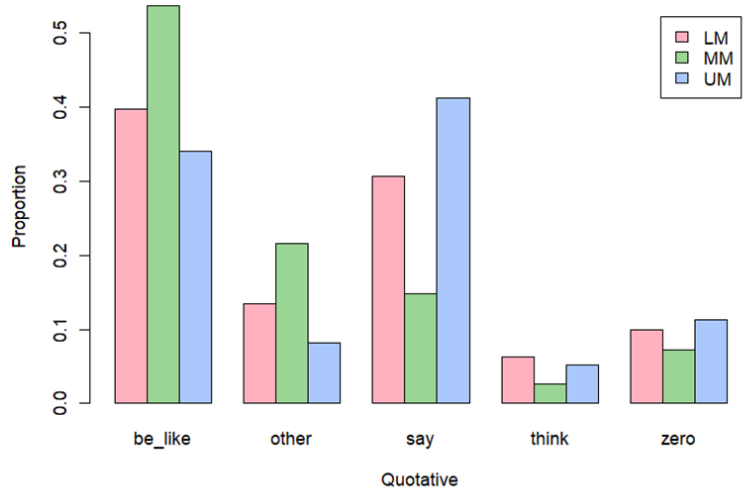


Figure 5.7: Normed quotative distribution across social class

The pattern for the upper middle class is similar to that of the older speakers and the men in the sense that they still prefer to use *say*, which could indicate that these three groups avoid socially stigmatised forms such as *be like* to some extent. The c-tree analysis (see section 5.1.6) will shed further light on the influence of social class on *be like*.

5.1.4 Quotative distribution across age and gender

Since only *be like*, *say* and *zero* are frequent enough to be divided further across social groups, only these three quotatives are considered for further analysis. Table 5.10 illustrates the stratification of the sample according to age and gender. Table 5.10 indicates that the sub-group with the most participants in the sample is the 18–20 female group (9) vs the male group (5).

	Age	
	18-20	33-48
Gender		
F	9	2
M	5	3
Total	**14**	**5**

Table 5.10: Stratification across age and gender

The results for the un-normed distribution of the three most frequent quotatives across age and gender are illustrated in Table 5.11 and Figure 5.8.

		Be like				Say				Zero				Other			
Gender		**F**	**%**	**M**	**%**	**F**	**%**	**M**	**%**	**F**	**%**	**M**	**%**	**F**	**%**	**M**	**%**
Age	18-20	153	**36**	25	**6**	41	**10**	26	**6**	12	**3**	18	**4**	49	**11**	17	**4**
	33-48	7	**2**	11	**3**	13	**3**	27	**6**	3	**1**	5	**1**	9	**2**	14	**3**
Total																	**430**

Table 5.11: Un-normed distribution of the quotatives across age and gender

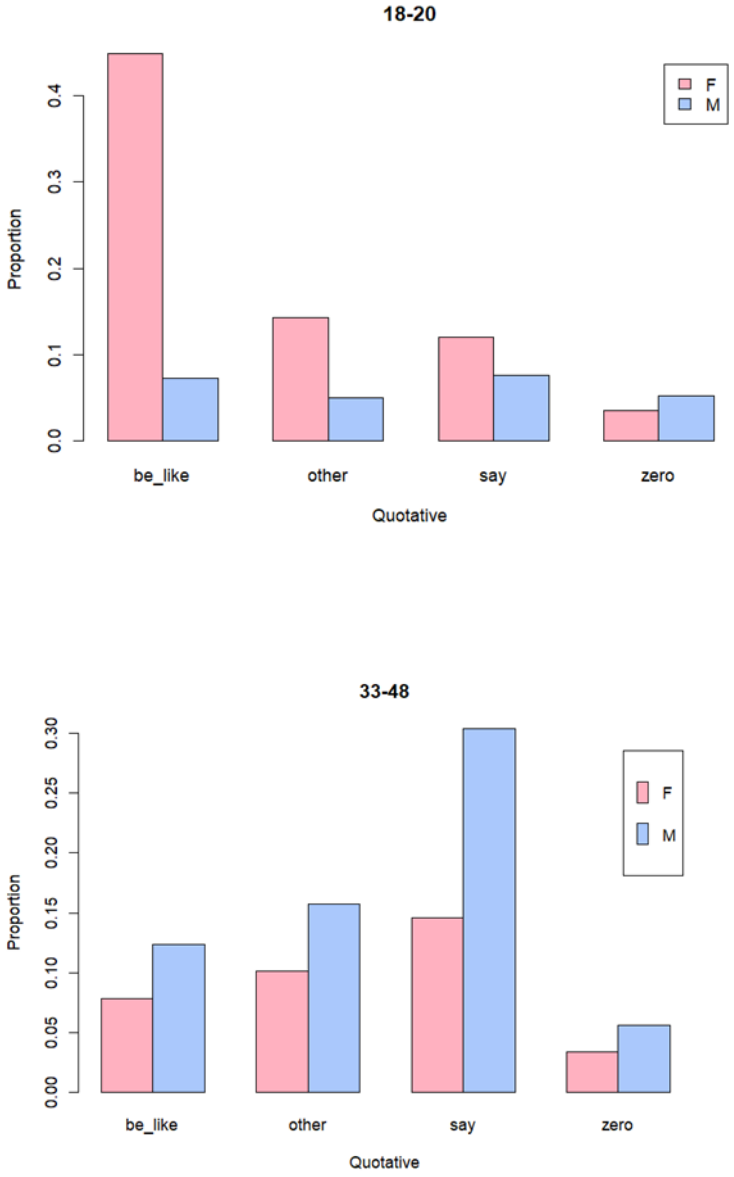


Figure 5.8: Un-normed distribution of the quotatives across age and gender

The un-normed results in Table 5.11 and Figure 5.8 suggest that the younger females produced the most instances of *be like*: 36% of the total quotatives. The sub-group that produced the second highest proportion of *be like* are the younger males: 6%. The young female group used *be like* more than *say*, while the young male group used an equal proportion of *say* and *be like*, indicating that the young male group is also contributing to the high proportions of *be like* use in WSAE.

Zero is only slightly more frequent in the younger group vs the older group. The older men obtained an equal proportion of *say* to the younger men. The normalised results will illuminate further differences between the social groups.

Table 5.12 and show the normalised results and confirm that the female speakers in the younger group are the ones with the highest proportion of *be like* usage. The results also suggest that there is not as great of a gender difference in the usage of *be like* among older speakers. This reflects Gardner *et al.*'s (2020:308) findings for York (EngE), for which they stated that the gender difference becomes more pronounced among younger speakers. Thus, older speakers do not show the same gender stratification of *be like* that younger speakers do.

18-20 Females

Quotative	%	N
be like	**60**	153
say	**16**	41
zero	**5**	12
other	**19**	49
Total		**255**

18-20 Males

Quotative	%	N
be like	**17**	25
say	**18**	26
zero	**13**	18
other	**12**	17
Total		**143**

33-48 Females

Quotative	%	N
be like	**22**	7
say	**41**	13
zero	**9**	3
other	**28**	9
Total		**32**

33-48 Males

Quotative	%	N
be like	**19**	11
say	**47**	27
zero	**9**	5
other	**25**	14
Total		**57**

Table 5.12: Normed distribution of the quotatives across age and gender

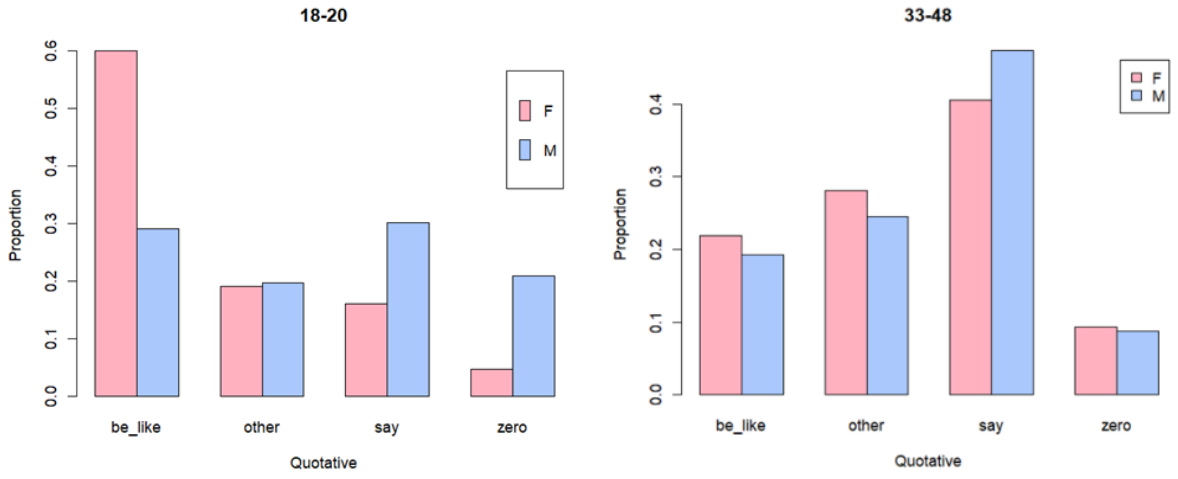


Figure 5.9: Normed distribution of the quotatives across age and gender

The results also indicate that the men in the younger group use the zero quotative the most out of all the other groups. Regarding *say*, the results show that the women and men in the older group use it relatively equally, and that the older men have a 6% lead in the use of *say*, while the women have a 3% lead in *be like.* The results are surprising for the younger male group as it indicates that they obtained a lower proportion of *be like* use compared to the older men. This result is partly explained by an older speaker Simon (33), who is grouped with the younger speakers in his use of *be like* in the c-tree results (see the c-tree illustrated in Figure 5.10, section 5.1.6) and is thus an outlier, since he patterns more towards the younger than older speakers.

The next section considers the distribution of the main quotatives across gender and social class, first combining the two age groups together and then considering the 18–20 group separately. Since this analysis considers the data when split into gender and age groups within three social classes, only the normed results are presented for further analysis.

5.1.5 Quotative distribution across gender and social class in the younger and older age groups

Table 5.13 illustrate the normed results for the distribution of the main quotatives across gender and social class for the entire dataset, however, percentages were calculated from the totals of each gender in each social class. Because there are multiple variables involved, it was not possible to present these results in a proportional bar graph. The female social groups presented in Table 5.13 comprise one older female participant in the middle-class group and one in the upper middle-class group. Therefore, there are only young participants in the lower-class female group. The upper middle-class male group consists of three younger speakers, which means this group also has no older representatives.

	be like	%	say	%	zero	%	N
UM							
Female	21	**38**	24	**43**	1	**2**	56
Male	12	**29**	16	**39**	10	**24**	41
MM							
Female	114	**61**	17	**9**	13	**7**	187
Male	5	**14**	16	**46**	3	**9**	35
LM							
Female	25	**57**	13	**30**	1	**2**	44
Male	19	**28**	21	**31**	10	**15**	67
Total							**430**

Table 5.13: Normed distribution of the quotatives across gender and social class

Even though the upper middle-class female and male participants were equal in number (3 for each), their percentages were obtained from the separate total quotatives for each gender in that social class. The other two social classes were analysed in the same way.

The results for both age groups combined indicate that females have the highest rate of *be like* usage across all three social groups and that the difference between the upper middle-class female and male groups for *be like* are not as pronounced as between the other two social groups. It also seems to indicate that the middle-class male group trails the furthest behind (14%) compared to the other male social groups (UM:29%, LM:28%).

The women in the upper middle class have a slightly higher frequency for *say* than the men and in the middle class, the men use *say* much more than the women (46% vs 9%). The fact that the UM female subgroup consists of one speaker from the older group and two from the younger, while the middle-class male group consists of one older and one younger speaker, seems to indicate that age is interacting with the result regarding *say* as well as regarding *be like* in the middle-class male group. The men in the upper middle and lower middle classes use the zero quotative more than the women, while it is more evenly distributed in the middle class, however, even here the men slightly lead with 9% vs 7%. Lastly, it is necessary to analyse the distribution of the quotatives according to social class and gender within the younger group specifically to determine if there were any age differences obscured in the combined results. See Table 5.14 below.

	be like	%	say	%	zero	%	N
UM							
Female	18	**45**	12	**30**	1	**3**	40
Male	12	**29**	16	**39**	10	**24**	41
MM							
Female	110	**64**	16	**9**	10	**6**	171
Male	5	**20**	7	**28**	3	**12**	25
LM							
Female	25	**57**	13	**30**	1	**2**	44
Male	8	**40**	3	**15**	5	**25**	20
Total							**341**

Table 5.14: Normed distribution of the quotatives in the 18–20 age group across social class

The male middle-class and lower middle-class groups produced the least amount of quotatives because each of these two groups only consists of one speaker. The decision was made to represent all three social classes in the younger group and not to combine the middle and lower middle-class groups together to have more male speakers in each group because the results for

social class in this chapter has so far revealed that there are differences between the lower-middle and middle-class groups that would be obscured if they were combined. It should be stressed therefore that the results for the groups with only one participant are thus merely impressionistic but might suggest similar patterns compared with community trends in other dialects.

With the older speakers removed, it can be seen that the young men in the upper middle-class group actually use *say* more than the young women in this social class. The younger upper middle-class women have a 16% lead over the younger men in the use of *be like*. The UM men have a 9% lead in the use of *say* and a lead of 11% in the use of the zero quotative. The men also prefer the zero quotative much more than the women in the other two social classes, indicating that gender influences the use of the zero quotative in favour of men. In addition, even with the oldest speaker in the male middle-class group removed, the result still indicates that the middle-class male speaker uses *be like* the least compared to the male speaker in the LM group and the three male speakers in the UM group. Even though this is based on one speaker, this could imply that middle-class male speakers use *be like* the least i.e. compared to male speakers in the UM and LM classes and of course compared to females. However, a bigger sample would be necessary to confirm this. Since the studies for social class distribution regarding *be like* are scarce and most found no significant result in this respect (see section 2.6.6.5), it is not possible to compare the current results with trends regarding middle-class male speakers in other English dialects. However, the current dataset suggests that there could be social class differences in the use of *be like* in WSAE across men and women. This topic remains a gap in the literature that would need to be explored further.

5.1.6 C-tree analysis of the social distribution of *be like*, *say* and *zero*

Figure 5.10 presents the c-tree results for the three main quotatives. As confirmed in Figure 5.10, social class, age and gender are significant factors influencing quotative choice. Only *be like*, *say* and *zero* were included directly in the analysis, with all other quotatives placed in the "other" category. Firstly, gender is the most significant factor, with a value of $p<0.001$. Within the female group, social class is the next significant factor. The c-tree confirms the proportional results in Table 5.14 and reveals that the difference between young middle-class female speakers and the other subgroups is significant: They are the leaders in the use of *be like* vs the other quotatives (node 7 vs nodes 4, 5 & 8). This result is partly due to the fact that the younger upper- and lower-class females (node 4) use *say* more and *be like* less than the young MM females (node 7), who use *be like* more and *say* the least (see the discussion for Table 5.14). Since there is no older female speaker in the LM group and only one in the UM group, node 5 shows the pattern for the one female UM speaker and node 8 shows the pattern for the one middle-class older female speaker and can thus be ignored.

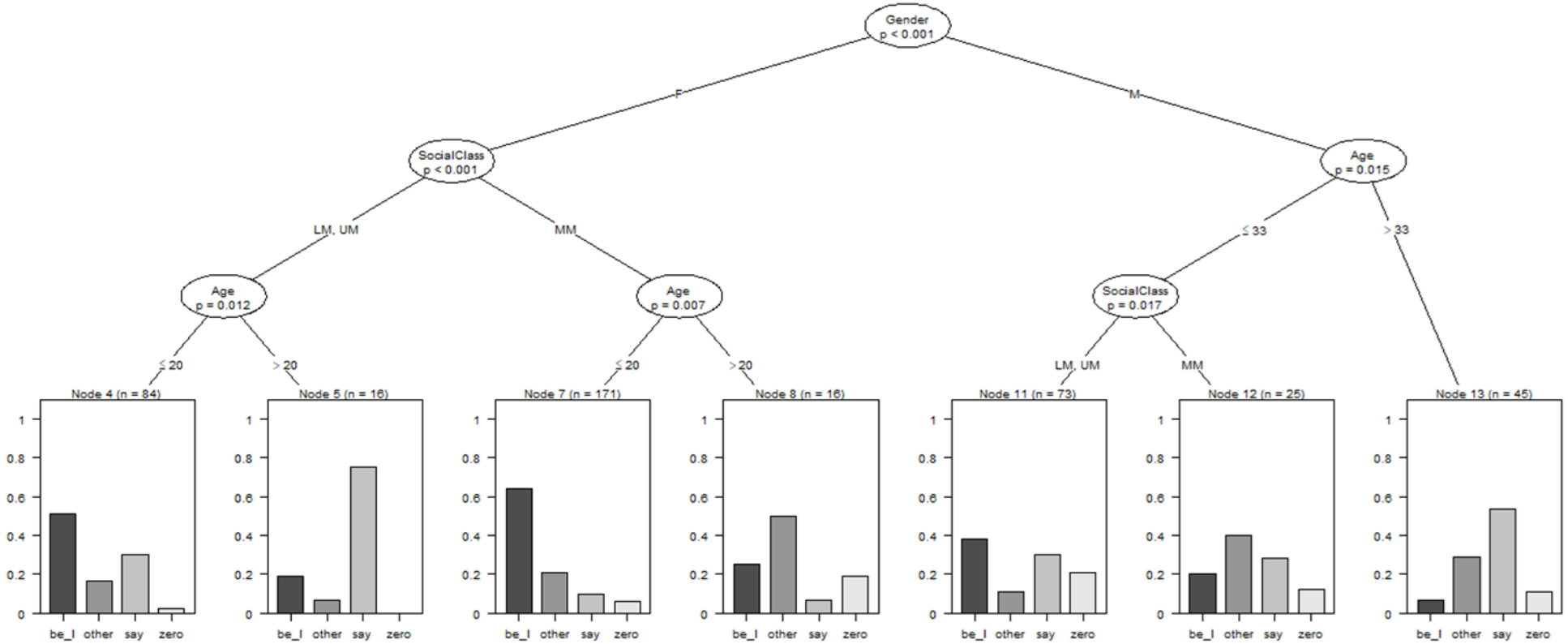


Figure 5.10: C-tree of the social distribution of be like, say *and* zero

Within the male group, age is significant at p = 0.015 and it is more important than social class, which is the opposite for the women: for them, social class is a more important factor. The age factor indicates that speakers 33 and younger have a different pattern regarding the three main quotatives than the speakers who are older than 33. With regards to the 33 and younger male group, social class is significant at p = 0.017. The results for social class should be regarded cautiously, because here as well, the middle class, 33 and under group is represented by only one speaker (Jeremy; node 12). The other male middle-class speaker is Ben (48). When nodes 11 & 12 are considered together, the results indicate that the younger men use *be like* more than the older men (node 13), but this is only the case when the outlier in the older group (Simon, 33) is relegated to the younger group as in this c-tree (see Table 5.12 for the distributional results which illustrated the opposite when Simon was grouped with the older speakers). The c-tree also confirms that the older female speakers (node 8) and the older male speakers (node 13) use *say* the most, as expected.

The c-tree analysis for *be like* in comparison to all other quotatives in the entire dataset is illustrated in Figure 5.11. As per Figure 5.11, the c-tree analysis for *be like* vs all other quotatives yields a similar result as the previous analysis, although social class is no longer a significant factor. In addition, where the men are grouped in the age range of 33 and younger, another significant age difference is observed in which node 8 indicates that Simon has a higher normed frequency of *be like* than the men in the 20 and younger age group, confirming that he is an outlier in the older group.

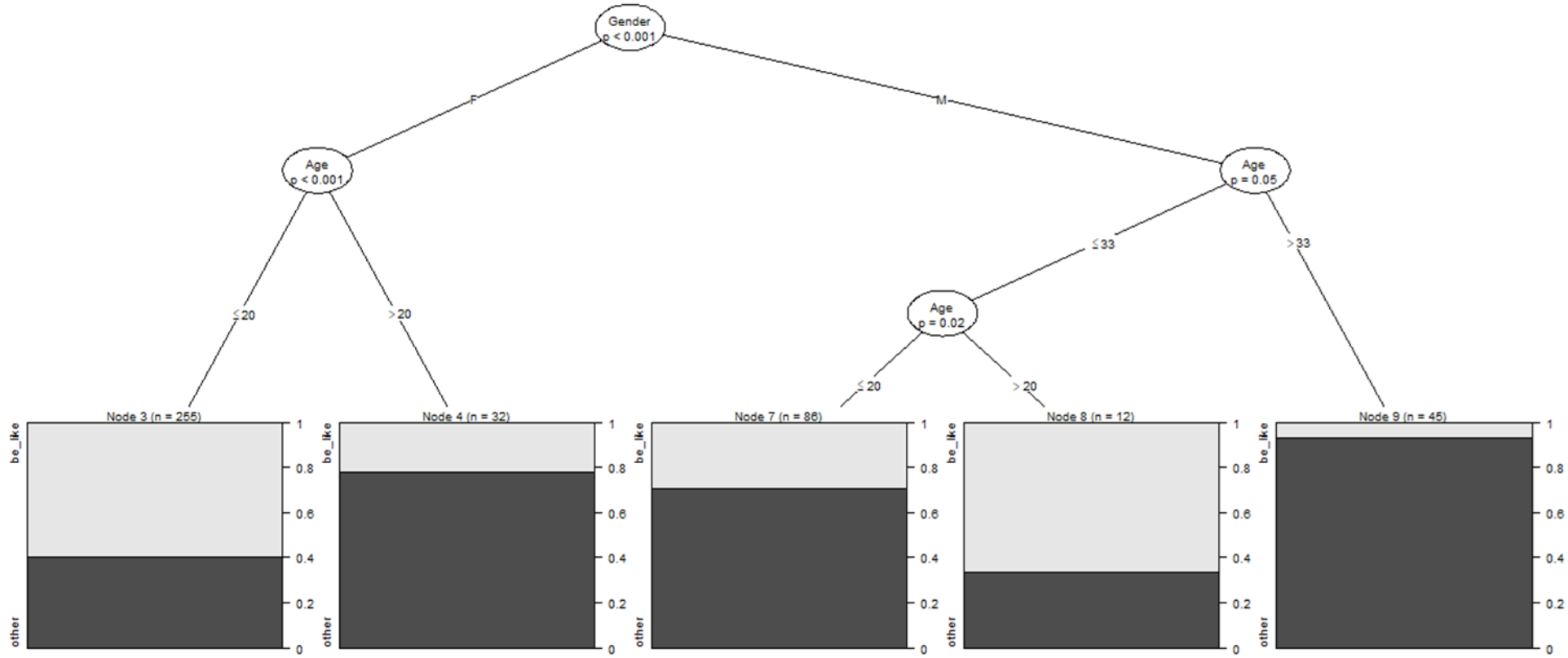


Figure 5.11: C-tree of the social distribution of be like*;* be like *is in grey.*

5.1.7 Distributional analysis of the content of the quote factor

This section provides an analysis of the quote content introduced by the quotatives in terms of thought vs speech reporting. Because this section focuses on only these two levels of this factor, anything that does not fit into the description of either thought or speech was placed in the n/a category. The n/a category is thus not included in the tables and graphs (except for Table 5.15) but is included in the total quotative count. An example of *be like* introducing thought is the following:

(124) So one day, like one of the music shops nearby [..]. And I **was like**, "I'm gonna go buy myself a ukulele." (<$N><s n=159>)

And an example of *be like* introducing speech:

(125) Then he walked in to meet [name] there, the cousin. And then **he's like**, "I'm going to buy everyone a tequila." (<$H><#><s n="148">)

In the entire data set, the quotative system is more prone to introduce direct speech (70%) rather than thought (26%). Below is a pie chart presenting this data.

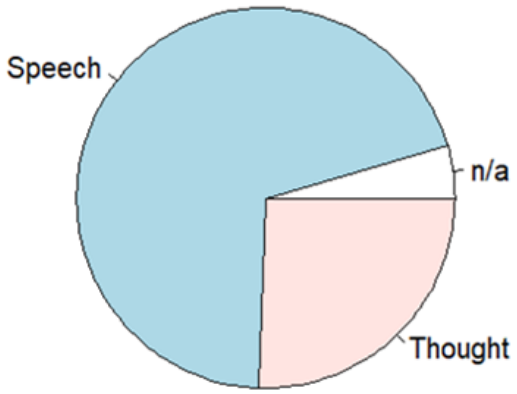


Figure 5.12: Thought vs Speech reporting in the dataset

Table 5.15 presents the results of the content of the quote factor; this is also illustrated in Figure 5.13.

Quotative	**thought** N	%	**speech** N	%	**n/a** N	%	
be like	60	**31**	121	**62**	15	**8**	
say	1	**1**	105	**98**	1	**1**	
zero	16	**42**	20	**53**	2	**5**	
ask	0	**0**	8	**100**	0	0	
be	2	**29**	5	**71**	0	0	
decide	4	**57**	3	**43**	0	0	
get like	1	**100**	0	**0**	0	0	
go	3	**17**	15	**83**	0	0	
tell	2	**12**	15	**88**	0	0	
think	18	**100**	0	**0**	0	0	
other	3	**23**	9	**69**	1	**8**	
Total	**110**	**26**	**301**	**70**	**19**	**4**	**430**

Table 5.15: Quotatives correlated to content of the quote factor

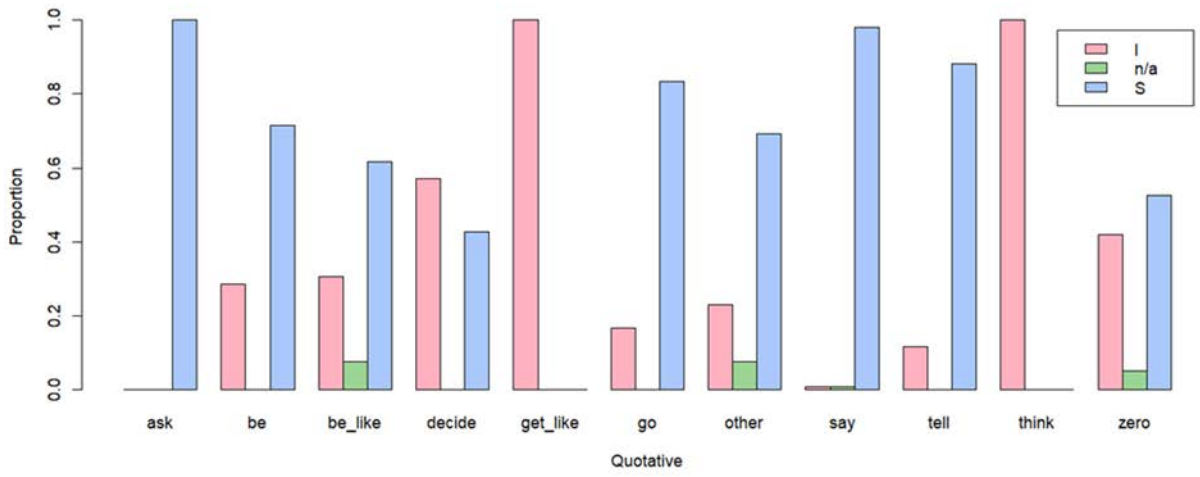


Figure 5.13: Quotative distribution for content of the quote

The results reflect what has been stated for *say*, *think*, and *go* (Blyth *et al.*, 1990:215, see discussion in section Chapter 2, section 2.7.4). Table 5.15 indicates that *say* mainly introduces direct speech (98%), *think* only introduces thought in the data (100%) and *go* mainly introduces speech (83%). In the only instance in which *say* introduces thought, the speaker makes it clear that it is thought/internal dialogue rather than speech that is being introduced:

(126) …after about a month, I **said** to myself "No this is too difficult". (<$S><s n=54>)

In Tagliamonte and Hudson's data (1999:163), in both EngE (1995) and CanE (1996), the zero quotative was found to introduce non-lexicalised sound the most, followed by direct speech and lastly internal dialogue/thought. At the time, the zero quotative was more frequently used in CanE (20%) than in EngE (10%). In Rickford's (2007:6) study comparing the speech of AmE high school and college students from California in two corpora (1990,1994; 2005), he found that the zero quotative declined from 15.9% in 1990,1994 to 10.7% in 2005. In other dialects such as EngE, CanE and AusE, the zero quotative seems to have risen minimally over the years: Rodríguez Luoro *et al.* (2020:336, AusE, recorded 2011–2015) found the zero quotative at 17% frequency for Perth and Buchstaller (2015:463) found that it only slightly rose in data for EngE, Tyneside from 1994 (15%) to 2011 (18%). D'Arcy (2012:364) found that although the zero quotative is frequently used with gestures and sounds, it has also expanded to introduce all content types. In Table 5.15, zero has diverse functions, introducing thought (42%) and speech (53%) at close frequencies to each other, but it tends to occur with speech 11% more. With regard to *be like*, in Horn's (2022:76) analysis of the Cape Town speech community, the young WSAE speakers also preferred reporting speech predominantly rather than thought with *be like*. In addition, in her data

the zero quotative is mostly used with speech reporting vs thought as well, whereas my findings indicate that zero is more evenly distributed between the two. This could possibly indicate regional differences in WSAE regarding how the zero quotative is used.

Where in earlier studies *be like* was mainly found to report internal monologue/thought, in the current data it is mainly found with direct speech (62%) compared to thought (31%). According to Tagliamonte & D'Arcy (2004:507), high frequencies of *be like* with direct speech indicates a later stage of its diffusion in a speech community in which a "bleaching of its pragmatic function" (of introducing internal thought) is observed, and thus indicates increasing grammaticalisation (see the discussion regarding this in section 2.6.6.1).

Figure 5.13 illustrates why *be like* has become so frequent in the speech of young people: it entered the territory of *say*, the second most frequent quotative in the data and has expanded into direct-speech reporting to such an extent that it is competing with *say* for this function, whereas its earlier function reported in the literature used to be limited mainly to thought reporting.

5.1.8 Grammatical person analysis

This section investigates the correlation of the quotatives with grammatical person. Quotative *be like* has previously been investigated for the effect that grammatical person has on its use and linguists have found that at its early stage it has been mainly used to report the speech of the speaker (first person), as in:

(127) I turned to him. **I was like**, "No, I'm still waiting". (<$H><s n=149>)

and was rarely used to report the speech of others (third person), as in:

(128) My other friend **was** also **like**, "Okay, no, that was too much for me". (<$N><s n=97>)

However, it has been observed to have expanded to third person with much higher frequencies in more recent studies (see Chapter 2, section 2.7.4.2). Figure 5.14 illustrates the overall distribution of grammatical person among all the quotatives. As shown in Figure 5.14, the preferred grammatical person among all the quotatives is first-person reporting (49%), where speakers report on their own thoughts or speech, closely followed by third-person reporting (41%), where speakers report on the speech and (sometimes) thoughts of others.

Second-person reporting, as in (129) was infrequent in the data. However, it was not as rare (N = 14) as reported in previous studies (e.g. Gardner *et al*., 2020:297).

(129) You just sit there and you GO, "How do people think like this?" (<$M><#><sn="77">)

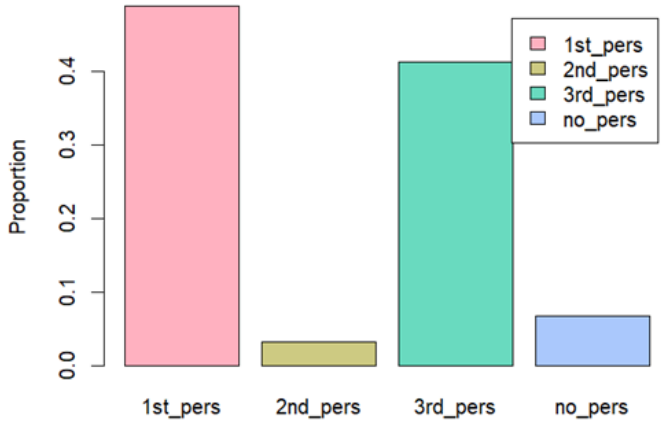


Figure 5.14: Distribution of grammatical person in the quotative system

There were also contexts in the data where no grammatical person was specified, as in:

(130) And there was a bit of a argument about “What, what's the problem if she moans at them?" (<$L><#><s n="46">)

Table 5.16 and Figure 5.15 present the results for grammatical person. The results indicate that *be like* appears much more frequently with first-person subjects (57%) rather than third person subjects (35%). *Say*, on the other hand, is found more with third-person than first-person subjects in the data (64% vs 35%). This result is usual for *say*, as it has been reported in the literature to be found mostly with third-person reporting (Blyth *et al.*, 1990:221). The zero quotative is almost equally as frequent between third and first person but indicates a slight preference for first person (37%). It is also found with second-person subjects and where there is no subject specified, indicating its versatility in the context of grammatical person (similar to the content of the quote). *Think* is found mostly with first-person reporting, which reveals that speakers usually tend to report their own thoughts rather than those of others. In this data for WSAE, it is apparent that *be like* has expanded to third-person subjects. However, its first-person bias is maintained, reflecting the constraint on grammatical person associated with *be like* in CanE and EngE (Gardner *et al.*, 2020:296).

Quotative	1st person	N	%	3rd person	N	%	2nd person	N	%	no person	N	%	
be like		111	**57**		68	**35**		2	**1**		15	**8**	
say		37	**35**		68	**64**		2	**2**		0	**0**	
zero		14	**37**		13	**34**		3	**8**		8	**21**	
ask		3	**38**		4	**50**		1	**13**		0	**0**	
be		1	**14**		3	**43**		0	**0**		3	**43**	
decide		6	**86**		1	**14**		0	**0**		0	**0**	
get like		1	**100**		0	**0**		0	**0**		0	**0**	
go		10	**56**		3	**17**		5	**28**		0	**0**	
tell		8	**47**		8	**47**		0	**0**		1	**6**	
think		15	**83**		2	**11**		1	**6**		0	**0**	
other		4	**31**		7	**54**		0	**0**		2	**15**	
Total		**210**	**49**		**177**	**41**		**14**	**3**		**29**	**7**	**430**

Table 5.16: Quotative distribution across grammatical person

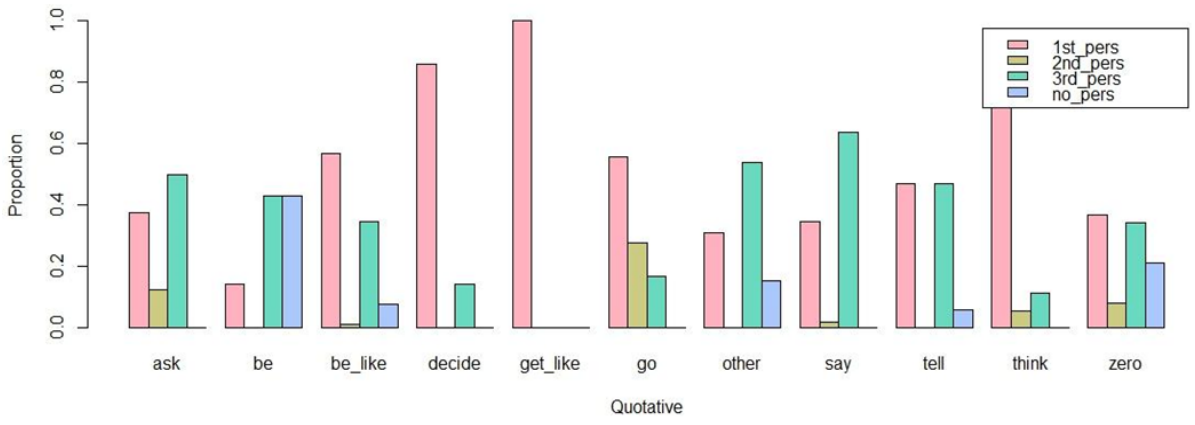


Figure 5.15: Percentage of quotatives across grammatical person

Horn's (2022:100) study of the quotatives in the speech of young people in Cape Town, found that among the young WSAE speakers, *be like* was distributed almost equally between first-person and third-person subjects, with a slight preference for first-person subjects. This seems to indicate that regional differences exist with regards to the choice of grammatical person with *be like* in WSAE.

5.1.9 Tense distribution of the quotatives

This section analyses the tense distribution of the quotatives. There are three environments that have been discussed in the literature in terms of quotative use (see section 2.7.4.3):

The present temporal reference with present tense morphology of the quotative verb:

(131) Yeah, my dad, he's, he **says** he's old. He's like an old man. **I'm like**, "No, dad, you're, you're thriving still." (<$M><s n=61>)

Past temporal reference with past tense morphology:

(132) but my brother and them, they still **ate** them, but I **was like**, "No, no, that does not look". (<$C><s n=40>)

And past temporal reference with present tense morphology, defined as the historical present:

(133) the Saturday night, we **had** the same conversation […] **he's like**, "But why don't we get married?". (<$L><s n=76>)

Exceptions to this categorisation scheme are modals and all other tense constructions that either do not occur with *be like*, such as the progressive and perfect aspects, or constructions with low numbers such as modals that occur with the infinitive form of the verb:

(134) My dad would only get upset if he, if I like told a lie and he found out, and **he'd be like** "You don't have to lie to me…" (<$N><s n=64>)

And future constructions, such as:

(135) If I try to introduce myself as Eve and someone knows me as [name], they're **gonna be like**, "Huh? What?" (<$C><s n=85>)

This analysis follows Singler's (2002:271) method of combining constructions such as the present progressive and perfect aspects (where they occur with other quotatives), modals with present morphology (e.g. *will*) and future constructions with the present tense and grouping constructions such as past progressive and perfect aspects (where they occur with other quotatives) and past forms of modals (*would*) with the past tense. First, an overall view of tense distribution in the data is provided.

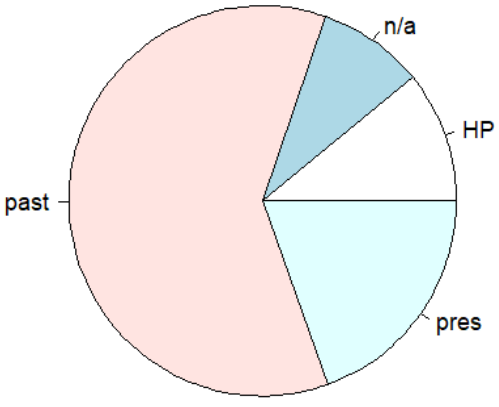


Figure 5.16: Tense distribution in the WSAE dataset

Figure 5.16 illustrates that the majority of quotatives are used in the past tense (61%), while the second most frequent context is the present tense (20%) and the historical present occurs the least (11%). and Figure 5o provide the results for tense distribution across the quotatives. The

n/a category represents the zero quotatives, which are excluded from the tense analysis since they do not have a quotative verb that indicates tense.

Quotative	present	N	%	past	N	%	HP	N	%	n/a	N	%	
be like		39	**20**		131	**67**		26	**13**				
say		13	**12**		85	**79**		9	**8**				
zero		0	**0**		0	**0**		0	**0**		38	**100**	
ask		1	**13**		6	**75**		1	**13**				
be		2	**29**		3	**43**		2	**29**				
decide		0	**0**		6	**86**		1	**14**				
get like		1	**100**		0	**0**		0	**0**				
go		14	**78**		2	**11**		2	**11**				
tell		7	**41**		8	**47**		2	**12**				
think		2	**11**		12	**67**		4	**22**				
other		5	**38**		8	**62**		0	**0**				
Total		**84**	**20**		**261**	**61**		**47**	**11**		**38**	**9**	**430**

Table 5.17: Tense distribution across quotatives

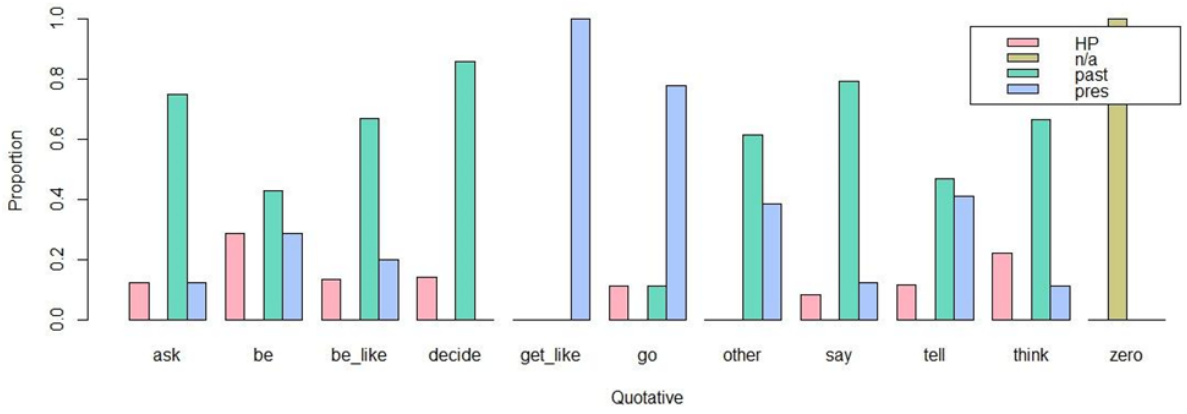


Figure 5.17: Tense distribution across the quotatives

The results in Table 5.17 indicate that *be like* is overwhelmingly preferred with the past tense (67%), while it is used second most frequently with the present tense (20%) and lastly with the historical present (13%). This result is surprising since it has been reported to be used most with the historical present tense (Tagliamonte *et al.*, 2016:835) even though it was stated that the past tense context has steadily increased with speakers born in the 1970s and later (Gardner *et al.*, 2020:299). For young Capetonian speakers of WSAE, Horn (2022:93) reported that the HP was used more with *be like* than the past tense. The different result for the current data compared to Horn's study and the studies mentioned for CanE in which the HP is preferred, means that either *be like* has become more prominent with the past tense among younger speakers as time has passed (and as a result the HP has decreased significantly in this context), or that the Port

Elizabeth speakers have a different narrative style in which the past tense is preferred over the HP. In this regard, *say* also prefers the past tense in my data, with the present tense second and HP third, whereas in Horn's (2022:93) data, the HP is the second most frequent quotative context for *say* and the present tense is third. In D'Arcy's (2012:358) longitudinal study of NZE across three corpora, she found that the HP context for *say* declined in the Canterbury Corpus (1994–2006) from 47% with speakers recorded in 1994–1997, 37% with speakers recorded between 1998–2001 and 25% with speakers recorded between 2002–2006. The corpus also indicates that the HP context in NZE for *go* declined as well while, in contrast, the HP rose from about 20% with *be like* in 1994–1997 to over 60% in 2002–2006 (no specific percentages are given) and D'Arcy (2012:363) states that the HP has come to be mostly found with *be like* in NZE as opposed to the other quotatives. This again is in contrast to the current results which indicate that *be like* is rare with the HP and much more frequent with the past tense. Thus, the current data for WSAE regarding tense correlation with *be like* reflects the trend of dialects such as BrE, (specifically Scottish English; Macauley, 2001:10), in favour of past tense usage rather than the HP, found for dialects such as AusE (Winter, 2002:11), AmE (Singler, 2001:271), CanE and NZE (Tagliamonte, 2016:835, 836); and EngE (York; Gardner *et al*., 2020:218).

5.1.10 C-tree analysis of the impact of internal factors on quotative usage

Figure 5.18 provides the c-tree analysis of the quotatives in relation to the internal factors discussed above. Among the three factors, the highest split is tense, though this is due to the zero quotative being excluded since it could not be coded for tense. The next significant factor is quote content, split between thought and speech. Within the thought predictor, *be like* is the most frequent out of the other quotatives, while *think* is the second most frequent (node 2). As indicated by the distributional results in Table 5.15, *say* and *tell* are rarely used to report thought.

Within direct-speech reporting, tense is a significant factor and indicates that the major difference is between HP and the past tense on one side and present tense on the other side. The c-tree illustrates that *be like* and *say* are found most often with the past and HP tenses combined than with the present tense (see Table 5.17). The exception is *tell*, which is found more often with the present tense. As illustrated in figure 15, the difference between first and third person is not large enough among the quotatives to be significant.

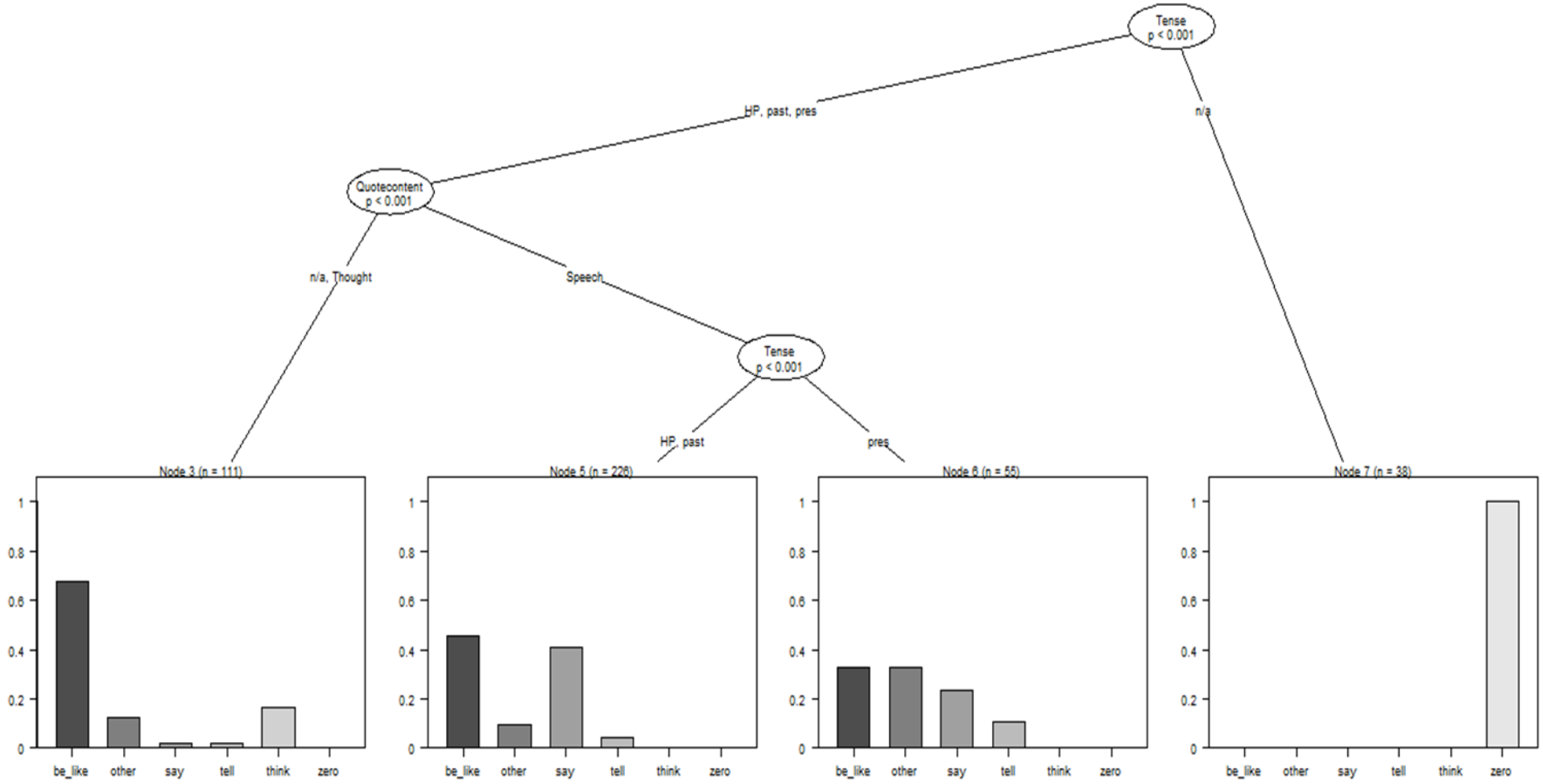


Figure 5.18: Impact of internal factors on quotative usage

5.1.11 Internal (linguistic) factors influencing *be like* in the context of social factors

This section aims to discover any differences regarding the linguistic factors influencing the use of *be like* across the younger and older age groups, as well as across gender and social class; in order to find out to what extent it has grammaticalized in each group. The other quotatives will be considered for comparison. Table provides the results for content of the quote for the male and female genders in the 18–20 age group.

Firstly, the results for content of the quote indicate that both genders in the younger group report speech more than thought overall – females (68%) and males (71%). Regarding the specific quotatives, the young males and females pattern mostly in a similar way: *be like* is used more frequently for reporting speech rather than thought by both genders. *Say* is predominantly used for speech reporting with both genders and *think* is categorically used by both genders for thought/internal monologue reporting (100% each), as expected. *Go* only appears with females and is used more for direct speech than thought. *Be* on its own only appears twice in the younger group and is used by men to report speech in both instances. *Tell* is only used by the women in the data and is used mostly for direct speech rather than thought. One difference is apparent between the two genders in both age groups (compare Table 5.18 with Table 5.19): the zero quotative is used differently. For both age groups combined, the women use the zero quotative more to report thought (60% / N = 9) rather than speech (27% / N = 4), while the men use it more to report speech (69% / N = 16) rather than thought (30% / N = 7). However, the numbers in both the young and older age groups for zero are relatively low when distributed across gender and a larger sample would tell if this difference holds more generally.

female speakers

Quotative	thought	N	%	speech	N	%	n/a	N	%	
be like		46	30		98	64		9	6	
say		0	0		40	98		1	2	
go		3	17		15	83				
zero		7	58		3	25		2	17	
tell		2	17		10	83				
think		8	100		0	0				
ask		0	0		3	100				
be		0	0		0	0				
get like		1	100		0	0				
decide		1	100		0	0				
other		1	17		5	83				
Total		**69**	**27**		**174**	**68**		**12**	**5**	**255**

male speakers

Quotative	thought	N	%	speech	N	%	n/a	N	%	
be like		7	28		14	56		4	16	
say		0	0		26	100				
go		0	0		0	0				
zero		7	39		11	61				
tell		0	0		0	0				
think		3	100		0	0				
ask		0	0		3	100				
be		0	0		2	100				
get like		0	0		0	0				
decide		3	50		3	50				
other		1	33		2	67				
Total		**21**	**24**		**61**	**71**		**4**	**5**	**86**

Table 5.18: Distribution of content of the quote in the 18–20 age group, stratified by gender.

female speakers

Quotative	thought	N	%	speech	N	%	n/a	N	%	
be like		1	14		6	86				
say		0	0		13	100				
go		0	0		0	0				
zero		2	67		1	33				
tell		0	0		2	100				
think		2	100		0	0				
ask		0	0		1	100				
be		1	50		1	50				
get like		0	0		0	0				
decide		0	0		0	0				
other		1	50		1	50				
Total		**7**	**22**		**25**	**78**		**0**	0	**32**

male speakers

Quotative	thought	N	%	speech	N	%	n/a	N	%	
be like		6	55		3	27		2	18	
say		1	4		26	96				
go		0	0		0	0				
zero		0	0		5	100				
tell		0	0		3	100				
think		5	100		0	0				
ask		0	0		1	100				
be		1	33		2	67				
get like		0	0		0	0				
decide		0	0		0	0				
other		0	0		1	50		1	50	
Total		**13**	**23**		**41**	**72**		**3**	**5**	**57**

Table 5.19: Distribution of content of the quote across quotatives in the 33-48 age group, stratified by gender.

The older speakers also report more direct speech than thought overall, indicating that this pattern is true for the entire speech community. *Be like* has a low frequency among the older speakers, but with respect to the few instances in which it occurs, the older women (2 participants) use it more to report speech than thought, similarly to the younger group. However, older men use it more to report thought than speech. This unexpected direction for the older men cannot be verified because of the low numbers of *be like* quotatives in this regard, but if it held true in a bigger sample, it would indicate the earlier pattern for *be like* in the speech of older men i.e. where it at first was mainly found to report internal thought than speech. This type of result would be in line with two established patterns in the theory of language change in progress. The first is that the older generation reflects earlier patterns of a language: the apparent-time hypothesis (Chambers, 2002:212) and that men in general lag behind in adopting linguistic innovations, whereas women usually are the leaders (Labov, 1972:301). This would explain why the older women pattern similarly to the younger generation and why the older men do not – the older men possibly lag behind the older women in the internal development of *be like,* thus reflecting its earlier pattern where the content of the quote is concerned.

Table Table 5.20 presents the results for speech vs thought reporting across social class. The lower, middle and upper middle classes all pattern in the same way in that *be like* is used more for reporting speech than thought. The same applies to *say*. Table 5.21 presents the results for speech vs thought reporting across gender and social class. When the social classes were split into the two genders, *be like* patterned in the same way for the women in all three classes – speech was preferred over thought.

			UM						
Quotative	**thought**	N	%	**speech**	N	%	**n/a**	N	%	
be like		11	**33**		19	**58**		3	**9**	
say		0	**0**		39	**98**		1	**3**	
zero		2	**18**		8	**73**		1	**9**	
other		7	**54**		6	**46**		0	**0**	
Total		**20**	**21**		**72**	**74**		**5**	**5**	**97**

			MM						
Quotative	**thought**	N	%	**speech**	N	%	**n/a**	N	%	
be like		36	**30**		75	**63**		8	**7**	
say		1	**3**		32	**97**		0	**0**	
zero		9	**56**		6	**38**		1	**6**	
other		15	**28**		39	**72**		0	**0**	
Total		**61**	**27**		**152**	**68**		**9**	**4**	**222**

			LM						
Quotative	**thought**	N	%	**speech**	N	%	**n/a**	N	%	
be like		13	**33**		27	**68**		4	**10**	
say		0	**0**		34	**100**		0	**0**	
zero		5	**45**		6	**55**		0	**0**	
other		11	**50**		10	**45**		1	**5**	
Total		**29**	**26**		**77**	**69**		5	**5**	**111**

Table 5.20: Distribution of content of the quote across social class

Female speakers

UM

Quotative	thought	N	%	speech	N	%	n/a	N	%	
be like		8	**38**		13	**62**		0	**0**	
say		0	**0**		23	**96**		1	**4**	
zero		0	**0**		0	**0**		1	**100**	
other		5	**50**		5	**50**		0	**0**	
Total		**13**	**23**		**41**	**73**		**2**	**4**	**56**

MM

Quotative	thought	N	%	speech	N	%	n/a	N	%	
be like		35	**31**		71	**62**		8	**7**	
say		0	**0**		17	**100**		0	**0**	
zero		8	**62**		4	**31**		1	**8**	
other		12	**28**		31	**72**		0	**0**	
Total		**55**	**29**		**123**	**66**		**9**	**5**	**187**

LM

Quotative	thought	N	%	speech	N	%	n/a	N	%	
be like		4	**16**		20	**80**		1	**4**	
say		0	**0**		13	**100**		0	**0**	
zero		1	**100**		0	**0**		0	**0**	
other		3	**60**		2	**40**		0	**0**	
Total		**8**	**18**		**35**	**80**		**1**	**2**	**44**

Male speakers

UM

Quotative	thought	N	%	speech	N	%	n/a	N	%	
be like		3	**25**		6	**50**		3	**25**	
say		0	**0**		16	**100**		0	**0**	
zero		2	**20**		8	**80**		0	**0**	
other		2	**67**		1	**33**		0	**0**	
Total		**7**	**17**		**31**	**76**		**3**	**7**	**41**

MM

Quotative	thought	N	%	speech	N	%	n/a	N	%	
be like		1	**20**		4	**80**		0	0	
say		1	**6**		15	**94**		0	0	
zero		1	**33**		2	**67**		0	0	
other		3	**27**		8	**73**		0	0	
Total		**6**	**17**		**29**	**83**		**0**	**0**	**35**

LM

Quotative	thought	N	%	speech	N	%	n/a	N	%	
be like		9	47		7	37		3	**16**	
say		0	0		21	100		0	**0**	
zero		4	40		6	60		0	**0**	
other		8	47		8	47		1	**6**	
Total		**21**	**31**		**42**	**63**		**4**	**6**	**67**

Table 5.21: Quotative distribution for content of the quote across social class and gender.

From Table 5.21 we can also see that for the men, some differences are observed: both the middle and upper classes used *be like* more for speech rather than thought. However, the pattern is reversed in the lower middle class where it is used with thought 47% of the time vs 37% of the time: there is a 10% difference between the two quote types. This could be linked to the fact that the LM group reported thought to a higher degree than the other two social classes overall: the UM and MM classes obtained 17% overall for thought reporting while the LM group obtained 21%. However, these results could also suggest that there might be a social-class difference in the larger community; but more data would be needed to corroborate this. The social class factor will not be investigated separately for the younger and older age groups since certain subgroups have only one representative and the token numbers regarding internal factors are too low to be divided further.

Moving on to grammatical person, Table 5.22 presents the results for quotative distribution regarding grammatical person across gender in the 18–20 age group. The result for the younger

female speakers is the same as the overall results for grammatical person and what is reported as being associated with *be like* in general: a first-person bias in, for example, CanE and EngE (see Gardner *et al.*, 2020:296). The result for the younger men, on the other hand, is surprising. Their use of *be like* is equally distributed between first- and third-person contexts, indicating that it has spread more into third-person usage with younger men than younger women. The results for the younger women regarding *say* also reflects what is reported – it favours third person subjects above first person (Blyth *et al.*,1990:221). However, the men again do not follow this trend: They use *say* equally for first and third-person subjects. An increase into first-person reporting is not attested in D'Arcy's (2012:358) study of the internal factors constraining *say* in the Canterbury Corpus for NZE over the years (1997–2006 corpora): it favours third-person subjects every time. Since the younger men also use *say* equally between first and third-person contexts, this could suggest gender differences in reporting the speech/thought of others vs that of oneself rather than an advanced grammatical development regarding *be like*. The zero quotative seems to be more versatile with grammatical person as well and is distributed evenly between first- and third-person reporting for both genders.

female speakers

Quotative	1st person	N	%	3rd person	N	%	2nd person	N	%	no person	N	%	
be like		90	**59**		55	**36**		1	**1**		7	**5**	
say		9	**22**		32	**78**		0	**0**		0	**0**	
go		10	**56**		3	**17**		5	**28**		0	**0**	
zero		5	**42**		5	**42**		1	**8**		1	**8**	
other		21	**68**		8	**26**		0	**0**		2	**6**	
Total		**135**	**53**		**103**	**40**		**7**	**3**		**10**	**4**	**255**

male speakers

Quotative	1st person	N	%	3rd person	N	%	2nd person	N	%	no person	N	%	
be like		9	**36**		9	**36**		0	**0**		7	**28**	
say		12	**46**		12	**46**		2	**8**		0	**0**	
go		0	**0**		0	**0**		0	**0**		0	**0**	
zero		5	**28**		6	**33**		2	**11**		5	**28**	
other		9	**53**		6	**35**		0	**0**		2	**12**	
Total		**35**	**41**		**33**	**38**		**4**	**5**		**14**	**16**	**86**

Table 5.22: Quotative distribution for the grammatical person factor across gender in the 18–20 age group.

Table 5.23 below provides the results for the distribution of grammatical person in the 33–48 age group, stratified by gender. The older women and men indicate the same pattern of usage for *be like* as the younger group: first person is favoured above third-person reporting. Similarly to the younger women, the older men and women both prefer third-person reporting above first person for *say*. This means that it is only in the younger male group where *be like* and *say* are evenly

distributed between first- and third-person contexts. Whether this difference is meaningful is not clear at this stage and would need further investigation.

female speakers

Quotative	1st person	N	%	3rd person	N	%	2nd person	N	%	no person	N	%	
be like		5	71		2	29		0	0		0	0	
say		6	46		7	54		0	0		0	0	
zero		1	33		1	33		0	0		1	33	
other		4	44		4	44		0	0		1	11	
Total		**16**	**50**		**14**	**44**		**0**	0		**2**	**6**	**32**

male speakers

Quotative	1st person	N	%	3rd person	N	%	2nd person	N	%	no person	N	%	
be like		7	64		2	18		1	9		1	9	
say		10	37		17	63		0	0		0	0	
zero		3	60		1	20		0	0		1	20	
other		4	29		7	50		2	14		1	7	
Total		**24**	**42**		**27**	**47**		**3**	**5**		**3**	**5**	**57**

Table 5.23: Distribution of grammatical person in the 33–48 age group, stratified by gender.

As can be seen from Table 5.24, both the LM and MM classes have a higher rate of first-person vs third-person speech reporting overall, while the UM have the exact same rate for both. However, the LM is not far behind the UM in that the difference between first- and third-person reporting is small (the rate of first-person reporting for the LM is only 7% higher than for the third person). All three groups used *be like* and *say* in the dominant patterns described above: *be like* was mostly used in first-person contexts and *say* was mostly used in third-person contexts. This suggests that statistical analysis would not find social class a significant factor in constraining the use of *be like* where grammatical person is concerned. Another difference is that the UM class uses the zero quotative the most when there is no grammatical person involved. These are examples such as:

(136) I'm good at, you know, **like**, "Here's how you do this thing. Do it". (<$A><s n=89>)

The UM use the zero quotative the second most with the third person and the least with the first person. On the other hand, the MM use it the most with the first person, while it is equally distributed between first and third person in the LM group. This indicates that the zero quotative is not as constrained by grammatical person as *be like* and *say* is since it is not associated with a specific grammatical person.

UM

Quotative	1st person	N	%	3rd person	N	%	2nd person	N	%	no person	N	%	
be like		14	**42**		8	**24**		0	**0**		11	**33**	
say		13	**33**		25	**63**		2	**5**		0	**0**	
zero		2	**18**		4	**36**		0	**0**		5	**45**	
other		10	**77**		2	**15**		0	**0**		1	**8**	
Total		**39**	**40**		**39**	**40**		**2**	**2**		**17**	**18**	**97**

MM

Quotative	1st person	N	%	3rd person	N	%	2nd person	N	%	no person	N	%	
be like		70	**59**		46	**39**		1	**1**		2	**2**	
say		14	**42**		19	**58**		0	**0**		0	**0**	
zero		8	**50**		5	**31**		1	**6**		2	**13**	
other		32	**59**		14	**26**		5	**9**		3	**6**	
Total		**124**	**56**		**84**	**38**		**7**	**3**		**7**	**3**	**222**

LM

Quotative	1st person	N	%	3rd person	N	%	2nd person	N	%	no person	N	%	
be like		27	**61**		14	**32**		1	**2**		2	**5**	
say		10	**29**		24	**71**		0	**0**		0	**0**	
zero		4	**36**		4	**36**		2	**18**		1	**9**	
other		6	**27**		12	**55**		2	**9**		2	**9**	
Total		**47**	**42**		**54**	**49**		**5**	**5**		**5**	**5**	**111**

Table 5.24: Distribution of grammatical person across social class.

presents the distribution of the quotatives across gender and social class regarding the grammatical person factor.

Female speakers

UM

Quotative	1st person	N	%	3rd person	N	%	2nd person	N	%	no person	N	%	
be like		12	**57**		5	**24**		0	**0**		4	**19**	
say		7	**29**		17	**71**		0	**0**		0	**0**	
zero		0	**0**		1	**100**		0	**0**		0	**0**	
other		8	**80**		2	**20**		0	**0**		0	**0**	
Total		**27**	**48**		**25**	**45**		**0**	**0**		**4**	**7**	**56**

MM

Quotative	1st person	N	%	3rd person	N	%	2nd person	N	%	no person	N	%	
be like		67	**59**		44	**39**		1	**1**		2	**2**	
say		5	**29**		12	**71**		0	**0**		0	**0**	
zero		6	**46**		4	**31**		1	**8**		2	**15**	
other		26	**60**		9	**21**		5	**12**		3	**7**	
Total		**104**	**56**		**69**	**37**		**7**	**4**		**7**	**4**	**187**

LM

Quotative	1st person	N	%	3rd person	N	%	2nd person	N	%	no person	N	%	
be like		16	**64**		8	**32**		0	**0**		1	**4**	
say		3	**23**		10	**77**		0	**0**		0	**0**	
zero		0	**0**		1	**100**		0	**0**		0	**0**	
other		1	**20**		4	**80**		0	**0**		0	**0**	
Total		**20**	**45**		**23**	**52**		**0**	**0**		**1**	**2**	**44**

MM

Quotative	1st person	N	%	3rd person	N	%	2nd person	N	%	no person	N	%	
be like		3	**60**		2	**40**		0	**0**		0	**0**	
say		9	**56**		7	**44**		0	**0**		0	**0**	
zero		2	**67**		1	**33**		0	**0**		0	**0**	
other		6	**55**		5	**45**		0	**0**		0	**0**	
Total		**20**	**57**		**15**	**43**		**0**	**0**		**0**	**0**	**35**

Male speakers

UM

Quotative	1st person	N	%	3rd person	N	%	2nd person	N	%	no person	N	%	
be like		2	**17**		3	**25**		0	**0**		7	**58**	
say		6	**38**		8	**50**		2	**13**		0	**0**	
zero		2	**20**		3	**30**		0	**0**		5	**50**	
other		2	**67**		0	**0**		0	**0**		1	**33**	
Total		**12**	**29**		**14**	**34**		**2**	**5**		**13**	**32**	**41**

LM

Quotative	1st person	N	%	3rd person	N	%	2nd person	N	%	no person	N	%	
be like		11	**58**		6	**32**		1	**5**		1	**5**	
say		7	**33**		14	**67**		0	**0**		0	**0**	
zero		4	**40**		3	**30**		2	**0**		1	**10**	
other		5	**29**		8	**47**		2	**12**		2	**12**	
Total		**27**	**40**		**31**	**46**		**5**	**7**		**4**	**6**	**67**

Table 5.25: Distribution of the quotatives regarding grammatical person across social class and gender.

When social class is stratified by gender, as illustrated in Table 5.25, the higher rate for third-person reporting in the LM class and the equal rate for first- and third-person reporting compared to the MM class becomes more specified. Both the UM women and men have almost equal rates for first- and third-person reporting, which means that they have a similar pattern in which they do not display a specific preference regarding quoting dialogue from themselves vs others. The UM males also have a higher rate of using quotes that have no specified or implied grammatical person. This is likely due to the fact that quotes that tend to be used with no grammatical person are the zero quotatives, *it's like* (which is categorically assigned with no person) and *be*, which are found most frequently with male speakers.

Moving to the factor of tense, Table 5.26 provides the results for the correlation between tense usage and the quotatives in the 18–20 age group, stratified by gender.

female speakers

Quotative	**Present**	N	%	**Past**	N	%	**HP**	N	%	**n/a**	N	%	
be like		28	**18**		110	**72**		15	**10**				
say		4	**10**		34	**83**		3	**7**				
zero		0	**0**		0	**0**		0	**0**		12	**100**	
ask		0	**0**		2	**67**		1	**33**				
decide		0	**0**		1	**100**		0	**0**				
go		14	**78**		2	**11**		2	**11**				
tell		5	**42**		6	**50**		1	**8**				
think		0	**0**		6	**75**		2	**25**				
other		5	**71**		2	**29**		0	**0**				
Total		**56**	**22**		**163**	**64**		**24**	**9**		**12**	**5**	**255**

male speakers

Quotative	**Present**	N	%	**Past**	N	%	**HP**	N	%	**n/a**	N	%	
be like		9	**36**		10	**40**		6	**24**				
say		5	**19**		15	**58**		6	**23**				
zero		0	0		0	**0**		0	**0**		18	**100**	
ask		0	0		3	**100**		0	**0**				
decide		0	0		5	**83**		1	**17**				
go		0	**0**		0	**0**		0	**0**				
tell		0	**0**		0	**0**		0	**0**				
think		1	**33**		2	**67**		0	**0**				
other		1	**20**		4	**80**		0	**0**				
Total		**16**	**19**		**39**	**45**		**13**	**15**		**18**	**21**	**86**

Table 5.26: Distribution of the quotatives regarding tense across gender in the 18–20 age group.

Table 5.26 indicates that the young male and female speakers both use *be like* more with the past tense than the present tense and use it with the HP the least. However, the young male speakers have a more even distribution between the past and present tenses where the difference is merely 4%, whereas the female speakers use it much more with the past than the present tense. Here the difference is 54%. This difference between the genders could indicate an earlier stage of grammaticalisation for the men, where it is still used more with the present tense (as reported in earlier studies (see Chapter 2, section 2.7.4.3)) as compared with the women's use that indicate a later stage, where its past tense use is rising. Even though the HP is not used very frequently in the younger age group, the results indicate that the younger men still use *be like* more with the HP (24%) than the women (10%), another indication that the men in the younger

age group might be lagging behind the women regarding the stage of grammaticalisation of *be like*; this also because previous studies reported that *be like* mainly favoured use with the HP (Winter, 2002:11; Tagliamonte & D'Arcy, 2007:205) and that its association with the past tense is rising among younger speakers (Gardner *et al.*, 2020:299). The results for *say* with both genders remains unchanged in the dataset from what has been reported: that it is mostly used with the past tense (e.g. Blyth *et al.*, 1990:219; Ferrara & Bell, 1995:274). However, similarly to *be like*, the younger men use *say* more with the HP (23%) than the younger women (7%). Previously it was stated that D'Arcy (2012:358) found that the HP context has declined with *say* over the years. Thus, the fact that the younger men use *say* more with the historical present than the women could reflect an earlier stage of the developments in the quotative system for the men in this context as well.

Table 5.27 provides the results for the correlation between tense usage and the quotatives in the 33–48 age group, stratified by gender.

female speakers

Quotative	**Present**	N	%	**Past**	N	%	**HP**	N	%	**n/a**	N	%	
be like		0	**0**		5	**83**		1	**17**				
say		1	**8**		12	**92**		0	**0**				
zero		0	**0**		0	**0**		0	**0**		3	**100**	
ask		0	**0**		1	**100**		0	**0**				
tell		0	**0**		2	**100**		0	**0**				
think		0	**0**		1	**50**		1	**50**				
be		1	**50**		1	**50**		0	**0**				
other		0	**0**		2	**100**		0	**0**				
Total		**2**	**6**		**24**	**75**		**2**	**6**		**3**	**9**	**32**

male speakers

Quotative	**Present**	N	%	**Past**	N	%	**HP**	N	%	**n/a**	N	%	
be like		1	**9**		6	**55**		4	**36**				
say		3	**11**		24	**89**		0	**0**				
zero		0	**0**		0	**0**		0	**0**		5	**100**	
ask		1	**100**		0	**0**		0	**0**				
tell		2	**67**		0	**0**		1	**33**				
think		1	**20**		3	**60**		1	**20**				
be		0	**0**		1	**33**		2	**67**				
other		1	**50**		1	**50**		0	**0**				
Total		**9**	**15**		**35**	**61**		**8**	**14**		**5**	**9**	**57**

Table 5.27: Distribution of the quotatives regarding tense across gender in the 33–48 age group.

The older group's usage of *be like* regarding tense is the same as the younger group. Both men and women use it more with the past tense than the present or HP tenses. Interestingly, the older male group's pattern is similar to the younger male group in the sense that they also have more present tense usage with *be like* than the older women (when combining the results for the present and HP tenses: 45% / N = 5, men vs 17% / N = 6, women). This similar pattern suggests that the older men also reflect a lag in the grammatical development of *be like* where it was previously found more with the present tense. This idea is further supported by the fact that, similar to the younger male group, the older male group also have more *be like* usage with the HP than the older women. However, since the numbers for *be like* in the older group for both genders are very low, this pattern would need to be verified in the context of a larger sample.

The older group's use of quotative *say* is similar to the younger group; in both age groups, both genders use it predominantly in the past tense. However, the older group did not use it in the historical present. Since the HP is infrequent in the dataset overall, this result is likely due to the low number of older participants, which makes it more difficult to capture HP usage. However, the males in both age groups consistently exhibit more use of the HP than their female counterparts. For example, the older male group obtained an overall percentage of HP usage of 14% vs 6% for the older female group and the younger male group obtained 15% vs 9% for the younger female group. If the HP use with quotatives has declined over the years, as observed by D'Arcy (2012:358) for *say* and *go*, then the female groups indicate the erosion of the HP with quotatives to a greater extent than the male groups. Table 5.28 shows the results for quotative distribution regarding tense across social class.

UM

Quotative	present	N	%	past	N	%	HP	N	%	n/a	N	%	
be like		10	**30**		21	**64**		2	**6**				
say		5	**13**		31	**78**		4	**10**				
zero		0	**0**		0	**0**		0	**0**		11		
other		4	**31**		7	**54**		2	**15**				
Total		**19**	**20**		**59**	**61**		**8**	**8**		**11**	**11**	**97**

MM

Quotative	present	N	%	past	N	%	HP	N	%	n/a	N	%	
be like		22	**18**		86	**72**		11	**9**				
say		2	**6**		27	**82**		4	**12**				
zero		0	**0**		0	**0**		0	**0**		16		
other		21	**39**		28	**52**		5	**9**				
Total		**45**	**20**		**141**	**64**		**20**	**9**		**16**	**7**	**222**

LM

Quotative	present	N	%	past	N	%	HP	N	%	n/a	N	%	
be like		7	**16**		24	**55**		13	**30**				
say		6	**18**		27	**79**		1	**3**				
zero		0	**0**		0	**0**		0	**0**		11		
other		7	**32**		10	**45**		5	**23**				
Total		**20**	**18**		**61**	**55**		**19**	**17**		**11**	**10**	**111**

Table 5.28: Distribution of the quotatives regarding tense across social class

All three social classes use *be like* more with the past tense than with the present and HP tenses. However, some differences can be observed. The middle class has the highest rate of past tense usage with *be like* (72%) which indicates that its advancement into past tense usage vs its

previously favoured contexts of present and HP indicates that its grammatical development seems more advanced in the middle class, which correlates with previous observations that the middle class is the origin point for the spread of this quotative. Another difference is that the highest percentage of HP use was obtained in the lower middle class (30%), which could suggest a slower rate of grammatical development compared to the UM and MM groups, since the higher HP percentage is accompanied by the lowest past tense usage out of the three social classes (55%) This result could support Cukor-Avila's (2012:622) suggestion that *be like* is diffusing hierarchically from the middle class to the lower classes. In all three social classes, *say* indicates the usual dominant association with the past tense. D'Arcy (2012:363) observed that in NZE, *say* used with the HP tense developed a correlation with social class, where it was significantly associated with non-professional speakers. However, this is not observed in the current dataset as all three social classes have very low numbers of *say* with the HP and it was used the least in the HP in the lowest social group – the LM. The lower middle class also obtained the highest frequencies for the HP overall, (17% vs 9% for the MM and 8% for the UM). This possibly indicates a slower rate of HP erosion with quotatives in the LM group, similar to both the younger and older male groups.

Table 5.29 provides the results for the correlation of tense with *say* and *be like* across social class, stratified by gender.

female speakers

UM

Quotative **present**	N	%	**past**	N	%	**HP**	N	%	**n/a**	N	%	
be like	3	**14**		17	**81**		1	**5**		0		
say	1	**4**		23	**96**		0	**0**		0		
zero	0	**0**		0	**0**		0	**0**		1		
other	4	**40**		4	**40**		2	**20**		0		
Total	**8**	**14**		**44**	**79**		**3**	**5**		**1**	**2**	**56**

MM

Quotative **present**	N	%	**past**	N	%	**HP**	N	%	**n/a**	N	%	
be like	22	**19**		82	**72**		10	**9**		0		
say	1	**6**		13	**76**		3	**18**		0		
zero	0	**0**		0	**0**		0	**0**		13		
other	20	**47**		20	**47**		3	**7**		0		
Total	**43**	**23**		**115**	**61**		**16**	**9**		**13**	**7**	**187**

LM

Quotative **present**	N	%	**past**	N	%	**HP**	N	%	**n/a**	N	%	
be like	4	**16**		16	**64**		5	**20**		0		
say	3	**23**		10	**77**		0	**0**		0		
zero	0	**0**		0	**0**		0	**0**		1		
other	1	**20**		2	**40**		2	**40**		0		
Total	**8**	**18**		**28**	**64**		**7**	**16**		**1**	**2**	**44**

male speakers

UM

Quotative **present**	N	%	**past**	N	%	**HP**	N	%	**n/a**	N	%	
be like	7	**58**		4	**33**		1	**8**		0		
say	4	**25**		8	**50**		4	**25**		0		
zero	0	**0**		0	**0**		0	**0**		10		
other	0	**0**		3	**100**		0	**0**		0		
Total	**11**	**27**		**15**	**37**		**5**	**12**		**10**	**24**	**41**

MM

Quotative **present**	N	%	**past**	N	%	**HP**	N	%	**n/a**	N	%	
be like	0	**0**		4	**80**		1	**20**		0		
say	1	**6**		14	**88**		1	**6**		0		
zero	0	**0**		0	**0**		0	**0**		3		
other	1	**9**		8	**73**		2	**18**		0		
Total	**2**	**6**		**26**	**74**		**4**	**11**		**3**	9	**35**

LM

Quotative **present**	N	%	**past**	N	%	**HP**	N	%	**n/a**	N	%	
be like	3	**16**		8	**42**		8	**42**		0		
say	3	**14**		17	**81**		1	**5**		0		
zero	0	**0**		0	**0**		0	**0**		10		
other	6	**35**		8	**47**		3	**18**		0		
Total	**12**	**18**		**33**	**49**		**12**	**18**		**10**	**15**	**67**

Table 5.29: Distribution of the main quotatives regarding tense across social class and gender

Table 5.29 shows that the male UM speakers have less of a difference between present and past reporting (10%) vs the other subgroups, which indicates that they use the present tense with quotatives more often than the other subgroups. Both the male and female speakers in the LM class use the HP more frequently than both genders in the other classes, indicating that the previous overall result in which the LM obtained the highest percentage of HP usage is not skewed to any one gender.

For the female speakers across all social classes, *be like* is used predominantly with the past tense, with little difference between them. This result is the same for the male speakers in the middle and lower middle classes. Thus, only the UM male speakers differ in their use of *be like* regarding tense. This result could indicate that the male UM speakers reflect the earlier grammatical stage of *be like* where it was used more with the present than the past tense. In addition, the male UM group does not have any older speakers and thus the difference is even more suggestive since older speakers are not skewing the results towards earlier quotative patterns. In fact, Table 5.29 illustrates that the older male speakers all pattern according to the rest of the community in using *be like* more frequently in the past than present tense. If the results for the present and HP tenses are combined as in the earliest studies of *be like*, then both the UM and LM male speakers obtain more present tense usage than past tense (58% / N = 8 for the UM and 67% / N = 11 for the LM), showing a clear difference for the LM and UM males as compared to the LM, MM and UM females and MM males. One caveat for the MM male group is that it consists of one younger and one older male, and the older male did not use *be like* at all. Thus, it is not to say that the same pattern of past over present reporting for MM males would hold with a larger sample. No difference between the two genders within the social classes are observed for *say.* It is uniformly mainly used with the past tense.

5.1.12 Statistical analysis of the linguistic and social factors of *be like*, *say* and *think*

The statistical analysis will first provide conditional inference trees for the two age groups separately in order to focus on the significant factors for each group before combining them. Secondly, the c-trees for each age group will not include the factor of social class since each group only has one speaker in certain classes and the older age group does not have UM male or female representatives. For the c-tree analysis, the zero quotative was excluded since it was not coded for tense.

Figure 5.19 presents the c-tree for the 18–20 age group and includes the factors of age, gender, quote content, grammatical person and tense. Since the HP is rare in the dataset, it has been grouped with “present” tense as in earlier studies in order to focus only on present vs past tense

usage of the quotatives. For grammatical person, second and “no person” has been combined in the category of “other” since these items are also rare in the data. Thus, the analysis will only focus on the binary choice of first vs third person. For quote content, the data was coded for “thought” and “speech”; anything else that did not fit into those two categories were coded under an “n/a” category. A detailed description for categorising the quotatives according to thought, speech and “other” was provided in Chapter 3, section 3.8.2.

The n/a label represents all quotatives that did not fit into either speech or thought reporting – these are quotatives that introduced non-lexicalised sounds and gestures:

(137) …cause I know like […] **they're like**, "guh-guf" <O>hand gesture</O> and then like you go over, hey (<$R><s n=105>)

quotes that were incomprehensible in the Zoom recording:

(138) they **were** just **like**, "<unclear>". (<$B><s n=36>)

unrealised quotes:

(139) (37) And I WAS just LIKE <,> I like lost my gasket and we broke up. (<$J><s n=70>)

and written dialogue, as in:

(140) I got that message where I think someone **said** "if you want to be involved in the study" or whatever. (<$C><s n=89>)

In Figure 5.19, the highest determining factor is tense, however, this relates purely to the the n/a category (which consists of the zero quotative which could not be coded for tense). Thus, the earliest factor of real significance ($p < 0.001$) is quote content, split between the internal and n/a category (which relates to the quote content that did not fit into either speech or thought, as described above) on the one hand and speech on the other.

Node 4 indicates that within the younger group, *be like* (16% / N = 53 vs zero: 4% / N = 14 and *think*: 3% / N = 11; percentages calculated out of the total quotatives for the men and women combined; see Table 5.18) is preferred above all the other quotatives for internal thought reporting. Within speech reporting, gender is significant at $p < 0.001$ and is the next significant factor for the younger speakers. Within speech reporting in the female group, grammatical person is significant ($p <0.001$) and is split into first person and other on the one side and third person on the other side. Within the context of grammatical person, tense is significant for the female speakers. First-person reporting (with the inclusion of the “other” category) is split into the past

and present tense and is significant at p = 0.001. Likewise, third-person reporting is split into the past and present tenses (p = 0.03). As Table 5.22 indicated, the younger female speakers use *be like* with the first person 59% of the time (vs 36% for third person). Table 5.22 also indicated that they use it with the past tense 72% of the time (vs 28% for the present and HP combined).

Nodes 8 and 9 reveal that the younger female speakers use *be like* much more in the past than the present tense when reporting their own speech (over 60% of the time). Within third-person reporting, nodes 11 and 12 indicate that the younger females use *be like* with the past and present tenses almost equally, with past tense slightly favouring.

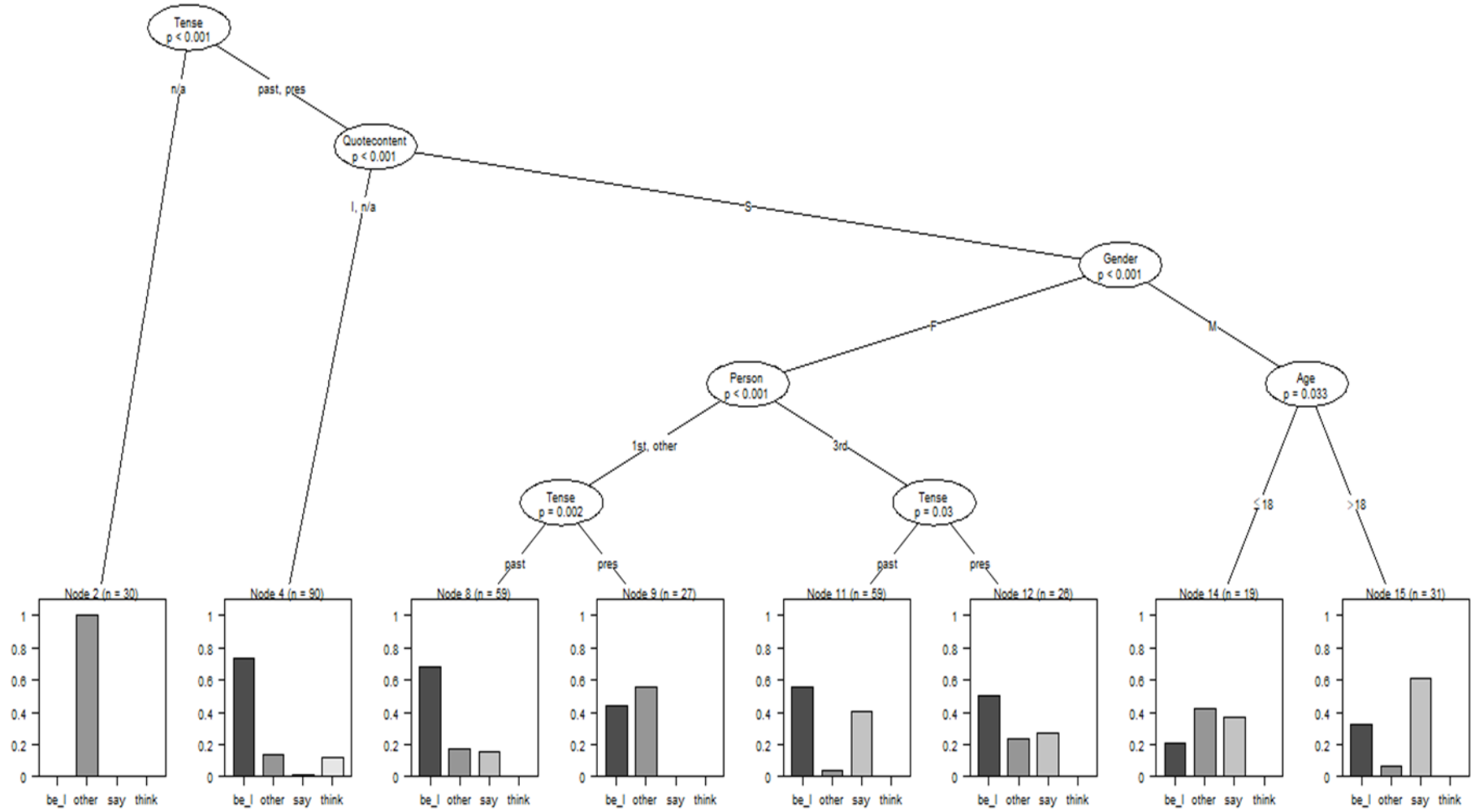


Figure 5.19: The social and internal factors constraining quotative use in the 18–20 age group.

This result shows that the younger females use *be like* more frequently in the present tense with the third person. The results also indicate (nodes 8 and 9) that when *say* is used to report the female speakers' own speech (first person), it is only used with the past tense and not in the present. For third-person reporting, which is what *say* is associated with overall (see Table 5.17), nodes 11 and 12 indicate that it is used more with the past than the present tense. For both first- and third-person reporting, nodes 9 and 12 indicate that *be like* is used more with the present tense than *say*. However, the distributional results (Table 5.26) indicated that younger female speakers tend to use both *be like* and *say* more with the past than the present tense and the c-tree confirms that this result is significant (nodes 8– 9 & 11–12). Thus, for the younger female speakers, the internal factors of content of the quote, grammatical person and tense are significant influences constraining the use of *be like* – it is most frequent with speech and first-person reporting in the past tense vs present tense; additionally, the female speakers use *be like* more frequently in the present tense with the third person (node 12) than with the first person (node 9).

For the male group, unlike the female group, the c-tree indicates that age is significant and divides the male speakers into 18 and younger and older than 18. There is only one 18-year-old male speaker in the younger group, thus node 14 displays the results for this speaker (Jeremy). For the male speakers who are older than 18, node 15 indicates that *say* is used significantly more than *be like* with speech reporting (see Table 5.18). This contrasts with the female group who use *be like* much more than *say* in every instance, from node 8 to 12. The distributional results (see Table 5.12) for the younger speakers indicated that the younger men used *say* (18%) more than *be like* (17%), but only with a 1% difference (vs the younger women who used *be like* 60% of the time vs 16% for *say*. Thus, the descriptive results already indicated that the younger women use *be like* much more than *say* and twice as much as the male group (a 30% difference), and the c-tree confirms this result.

It is evident that internal thought reporting is not a variable context for *say* because it is almost categorically excluded from this context in the dataset, only being used once. Rather, the two quotatives that compete with *be like* (56%) for reporting thoughts/internal monologue in the entire dataset are *think* and the zero quotative (see Table 5.15). Regarding the age division in the younger male group, since the c-tree divided the speakers between one 18-year-old and the rest of the male speakers over 18, it can be concluded that, similar to the younger females, age is not a significant factor for the male speakers either (in the context of finer-grained age differences).

In addition, unlike the female group, the grammatical factors of person and tense were not significant in the male group. This was suggested by the distributional analysis which indicated that the younger men reached equal proportions for *be like* with first and third person (see Table 5.22) and almost equal proportions for it with the past and present tenses (see Table 5.26). This

result, however, is likely due to the low number of quotatives obtained for the younger male speakers.

Figure 5.20 presents the c-tree results for the internal and external factors influencing *be like*, *say* and *think* in the 33–48 age group. In the older group, tense is the most significant factor, however, this can be ignored since it involves the n/a category (consisting of the zero quotative). Thus, the most meaningful factor for the older speakers (similar to the younger speakers; see Figure 5.19) is quote content ($p < 0.001$), divided between thought combined with the n/a category (which consists of non-lexicalised sounds etc.) and speech reporting. Table 5.19 revealed that the older women and men have almost equal rates of reporting thought for all quotatives combined (22% for the women and 23% for the men). Within speech reporting, tense is a significant factor for the older speakers. For the past tense, the c-tree indicates that gender is a significant influencing factor for quotative usage. Node 7 indicates that the older women use *be like* more with the past tense as compared with the older men (node 8). This is largely due to the fact that the older men used *be like* more for thought reporting than the women (see Table 5.19). They also used *think* more (9% / N = 5) than the women (6.25% / N = 2; see Table 5.19 for the Ns of *think*. The percentages reported here for *think* were calculated out of the total quotatives for each gender in the older group in Table 5.19). In addition, one instance of *say* was used by a male speaker to report thought: the only instance of *say* in this category in the entire dataset. Node 4 indicates that *think* is the second most frequent quotative in the thought category (see the N results in Table 5.15). The c-trees for the younger and older groups also illustrate that older speakers use *think* more than younger speakers (compare node 4 in Figure 5.19 to node 4 in Figure 5.20. Unlike the younger speakers (Figure 5.19, node 4), *think* competes with *be like* to some extent in the context of thought reporting in the speech of the older speakers (Figure 5.20, node 4).

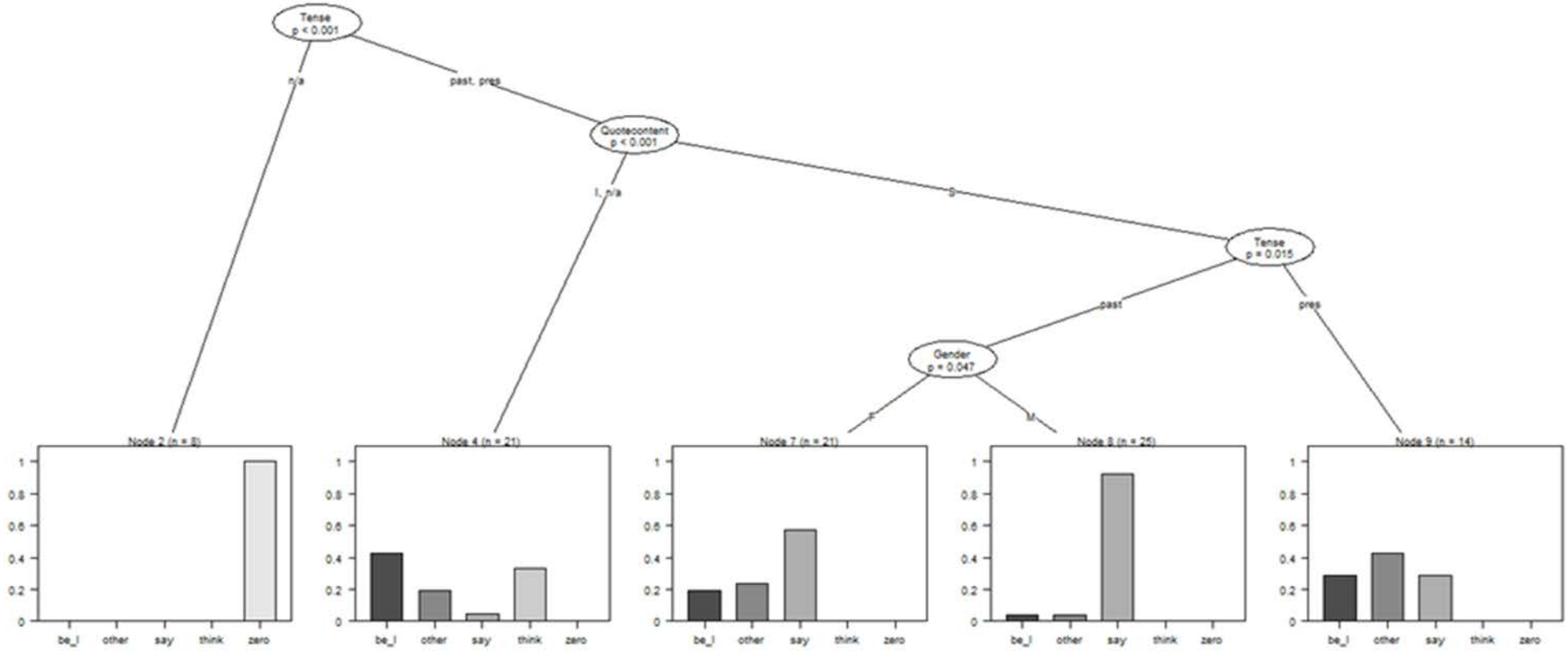


Figure 5.20: The social and internal factors constraining quotative usage in the 33–48 age group.

5.1.13 Statistical analysis: *Be like* vs all other quotatives

This section provides a conditional inference tree of *be like* vs all other quotatives with regards to internal and external factors. The purpose of the c-tree is to discover the significant factors constraining *be like* use in the WSAE sample. Figure 5.21 displays the results of the c-tree. Figure 5.21 shows that, when all factors are considered, gender is the most important factor influencing the use of *be like*. However, the role of gender is only understandable when correlated with age in the female group and (mainly) quote content in the male group. Within the female group, the highest determining factor is tense. However, this relates purely to the n/a category which is a collection of disparate categories and thus not easy to analyse. The next significant factor in the female group is age and indicates that the younger females use *be like* vs the other quotatives more than the older females (nodes 5 & 6). Thus, it confirms the distributional results in Table 5.12).

Within the male group, quote content is significant. However, speech and thought are combined into one group, indicating that there is no important difference regarding speech vs thought reporting for the men. This result is suggested by the distributional results (Table 5.18 & Table 5.19) where the tables indicate that, although both genders in the younger group used *be like* more with speech vs thought, the older men use it more with thought vs speech. Whereas tense (regarding past vs present contexts) is not significant for female speakers, it *is* significant for male speakers (nodes 10 & 11), indicating that the men use *be like* more with the present (node 11) than the past tense (node 11). This confirms the distributional results which suggested that men use the present tense more than the past tense with *be like* as compared to the women, who did not obtain a significant tense effect (see Table 5.26 & Table 5.27 for present & HP Ns combined in the younger and older male groups). In addition, that the younger men used *be like* more with the present than the past tense (Table 5.26; when combining Ns for the present and HP tenses) contributes to the significant tense effect for the men. The c-tree also reports that there is no significant difference between speech and thought reporting for either gender in relation to the use of *be like*. Another factor that fails to be significant is grammatical person. Thus, the only factor that significantly influences *be like* use within the WSAE sample when the social and internal factors are considered together is gender – with age as a significant factor in the female group and tense as a significant factor in the male group.

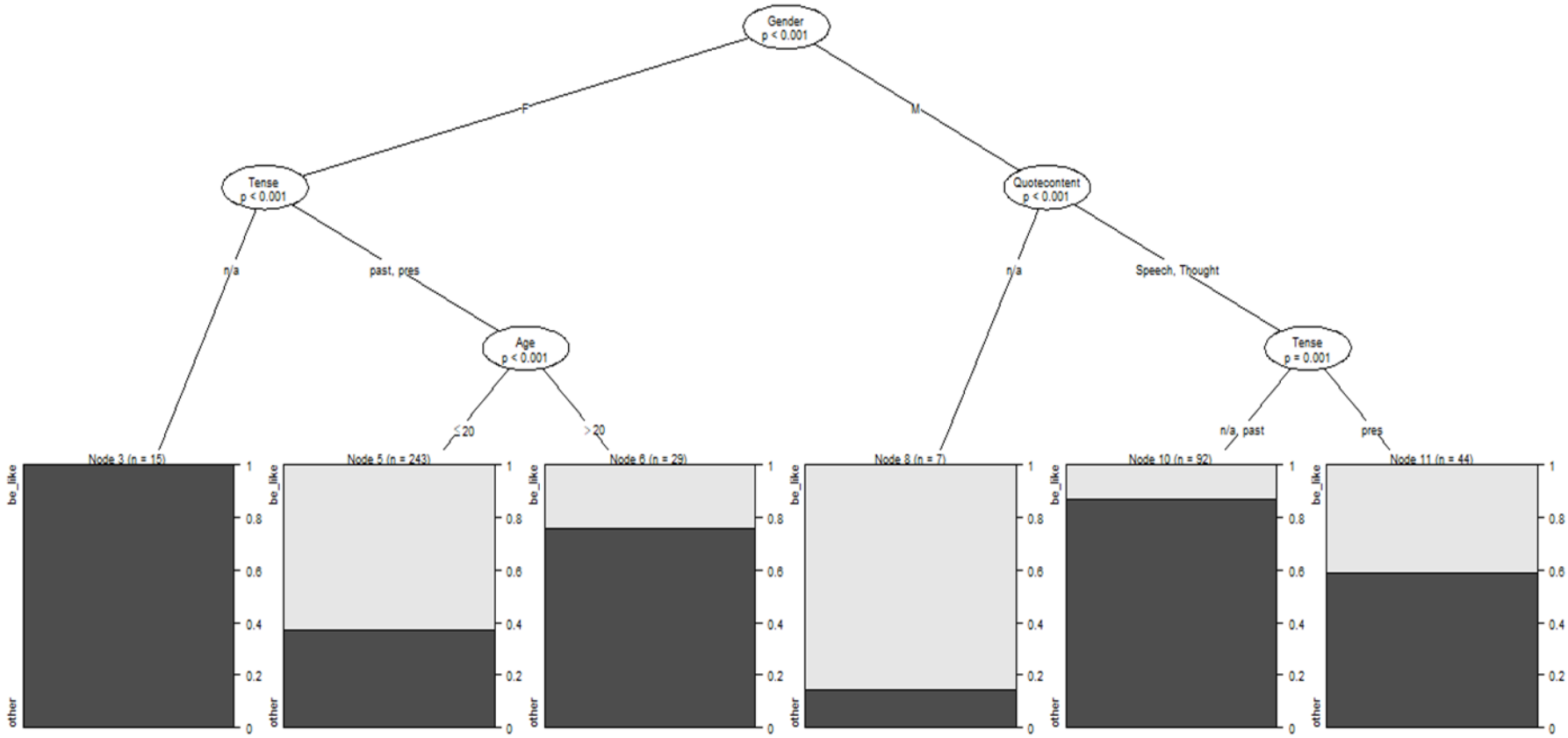


Figure 5.21: be like *vs all other quotatives correlated with all factors*[36]

[36] (Both structural (internal) and social (external); grey is *be like*.

5.2 Analysis of DM-*Like*

This section investigates the uses of *like* as a discourse-marker in the Port Elizabeth, WSAE sample. These are uses of the word that have a pragmatic effect and little semantic meaning. The older and more traditional categories of *like* that were found in the dataset are the following:

1. Verb: I do **like** to swim. (<$M><s n=70>)

2. Preposition: My brother just treats me **like** a little baby. (<$H><s n=78>)

3. Conjunction: I feel **like** we dodged a lot of drama when we were younger (<$J><s n=74>)

4. Compound adjective: It's like a gathering of people that are all **like-minded**, you know (<$R><s n=58>)

In this preliminary investigation, the variable context of *like* will not form part of the analysis (for example, its distribution vs other discourse markers such as *so* and *just* (see Tagliamonte, 2005). This section focuses only on the social factors influencing the use of *like* (gender, age and social class), as well as the internal factor of its structural distribution i.e. whether it appears in either clause-initial or clause-medial position (discussed in more detail later in this chapter).

Overall, there were 3,033 instances of *like* in the dataset out of a total of 115,691 words. Of these, 2423 were instances of discourse-marker *like* (including all approximative uses, see examples above); as well as 127 verbs, 215 prepositions, 196 quotative *be like*, 68 conjunctions and 4 compound-adjectives. It is apparent that the DM uses of *like* were far more than any of the other categories, while prepositions were the second most frequent category, followed by quotative *be like* and the verb uses. Table 5.30 and Figure Figure 5.22 display the percentage of DM-*like* vs Non-DM-*like* out of the total *like* tokens in the dataset.

Like	DM like	%
DM-*like*	2423	**80**
Non-DM-*l*	610	**20**
Total:	3,033	

*Table 5.30: Distribution of DM-*like *vs Non-DM-*like

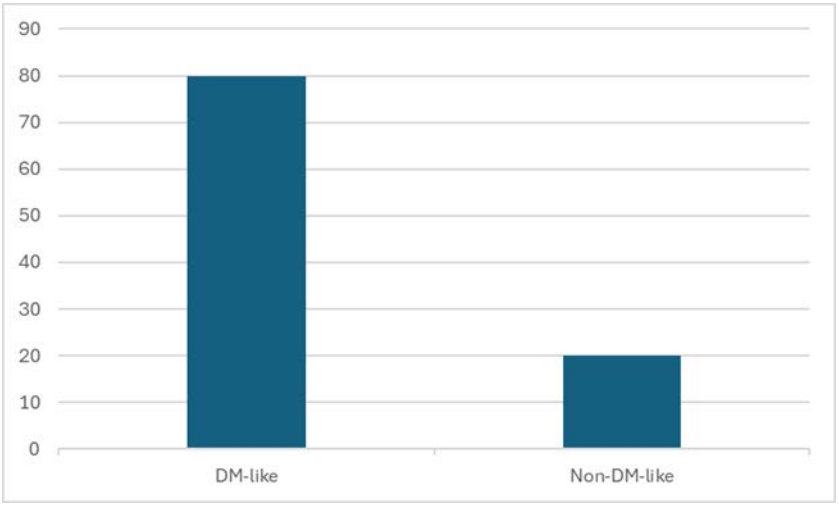


*Figure 5.22: Percentage of DM-*like *vs Non-DM-*like

5.2.1 Distributional results of DM-*like*

Following the method of Tagliamonte (2005:1904), normalisation of the results for DM-*like* was achieved by calculating the percentages of *like* out of the total number of words in the dataset for each social group, obtained through the concordance programme AntConc (the methodology chapter explains this in more detail). Table 5.31 and provide the results for the distribution of discourse-marker *like* across gender in the entire dataset (young and old speakers combined).

The results for the distribution of DM-*like* across gender indicate that, similarly to quotative *be like*, it is also used more by women than men, indicating that in WSAE, women lead in the spread of discourse-marker *like* as well. Thus, the speakers of the current dataset for WSAE pattern according to EngE and CanE, Toronto in Tagliamonte (2005:1912) but not in D'Arcy (2005:43), also for Toronto. Lochner's study (2019:38), which investigated DM-*like* in WSAE with speakers residing in the Western Cape also found that females used it more than men, thus supporting the results for the current data.

Gender	DMlike	%	Total words
Female	1520	2.4	**63025**
Male	903	1.7	**52666**
Total:	**2423**		**115691**

*Table 5.31: Distribution of DM-*like *across gender*

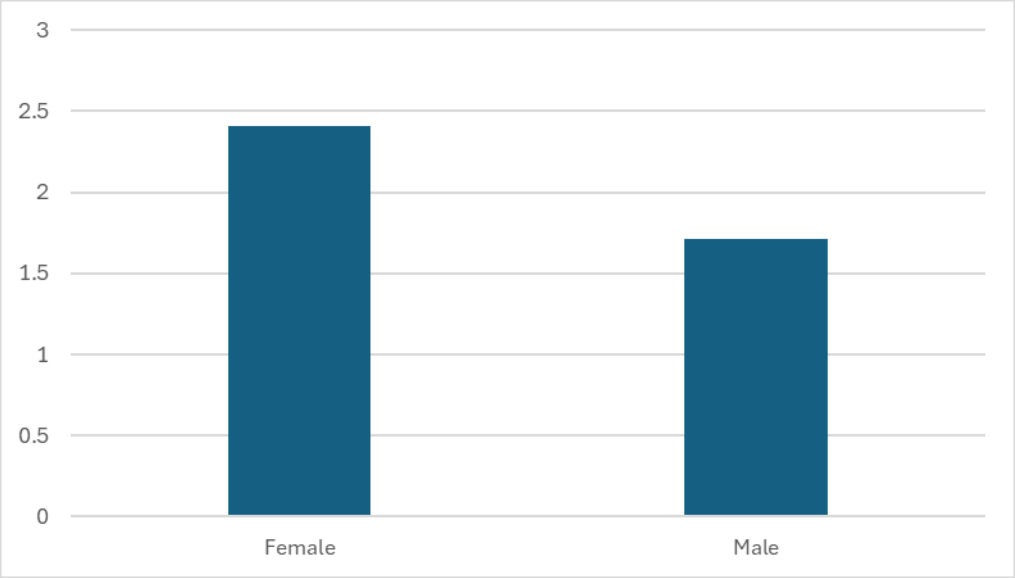


*Figure 5.23: Percentage of DM-*like *distribution across gender*

Table 5.32 and Figure 5.24 presents the distributional results for DM-*like* across age. Not surprisingly, Table 5.32 suggests that the younger speakers use DM-*like* to a much greater extent than the older speakers. Thus, this patterns according to most studies for DM-*like* regarding younger vs older speakers, which suggests that DM-*like* is age-graded in WSAE. Schweinberger (2015:211, 233) also found signs of age-grading rather than real change in progress for DM-*like* in AmE and CanE; however, Schweinberger (2015:243) and Beeching (2016:144) found evidence for language change in EngE regarding the frequencies of DM-*like* between gender and age (see the discussion of these studies in Chapter 2, section 2.7.5.1).

Age	DMlike	%	Total words
18-20	2276	2.6	**85634**
33-48	147	0.4	**30057**
Total:	**2423**		**115691**

*Table 5.32: DM-*like *distribution across age.*

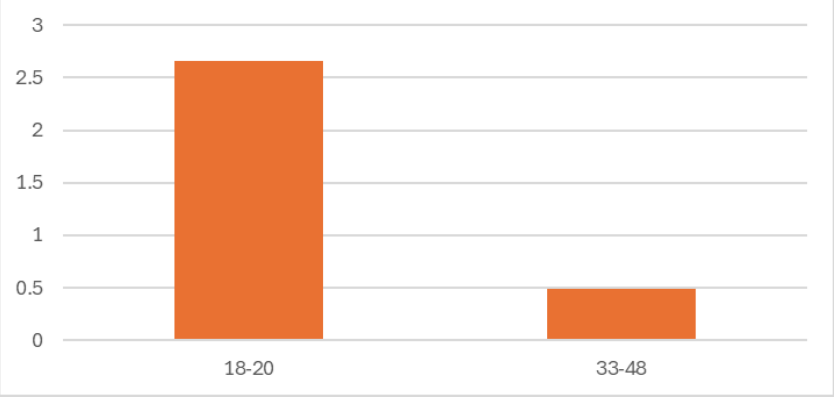


*Figure 5.24 Percentage of DM-*like *across age*

Table 5.33 and Figure Figure 5.25 presents the distributional results for DM-*like* across age and gender. As expected, within the younger group, the females obtained a higher frequency of DM-*like* than the males, however, the young men are not far behind. The younger female lead reflects previous findings in which young women are identified as the group most likely to use DM-*like* (e.g. Andersen, 2001:288; Beeching, 2016:141; D'Arcy, 2017:120). What is surprising is that the older men use more DM-*like* than the older women. This could be due to the fact that there were more male than female speakers (3 vs 2). However, it is more likely because the youngest male speaker in the older group (Simon, aged 33, N = 87), produced the highest number of DM-*like* tokens among the older speakers and rather seems to behave linguistically more like the younger speakers (he was also an outlier in the analysis for *be like*, see the discussion for Figure 5.10 in section 5.1.6).

18-20 age group			
Gender	DMlike	%	Total words
Female	1474	2.8	**52141**
Male	802	2.3	**33493**
Total:	2276		**85634**

33-48 age group			
Gender	DMlike	%	Total words
Female	46	0.4	**10884**
Male	101	0.5	**19173**
Total:	147		**30057**

*Table 5.33: Distribution of DM-*like *across age and gender*

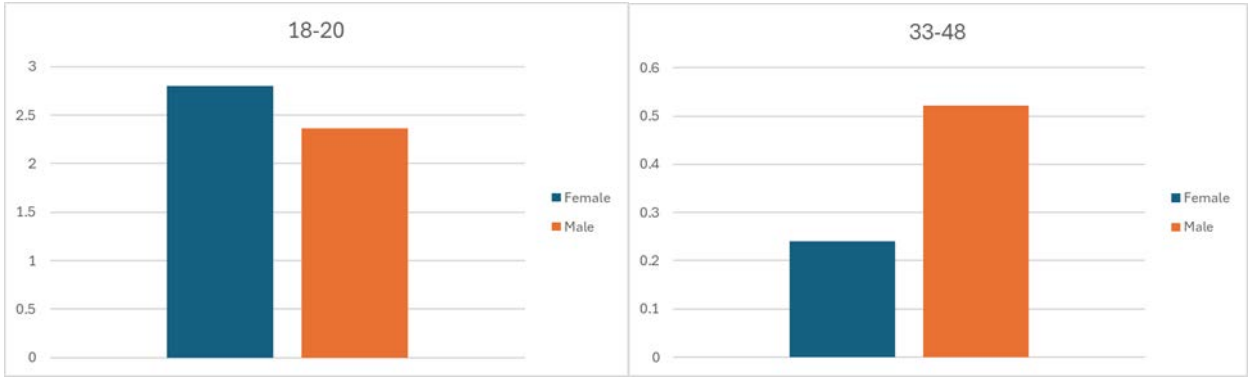


Figure 5.25: Percentage of DM-like across age and gender

When this speaker is taken out of consideration, the two older females also produce more instances of DM-*like* than the two older male speakers (N = 46 vs 13). The fact that Beeching's (2016:141) results for EngE indicated a higher frequency for DM-*like* usage for her 25–34 age group than for her 35–44 age group suggests that the one older speaker who had the highest rate of DM-*like* likely reflects the pattern of speakers in their early thirties and contrasts with the rest of the group whose age starts at 37. In addition, Andersen (2001:225) observed that, also for

EngE (BNC corpus, recorded between 1991–1994), three of his older speakers (34, 38 and 41) produced almost half the tokens of the marker and thus suggests the possibility that it first becomes “established in the speech of certain members of an age group before it spreads to other members of that age group”.

Table 5.34 and Figure 5.26 present the distributional results for DM-*like* across social class. The results for social class indicate that, similar to the quotative, the middle class shows the highest frequency of DM-*like* usage (see Table 5.9 and Figure 5.7), indicating that for both DM-*like* and quotative *be like*, the middle class lead in the spread of both variants. The LM and UM classes obtained equal proportions of DM-*like* tokens and do not trail far behind the MM class. Thus, the middle-class lead is not very large. Andersen (1996:49-50; 2001:290) found it to be used mostly by middle to upper-middle class speakers in EngE. However, he did not obtain a significant difference for social class in statistical testing. On the other hand, Schweinberger (2015:314) found that in NZE DM-*like* was used the most by working-class speakers and that upper-middle class speakers use it the least, while his results for CanE indicated that the middle class use it the most (see the discussion of this study in chapter 2, section 2.7.5). Thus, WSAE patterns towards EngE and CanE regarding middle-class speakers, however, the LM and UM speakers also participate to a great extent in its use. Lochner’s (2019) study of DM-*like* in WSAE did not test for social class and thus it is not possible to compare the results for social class distribution to an additional WSAE sample.

Social Class	DMlike	%	Total words
UM	668	2	**33080**
MM	1055	2.2	**47697**
LM	700	2	**34914**
Total:	**2423**		**115691**

*Table 5.34: Distribution of DM-*like *across social class*

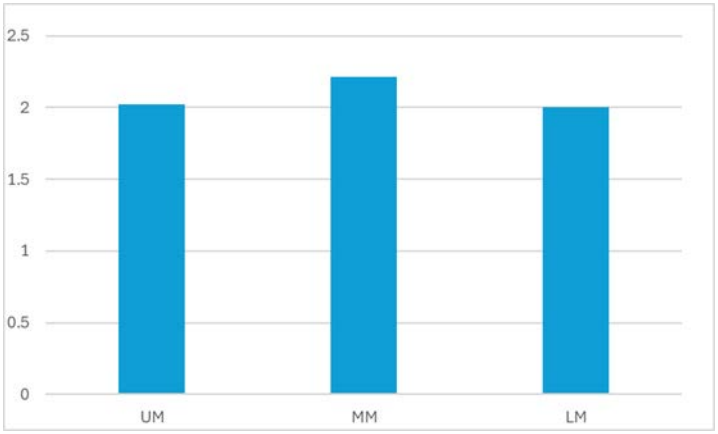


*Figure 5.26: Percentage of DM-*like *distribution across social class*

The next section investigates if there are any differences in social class distribution when the speakers of each gender are viewed separately. Table 5.35 and Figure 5.27 provide the results for the social class distribution of DM-*like* within each gender.

Male			
Social Class	DM like	%	Total words
UM	363	2.3	**15267**
MM	113	0.9	**12322**
LM	427	1.7	**25077**
Total:			**52666**

Female			
Social Class	DM like	%	Total words
UM	305	1.7	**17813**
MM	942	2.6	**35375**
LM	273	2.7	**9837**
Total:			**63025**

*Table 5.35: Distribution of DM-*like *across social class and gender*

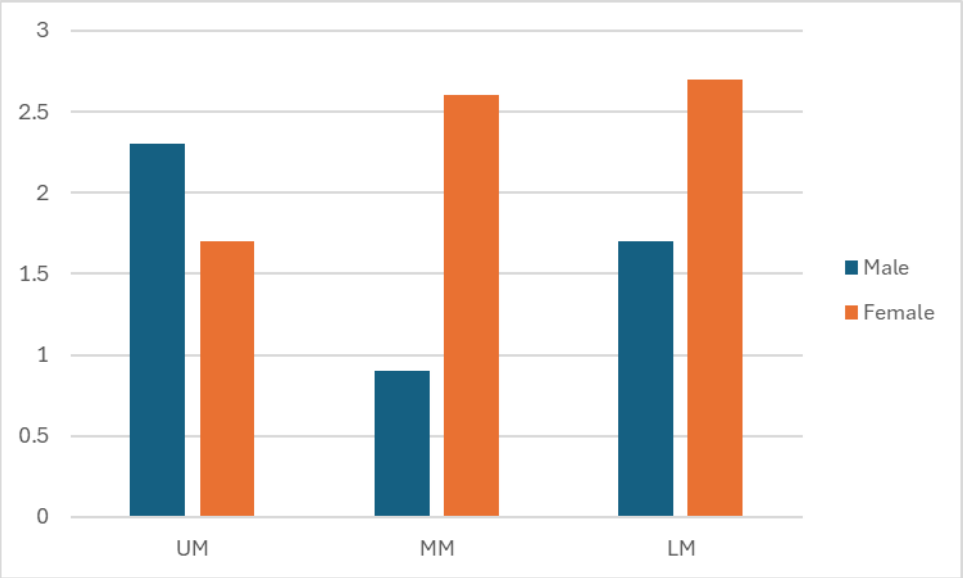


*Figure 5.27: Percentage of DM-*like *across social class and gender*

Table 5.35 indicates that, for the female group, a slightly different picture emerges. DM-*like* is highly frequent among all three social class female groups. However, the LM females lead with the highest percentage of DM-*like*, with the MM females behind by one percent. This seems to reflect the pattern for a change from below, which refers to innovative linguistic features being spread from the lower classes to the higher social classes (Labov, 1966:206). The results also indicate that it is specifically the UM female speakers that use DM-*like* the least (compare Figure 5.26 with Figure 5.25). Surprisingly, the UM males use DM-*like* more than the UM females, whereas the female groups in the other social classes consistently have higher frequencies than their male counterparts. The distributional result for the male MM group could be due to the fact

that the two speakers representing this group consist of one younger speaker and the oldest speaker in the group; the older speakers have tended to use DM-*like* relatively infrequently (see Table 5.32).

When looking at the results for individual speakers, it can be seen that the younger MM speaker (Jeremy) produced the second least DM-*like* tokens (N = 104) out of the other younger male speakers from different social classes. However, it is not possible to say if this is reflective of the pattern for other younger MM male speakers without a comparative corpus or more speakers in this category. The upper-middle class speaker, Alan, produced the fewest (N = 75) tokens of DM-*like* while the lower-middle class male speaker, Owen, produced much more tokens than all of the other younger men (N = 335). He even produced more tokens than most of the female speakers, except the highest DM-*like* user among the females, Michelle (N = 338), who is a middle-class speaker. This suggests that Owen is an outlier among the younger men, patterning more towards the female usage. Thus, it is also due to Owen (N = 335) that the LM male frequency is high (1.7%) since he produced the most tokens of *like* out of the LM male group: Greg (N = 5), Simon (N = 87).

For EngE, Beeching (2016:143) found that with respect to her 15-24-year-old male group, while the working-class men produced the highest rate of DM-*like* use, the upper-middle class men produced the second highest rate (59.78%), with the lower-middle class men following close behind (54.84%). The middle-class men in her sample obtained the lowest rate among the male group (36.64%), which suggests the possibility that Jeremy might reflect the pattern for young middle-class male speakers in general and it might then be the case that this social group favours *be like* and DM-*like* the least out of the other male social classes (refer to Table 5.13 for the results of quotative *be like* across gender and social class). Interestingly, similar to my own data, Beeching (2016:143) found that the lower-middle class female speakers in the 15–24 age group used DM-*like* more (88.19%) than the middle-class female speakers (77.84%). However, her data also indicated that the UM female speakers used it more than the MM female speakers, while my own data indicates that the UM female speakers use it the least. In EngE, Beeching's data also indicated that the UM, MM and LM female speakers used DM-*like* more than their male counterparts (except for the working-class speakers, where the pattern is reversed), while my data suggests that the UM male speakers use it more than the UM female speakers. The next section provides the statistical analysis for the social factors correlated with DM-*like*.

5.2.2 Statistical analysis of the social factors of DM-*like*

Because it was not possible to include the total words of each social group in the c-tree analysis, provides the results for the discourse-marker uses of *like* vs the non-discourse-marker uses (e.g. verb, preposition, etc.).

Figure reports that age is the most significant factor regarding the use of the DM. Within the age group of 33 and under, gender is a significant factor, but only when considering age and social class. The female group is split into a further age divide, where the 19-year-old and younger (18–19) speakers (nodes 5 & 6) differ significantly from the 20-year-olds (nodes 8–9). Within this age split, social class is significant. The c-tree reports that the two LM female speakers differ in their use of *like*. Harriet (19) obtained a lower rate of DM-*like* (node 5[37]) compared to the MM and UM female speakers in the 19 and under age group (node 6). On the other hand, in the 20-year- old age group, the LM female speaker, Kiana, obtained a higher rate for DM-*like* (node 8) than all the other female speakers, which suggests that this speaker is an outlier with a higher-than-average DM-*like* usage). Table 5.35 suggested that the LM female group obtained the highest rate of DM-*like* use vs the MM female speakers. It is apparent that the one LM female speaker (Kiana) contributes the most to this result (compare Kiana, (N = 213) and Harriet (N = 60). Overall, the young middle-class female speakers (18–20) and possibly the young lower-class female speakers produce high frequencies of DM-*like* compared with the UM female speakers. The c-tree also confirms the distributional results which indicated that the UM female speakers use DM-*like* the least (see Table 5.35).

Within the 33 and under male group, social class is also significant and indicates that there is a difference between the LM and UM males' use of DM-*like* and the one MM male (Jeremy). In the LM and UM male group, age is a significant factor and indicates that the 19-year-old and younger men use DM-*like* the most (the grey portion, node 12) out of all the other *like* categories (the dark portion). The c-tree also indicates that this group of young men use DM-*like* the most vs all the other subgroups (compare node 12 with nodes 5–6, 8–9, 13–14, 16–17). This result is likely due to the high proportion of DM-*like* obtained by the UM men and the LM male speaker in the younger group (refer to Table 5.34 and Figure 5.26 for these results). Unlike the results for quotative *be like* (see Figure 5.11, nodes 3 & 7), the c-tree indicates that the younger men also produce DM-*like* at high frequencies (nodes 12–14) as compared with the younger female group (nodes 5–6 & 8–9).

[37] DM-*like* is represented by the light portion while non-DM-*like* is represented by the dark portion. The y-axis indicates the total like tokens (DM-*like*, verb, preposition etc.).

Within the over 33 group, gender is significant and indicates that the older females use more DM-*like* (node 16) than the male speakers (node 17), as was suggested in the discussion of the distributional results for age and gender (see Table 5.33). It is apparent that the older speakers (nodes 16 & 17) do not use as high frequencies of DM-*like* than the younger speakers (all the other nodes), which suggests a degree of age-grading in the use of DM-*like* in WSAE (also suggested by Lochner's study (2019, see discussion in chapter 2, section 2.7.7). For AmE, Schweinberger (2015:197) states that "the older a speaker, the less likely it is that he or she uses this discourse feature" and this finding is apparent in the c-tree results. Schweinberger (2015:198) also states that age-grading is usually not accompanied by social stratification, and the fact that gender and social class obtained significant results in the c-tree suggests that DM-*like* might be undergoing a linguistic change.

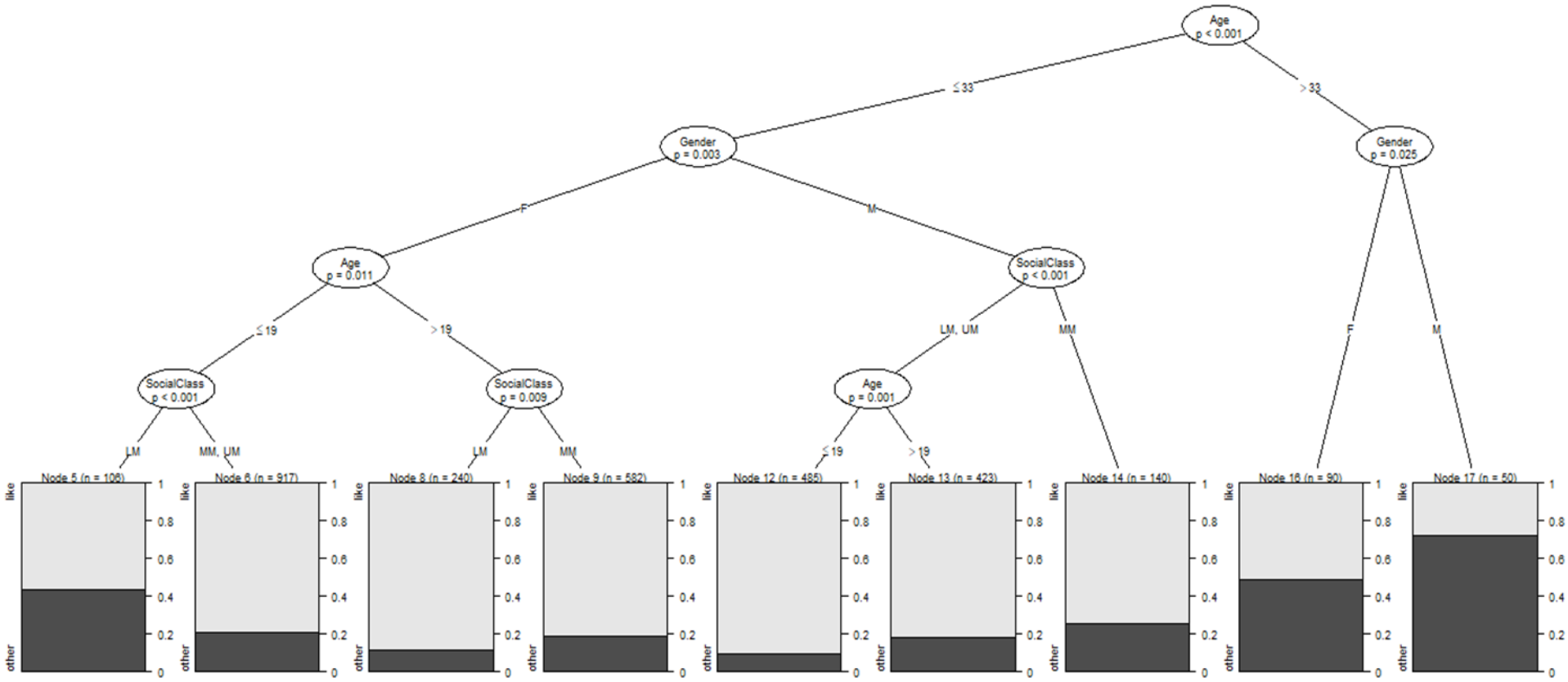


*Figure 5.28: C-tree of social factors correlated with DM-*like *vs non-DM-*like

5.2.3 Structural analysis of DM-*like*

This final subsection investigates the structural position of DM-*like* (in clause-initial or clause-medial position) and aims to discover if the social factors of age, gender and social class influence the choice in structural position.

Table 5.36 and Figure 5.29 illustrate the percentages for clause-initial and clause-medial *like* out of the total DM-*like* tokens.

Clause position	DM like	%
Initial	1010	**42**
Medial	1413	**58**
Total:	**2423**	

*Table 5.36: Structural positions of DM-*like

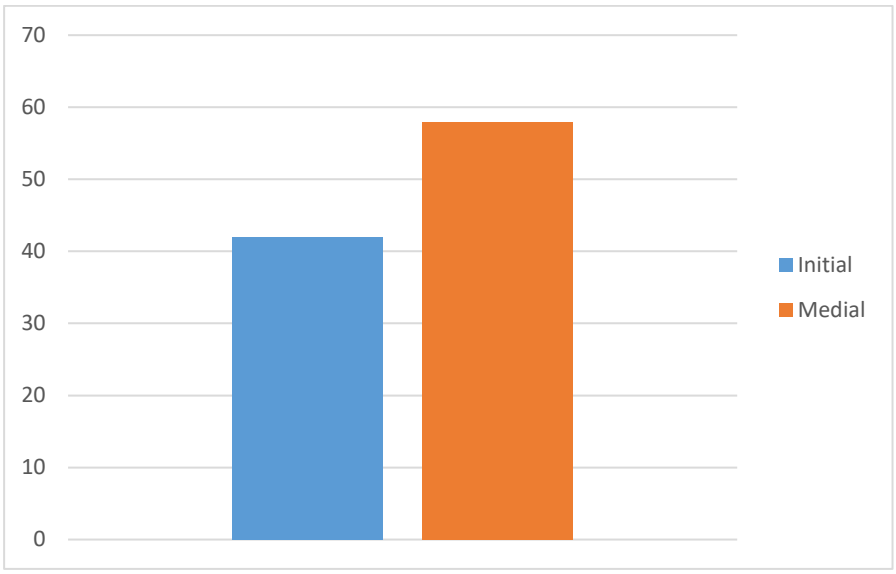


*Figure 5.29: Percentage of the structural positions of DM-*like

The results indicate that clause-medial position is the most frequent structural position for *like* (58%), whereas clause-initial position is not as frequent (42%). This is not surprising since previous studies have found clause-medial position to be the most frequent syntactic context for *like* across most varieties of English (e.g. Andersen, 2001:273, Tagliamonte, 2005:1901; Schweinberger, 2015:179; see Chapter 2, 2.7.5 for a discussion of these studies). The next step is to explore how the clausal position is distributed between the two age groups.

Table 5.37 and Figure 5.30 provide the results for the younger and older age groups. The results reveal that both age groups pattern in a similar manner regarding the clause position of DM-*like*. The younger and older speakers use *like* in clause-medial position the most, followed by clause-initial position. This is similar to D'Arcy's (2005:212) study of the Toronto (CanE) speech community in which it was revealed that younger and older speakers do not differ in their pattern regarding structural position. It is simply that younger speakers have higher frequencies in the same contexts as older speakers.

Age	Clause-initial	%	Clause-medial	%	Total words
18-20	964	1.1	1312	1.5	**85634**
33-48	46	0.1	101	0.3	**30057**
Total:	**1010**		**1413**		**115691**

*Table 5.37: The structural position of DM-*like *distributed across age*

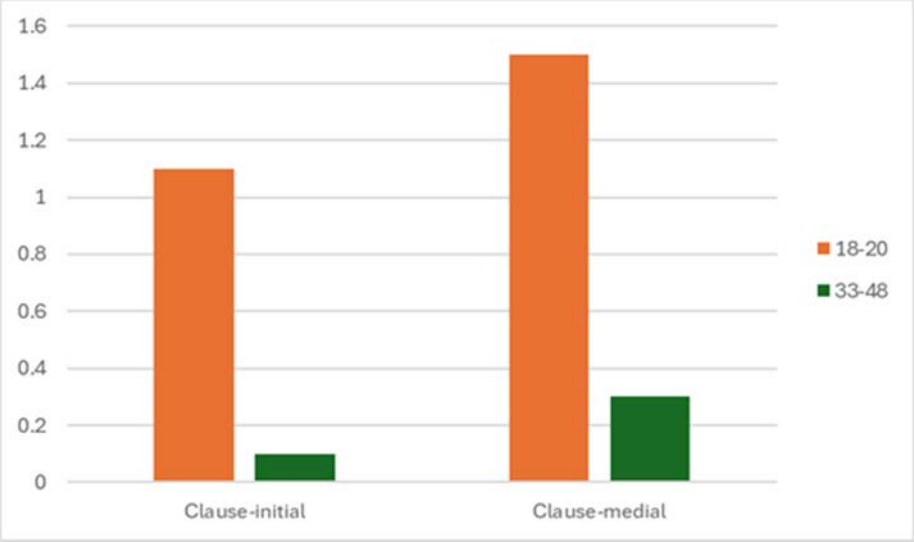


*Figure 5.30: Percentage of the structural position of DM-*like *across age*

Table 5.38 and Figure 5.31 provide the distributional results for clause position across gender, which might reveal differences obscured by the overall results for age. The results for gender regarding the distribution of structural position do not reveal any meaningful differences other than indicating that the women in the sample use DM-*like* more than the men and that both genders use clause-medial *like* more than clause-initial *like*. This result is contrary to what D'Arcy (2005:222) found in CanE where clause-initial position was significantly correlated with female speakers and clause-medial *like* was correlated with male speakers.

Gender	Clause-initial	%	Clause-medial	%	Total words
Female	634	1	886	1.4	**63025**
Male	376	0.7	527	1	**52666**
Total:	**1010**		**1413**		**115691**

*Table 5.38: The structural position of DM-*like *distributed across gender*

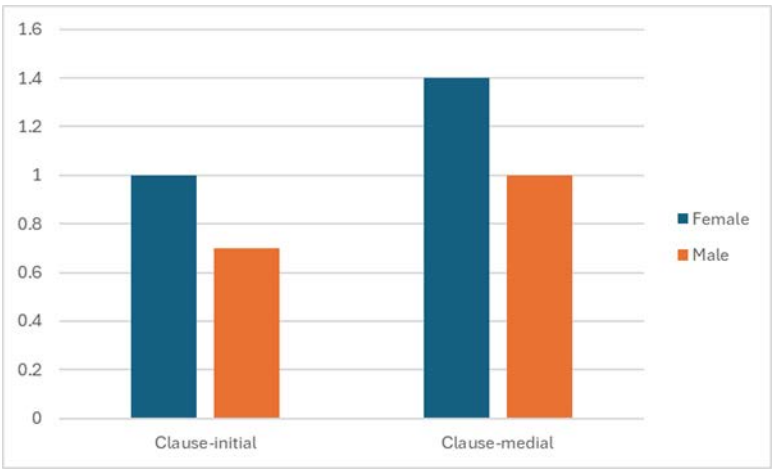


*Figure 5.31: Percentage of the structural position of DM-*like *across gender*

A similar result to D'Arcy's was found in Levey (2006:433) for EngE among preadolescent speakers. However, other studies did not find a significant gender effect for clause-initial or clause-medial *like* (D'aily O'Cain, 2000:66, AmE; Schweinberger, 2015:203, 216, 226, 328–329, AmE, CanE, NZE). Table 5.39 and Figure 5.31 provide the distributional results for clause position across gender and age.

18-20 age group						
Gender	clause-initial	%	clause-medial	%	DM-*like*	Total Words
Female	621	1.1	853	1.6	**1474**	**52141**
Male	343	1	459	1.3	**802**	**33493**
Total:	**964**	1.1	**1312**	1.5	**2276**	**85634**
33-48 age group						
Gender	clause-initial	%	clause-medial	%	DM-*like*	Total words
Female	13	0.1	33	0.3	**46**	**10884**
Male	33	0.1	68	0.3	**100**	**19173**
Total:	**46**	0.1	**101**	0.3	**146**	**30057**

*Table 5.39: The structural position of DM-*like *distributed across age and gender*

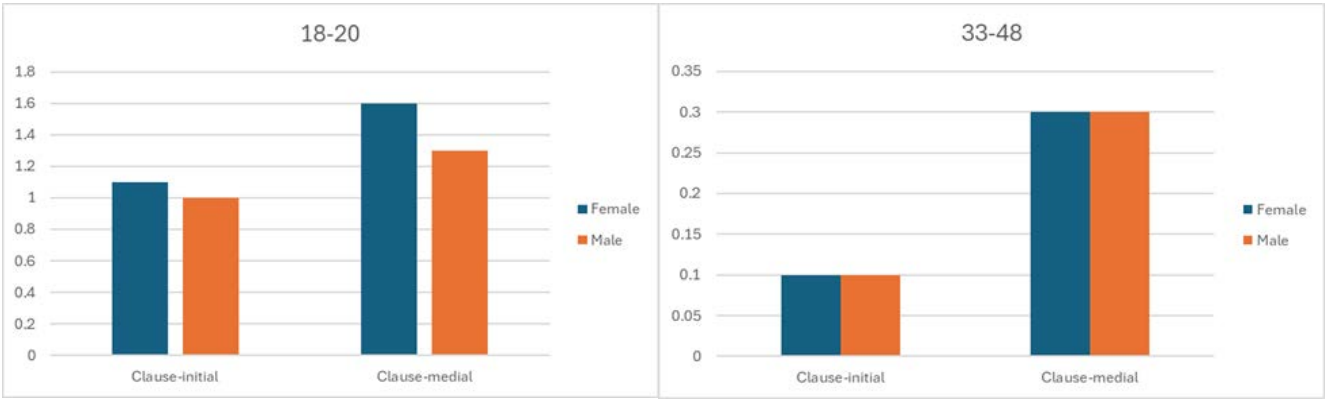


*Figure 5.32: Percentage of the structural position of DM-*like *across age and gender*

Regarding clause position, Table 5.39 reveals that the younger men and women use clause-medial *like* more than clause-initial *like*. The results also reveal that both the older men and women have the exact same rate of clause-initial vs clause-medial position and that they do not differ in this respect. The results therefore reveal that both the clause-initial and clause-medial uses of *like* seem to be age-graded, decreasing in rate from young to old. This result is similar to Schweinberger's finding for clause-initial *like* in AmE in which he found a gender effect among speakers in their twenties and thirties, but no such effect among speakers who are 41 and older. The following section investigates the distribution of clause structure across social class. Table 5.40 and Figure 5.33 provide these results.

The results for social class show no difference between the UM and MM groups regarding clause-initial *like* use, their rates are the same. The LM group, however, displays a higher rate of this feature compared to the other two social classes, suggesting that the lower middle-class speakers tend to use clause-initial *like* the most out of the three social classes. Where clause-medial *like* is concerned, it is the middle class who use it the most, followed by the upper-middle class and lastly the lower-middle class. However, all three social classes use clause-medial *like* more than clause-initial *like*.

Social Class	Clause-initial	%	Clause-medial	%	Total words
UM	274	0.8	394	1.1	**33080**
MM	394	0.8	661	1.3	**47697**
LM	342	0.9	358	1	**34914**
Total:	**1010**		**1413**		**115691**

*Table 5.40: The structural position of DM-*like *distributed across social class*

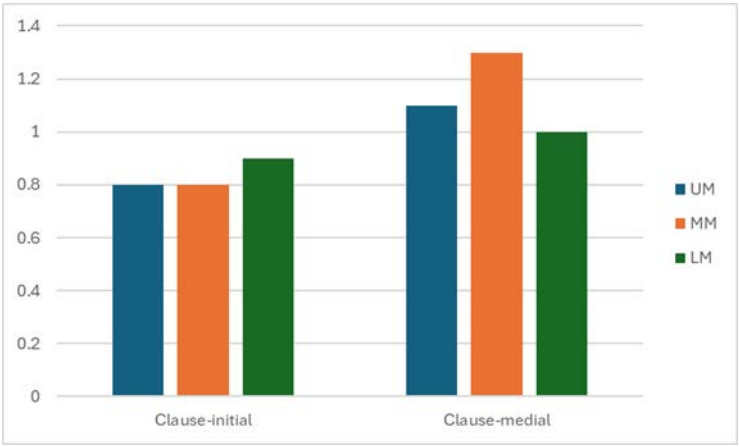


*Figure 5.33: Percentage of the structural position of DM-*like *across social class*

The current data seem to pattern more according to the AmE results in Schweinberger (2015:203, see chapter 2, section 2.7.5.2 for the discussion of this study) and it might be that social class is not a significant factor influencing the use of clause-initial vs clause-medial *like*.

Table Table 5.41 and Figure 5.33 provide the results for social class and gender vis-à-vis the structural position of DM-*like*. Firstly, Table 5.41 indicates that across social class, the UM and MM female speakers use clause-medial *like* more than clause-initial *like*. Surprisingly, the LM female group obtained the opposite pattern, where clause-initial *like* has the higher frequency. It is also the LM female group who use clause-initial *like* the most compared to the other social classes. For clause-medial *like*, the middle-class female speakers use it the most.

Women					
Social Class	Clause-initial	%	Clause-medial	%	Total words
UM	136	0.7	169	0.9	**17813**
MM	351	0.9	591	1.6	**35375**
LM	147	1.4	126	1.2	**9837**
Total:	**634**	1	**886**	1.4	**63025**

Men					
Social Class	Clause-initial	%	Clause-medial	%	Total words
UM	138	0.9	225	1.4	**15267**
MM	43	0.3	70	0.5	**12322**
LM	195	0.7	232	0.9	**25077**
Total:	**376**	0.7	**527**	1	**52666**

*Table 5.41: The structural position of DM-*like *distributed across social class and gender*

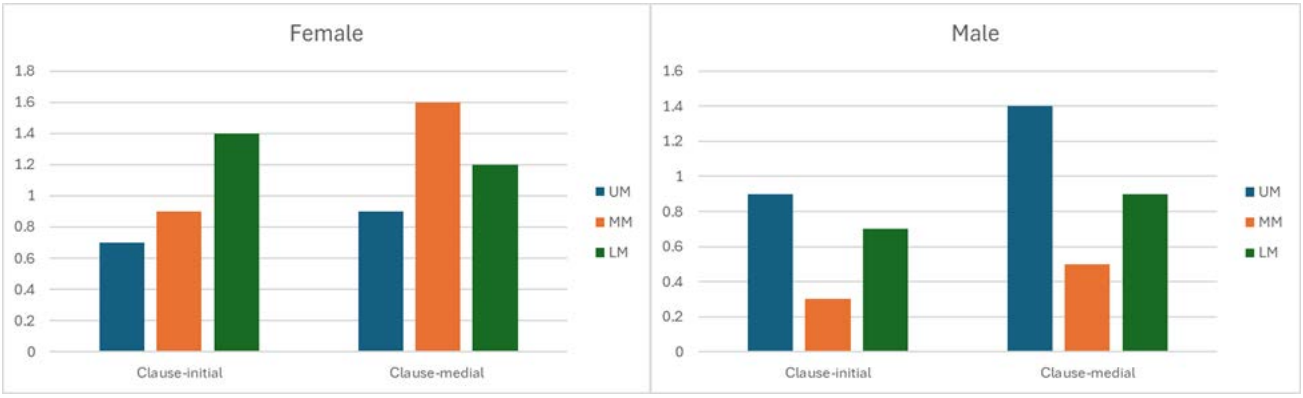


*Figure 5.34: Percentage of the structural position of DM-*like *across social class and gender*

Interestingly, the upper-middle class female speakers use clause-initial and clause-medial *like* the least compared to the other social classes, indicating that in both structural contexts, the UM female class tend to be more conservative in their use of DM-*like* compared to the other social classes. Within the male group, the upper-middle class men obtained the highest proportion of DM-*like* use for both structural contexts. In addition, the male UM class use clause-medial *like* more than female LM class. However, it is not clear at this stage if this result means anything in terms of previous findings (such as D'Arcy, 2005; see discussion of this study in chapter 2, section 2.7.5.3) in which women use clause-initial *like* more than men while men use clause-medial *like* more than women since the MM female group use clause-medial *like* the most in the current data.

In the discussion for Table 5.34, it was stated that the current data seem to indicate a trend similar to Schweinberger's (2015:314) results for NZE in which he found that that the upper-middle class used DM-*like* the least out of the social classes. The fact that clause-initial and clause-medial *like* are used the least by UM females vs LM, MM females and UM males, indicates that the lower rate for the UM class (illustrated in Table 5.34) is concentrated among the female speakers. In addition, the UM male speakers obtained a much lower rate for clause-initial vs clause-medial *like*. Schweinberger (2015:314) suggests that results such as these could be due to DM-*like* being stigmatised at certain social levels. The current results suggest a certain degree of stigmatisation for DM-*like* among UM female speakers and for clause-initial *like* among both the UM male and female speakers.

Labov (2001:264) found in his Philadelphia Neighbourhood Study that for the stable variable (-ing), his two middle-class groups indicated extreme gender differences where men used high frequencies of the variable in casual speech while women used low frequencies for both casual and formal speech. The current results in the WSAE sample for clause-initial *like* seem to follow such a pattern in the upper-middle class group and suggests that because of stigmatisation, the upper-middle class women avoid this form much more than the upper-middle class men. Labov (2001:266) provided evidence from numerous sociolinguistic studies indicating that men tend to

use more stigmatised forms than women, which explains the higher use of clause-initial *like* among the UM men vs the UM women. Further evidence suggesting that clause-initial *like* is stigmatised among the participants is the fact that the most represented group in the dataset, the MM female group, obtained a greater difference between clause-initial and clause-medial *like* compared to the other female groups: they are the most frequent users of clause-medial *like* but have a 0.7% "lag" for clause-initial usage in comparison to the other two groups. This result suggests that the MM female speakers are also trying to avoid this form more than clause-medial *like*, similar to the upper-middle class speakers and unlike the LM female speakers.

5.2.4 Statistical analysis of DM-*like* according to structural position

The results of the conditional inference tree for the structural distribution of DM-*like* (with the various social factors as independent variables) are illustrated in Figure 5.34. The results of this c-tree indicate that where clause-initial vs clause-medial *like* use is concerned, social class is the most significant factor. It indicates (see node 5[38]) that the UM and MM groups both use more clause-medial than clause-initial *like* (just above 60%; refer to Table 5.40 for the distributional results indicating this). The LM speakers in the older group obtained a similar result to the overall UM and MM groups and display more clause-medial than clause-initial *like* usage (see node 4). It is only the LM speakers from the younger group (node 3) that obtained a different pattern and show an equal amount of clause-initial vs clause-medial *like* (see Table 5.41 which indicates that it is mostly the LM females who contribute to this result). The c-tree indicates that gender is not a significant factor influencing the choice between clause-initial vs clause-medial *like* usage while social class and gender are.

Labov (2001:272) states that "The tendency to avoid stigmatized forms and prefer prestige forms is greatest for the women of the lower middle class and is often minimal for the lower class and upper middle class". The distributional results (Table 5.40) and the c-tree indicate that this is not the case for the younger LM class (see also the distributional results for the LM female speakers) since they are the group who do not avoid clause-initial *like*, which seems to be stigmatised by the upper middle-class speakers and to some degree the middle-class speakers. This result might suggest that clause-initial *like* is not stigmatised among the LM speakers. It is also true that the overall linguistic tendency in the current dataset as well as in many dialects (except NZE), (as discussed in Chapter 2, section 2.7.5), is to use clause-medial rather than clause-initial *like*, which also play a part in the results. It was only the LM female speakers who obtained a higher rate for clause-initial vs clause-medial *like*. However, as mentioned before, gender was not a significant

[38] Clause-initial *like* is the grey portion while clause-medial *like* is the dark portion.

factor in the statistical analysis. Even so, there are significant differences in the choice of clause-initial vs clause-medial *like* between the social classes.

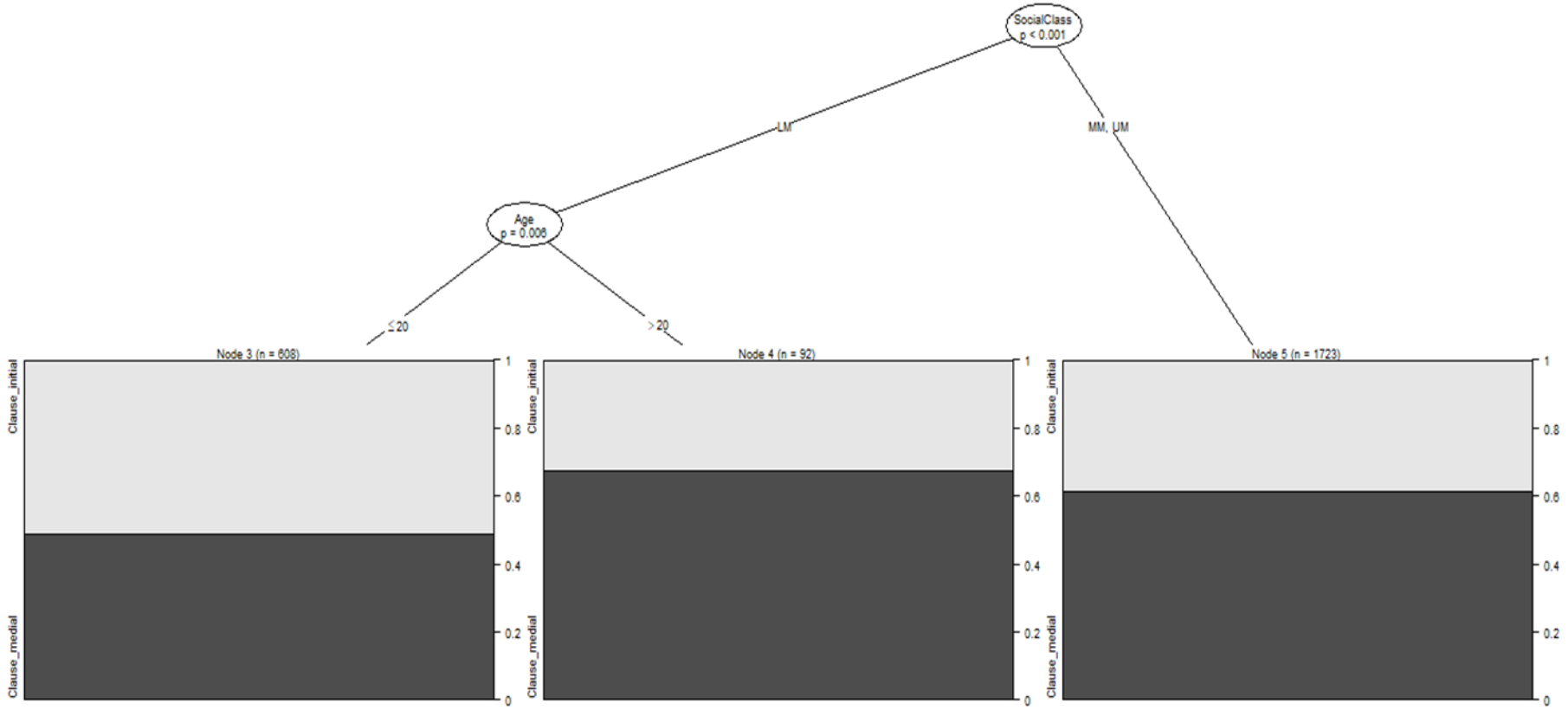


*Figure 5.35: Distribution of the structural position of DM-*like *correlated to social factors*

In conclusion, this chapter has investigated the social (external) and linguistic (internal) factors that influence the use of quotative be like and discourse-marker *like* in the WSAE sample. Distributional analyses for both quotative *be like* and discourse-marker *like* were conducted. C-tree analyses were included to provide a more detailed statistical component. In terms of internal factors, quotative *be like* was compared to other quotatives (such as *say*, *think* and the zero quotative) in the form of distributional proportions. It was also investigated vis-à-vis the other quotatives regarding the contexts of quote content, grammatical person and tense. In terms of social factors, *be like* vs other quotatives was examined to discover if age, social class and gender influence its use. The first c-tree illustrated the social factors influencing *be like* vs selected other quotatives directly included. The second c-tree illustrated how the internal factors influence its use. The third one illustrated how the social and internal factors combined influence its use vs selected quotatives directly included (while others were combined in the "other" category). A final c-tree illustrated how internal and external factors influence the use of *be like* when all other quotatives were grouped in the "other" category.

Discourse-marker *like* was similarly investigated by conducting a distributional analysis of external (gender, age social class) and internal factors (the structural positions of either clause-initial or clause-medial. C-tree analyses were also conducted for DM-*like*, first to discover the social factors that influence its use and secondly to discover how the external factors and internal factor of clause-position combined influence its use. The specific results and insights that were discussed in this chapter is summarised in the conclusion to this dissertation.

CHAPTER 6: CONCLUSION

This study aimed to examine how pragmatic *like* (quotative *be like* and DM-*like* patterns according to external and internal factors in White South African English in the Port Elizabeth region. The three research questions were the following:

1. How does pragmatic *like* (quotative *be like* and discourse-marker *like*) usage in White South African English (WSAE) in the Port Elizabeth region pattern according to external and internal/grammatical factors?

2. Once revealed, do the internal and external factors of pragmatic *like* in WSAE in Port Elizabeth English suggest anything about its current stage of development as compared to international English varieties?

3. Are there any similarities or differences regarding the social constraints on quotative *be like* and discourse-marker *like*?

First, a summary of the main findings is provided in answer to research question 1. This is followed by the implications of these findings with regards to the current stage of diffusion and grammaticalisation of pragmatic *like* in WSAE as compared to other English varieties, in answer to question 2. Finally, a comparison of the results of the social factors for quotative *be like* and discourse-marker *like* is provided.

6.1 Quotative *Be Like*

When quotative *be like* was investigated with regards to how it patterns as compared with the other quotatives, it was found that *be like* was the most frequent quotative in the sample, followed by *say* and the zero quotative. Compared with Horn's study (2022:69) of WSAE in Cape Town, the two studies were similar regarding the stated quotative pattern, however, in her data, *be like* was more frequent than in the current study. This indicates that WSAE speakers from various regions in S.A. favour similar quotatives above others in their speech.

The results for the distribution of *be like* across the younger and older age groups indicated that younger speakers (52%) use it much more than older speakers (20%). This result suggests that *be like* is age-graded in WSAE, where its use decreases with age. With regards to the use of *be like* across gender, the results indicated that women use it more than *say* while men use *say* more than *be like*. Horn (2022:287) obtained a similar result, which suggests that in WSAE, female speakers lead in the use of the quotative, while men display a more conservative, older quotative pattern in which *say* is the main quotative (e.g. D'Arcy, 2012:350). In addition, *be like* is in the process of replacing *say* as the main quotative choice among younger female speakers. For the

factors of social class and gender, therefore, the results found that *be like* is the most frequent among young middle-class female speakers, indicating that they are the leaders in its spread in the community. In addition, all three of the female social class groups use it more than their male counterparts, which establishes the female lead in WSAE further. It was also found that the gender difference was the smallest between the male and female upper-middle class speakers, indicating that the upper-middle class female speakers use it less frequently than their MM and LM counterparts. In terms of the results for the internal factors influencing the use of *be like*, in the WSAE sample, *be like* mainly introduces speech rather than thought while *say* is almost categorically used with speech reporting. The distributional results for grammatical person indicated that *be like* is expanding into third-person contexts, but it is still used more with the first person as compared to the third person. Horn's (2022) results indicated that *be like* was used at almost equal rates between first- and third- person contexts, but that it slightly favoured the first person.

With regard to the result for tense, it was found that *be like* is used much more with the past tense (67%) than the present tense (20%) in the sample. In contrast to Horn (2022)'s study, it was rare in the historical present in the current data. For the investigation of the quote content in relation to social factors, the majority of the social groups (both genders and all social classes) use it more with speech rather than thought reporting. However, older men indicated a pattern in which they used it more with thought as compared to speech. With regards to grammatical person, the results indicated that *be like* is used more with the first person vs the third person among the younger women, but that the younger men used it with equal proportions with first and third person. This could indicate gender differences in WSAE in the way that men and women choose to report speech. While younger women prefer to report more on their own speech (first person), men do not show a preference towards either reporting their own or other people's speech.

For tense, the majority of the social groups prefer using *be like* with the past tense as compared to the present tense. However, both younger and older men tend to use it with the present tense more than the younger and older female speakers. When social class was considered, it was revealed that the middle-class speakers obtained the highest proportion of *be like* use with the past tense while the lower-middle class speakers obtained the highest proportion of *be like* with the historical present tense.

The c-tree results for the younger group revealed that quote content, grammatical person and tense were significant factors influencing their use of *be like*. It indicated that *be like* was the most frequent quotative in the context of thought reporting, indicating that it competes with the zero quotative and *think* in this context. Within speech reporting for the younger female group, grammatical person and tense were significant factors and indicated that younger female speakers use *be like* in the past tense much more frequently when reporting their own speech

(first person). The younger women used *be like* with the present tense at a higher frequency when reporting the speech of others (third person). Unlike the female group who used *be like* the most with direct speech, the younger male group used *say* the most within speech reporting. There were no other significant factors for the male group.

For the c-tree analysis of the older group, the factors of quote content and tense were significant and indicated that the older women used *be like* more with the past tense than the older men. Compared with the younger group, the older speakers used *think* more, indicating that their choice in quotatives is more varied since the younger group mostly used *be like*, *say* and the zero quotative and to a much lesser degree *think*. Thus, the older speakers also participate in these quotatives along with *think*.

To answer question 2 with regards to *be like*, the following observations were made based on the results. The fact that *be like* overtook *say* as the most frequent quotative in the WSAE sample indicates that it patterns similarly to English dialects such as AmE, CanE and AusE, for which previous studies found a similar result. This result for WSAE suggests that *be like* is at an advanced stage of diffusion in the WSAE, Port Elizabeth speech community since a high frequency of an innovative linguistic variant is indicative of further spread.

Within the older group, men and women use *say* more than *be like*, however, the older women used *be like* at a higher rate than the older men. The younger female group used *be like* the most while the younger male group used *say* more, but only by 1%. These results indicate that older women in WSAE tend to pattern more towards the trend of the younger speakers while the older men are more conservative in their use of *be like*. In addition, it was noted that middle-class females use *be like* the most and are thus the leaders of this innovative feature. On the other hand, the upper middle class used *say* the most, also patterning more towards a conservative usage and thus more along the lines of the older men in the sample. Because the upper middle class obtained a smaller gender difference compared with the other social classes, it was revealed that the UM women use *be like* the least compared to the women in the other social classes and thus also illustrating a more conservative pattern among these particular women. Overall, the female lead in the use of *be like* follows the trend in dialects such as CanE, NZE and certain regions of EngE.

The results for content of the quote indicated that *be like* is used more with speech rather than thought. Older studies of AmE showed that *be like* was mostly used with thought rather than speech, however, later studies showed that it is increasing significantly in the context of direct speech. Tagliamonte & D'Arcy (2004:507) observed that high frequencies of *be like* with direct speech is indicative of a later stage of diffusion since its main pragmatic function (internal monologue reporting) is being bleached, thus indicating increasing grammaticalisation. This

explains why *be like* is competing with *say*: *be like* has overtaken *say* in the context of direct-speech reporting (a context in which *say* formerly was the main quotative) – it entered the same grammatical context of *say* and especially younger speakers started to use it more frequently in this context.

It has been stated that where *be like* is used more with third-person vs first-person reporting, an advanced stage of diffusion and grammaticalisation is suggested for the specific community. The results for the WSAE sample suggest that it has extended to the third-person context. However, it is still used more with first rather than first person. Since Horn's (2022) study found an almost equal rate for *be like* with first- and third-person contexts, this could suggest that the quotative is at a more advanced stage of grammaticalisation in Capetonian WSAE speech as compared to Port Elizabethan WSAE.

Regarding tense, earlier studies of AmE showed that *be like* was more frequent in the present tense (which included the historical present) as compared to the past tense. Thus, later studies (e.g. Gardner *et al*. 2020:299), observed that its more frequent use with the past tense over the years indicates further diffusion and grammaticalisation in a community. The current results found that *be like* is used most frequently with the past as compared to the present tense, indicating a later stage of grammaticalisation. Note though, that while the younger subjects used *be like* more in the past as compared to the present tense, the men use it more in the present tense as compared with the past tense, which might indicate that the younger men trail behind the women in this trend. Furthermore, the older men use it more in the present tense, which is indicative of an earlier stage of grammaticalisation.

Regarding social class and tense, the middle-class speakers obtained the highest proportion of *be like* use with the past tense, indicating that as the leaders, it is more grammatically developed in their speech as compared with the other social classes. The LM class displayed the highest proportion of *be like* use with the present tense, which might suggest that they also display an earlier stage of development. This slower rate of development which is suggested for the LM speakers possibly supports Cukor-Avila's (2012:622) observation that *be like* is diffusing hierarchically from the middle to the lower classes. *Be like* has also been found to be more frequent with the past rather than the present tense in BrE (Scottish English), whereas it has been found to be more frequent in the present tense in AusE, AmE, CanE and NZE.

6.2 Discourse-marker *Like*

The results for discourse-marker *like* indicated that it is used much more in the sample than non-discourse marker uses. The results for gender indicated that women use it more than men. This result is similar to what has been found for EngE and some studies of CanE. When the factor of

age was investigated, it was found that DM-*like* is used much more by the younger speakers as compared to the older speakers. This pattern indicates a similar trend found in all dialects for which age distribution was investigated.

The social class results revealed that DM-*like* is the most frequent in the middle class. However, when the interaction between gender and social class was investigated, it was revealed that it is the LM females who use it the most, which suggests a change from below.

For the structural analysis of DM-*like*, it was found that in the entire dataset, clause-medial *like* is more frequent than clause-initial *like*. This is a similar pattern to most dialects for which this structural pattern was investigated, except for Irish English, in which clause-final *like* was found to be the most frequent.

For the investigation of the social factors influencing structural choice, it was revealed that both younger and older speakers use clause-medial *like* the most, which indicates the main linguistic pattern for DM-*like* in general. Contrary to D'Arcy's results for clause-position across gender, the men in the WSAE sample do not use clause-medial *like* the most. The women use it the most in both contexts, solidifying the female lead for this feature. When gender and social class were investigated in combination, it was revealed that the upper-class women use clause-medial and clause-initial *like* the least, while the upper-class men use it more than them. This result suggests a certain degree of stigmatisation in the upper-middle class for DM-*like* in general.

The c-tree result for DM-*like* indicated that age was the most significant factor and the fact that the use of this feature decreases with age suggests a degree of age-grading. However, the results indicated that gender and social class were significant factors constraining its use which suggests that it is undergoing a linguistic change.

The c-tree results for the social factors influencing the structural position of DM-*like* revealed, surprisingly, that gender was not significant. This is likely due to the high frequency of DM-*like* among the upper-middle class male speakers regarding clause-medial *like*. Even so, social class and age were significant factors influencing its use.

For research question 3, regarding whether there are any similarities or differences in the social distribution of quotative *be like* and DM-*like*, the following was observed. The younger female speakers use both quotative *be like* and DM-*like* more than the younger men. However, when gender and social class are considered, the middle-class female speakers took the lead in the use of *be like*, while the lower-middle class females use DM-*like* the most. Regarding age, both variants are much less frequent among the older speakers, suggesting a certain degree of age-grading.

In conclusion, the research questions for this study were answered by the following results:

For the first hypothesis, younger speakers were found to be the most frequent users of both quotative *be like* and discourse-marker *like*. For the second hypothesis, the expectation was met that pragmatic *like* is used more by female speakers compared to male speakers. However, contrary to this expectation, discourse-marker *like* was more frequent among upper-middle class men compared to their female counterparts. For the third hypothesis, middle-class speakers were the most frequent users of both *be like* and discourse-marker *like*. However, middle-class female speakers used quotative *be like* the most while lower-class female speakers used DM-*like* the most.

The expectation of hypothesis 4 was that, for the influence of internal factors on *be like*, it would be used the most with first-person subjects, internal thought reporting and the present tense. The results indicated that it was used more with first-person subjects. However, contrary to expectation, it was used more with direct speech as well as the past tense. The expectation for the most frequent structural position of discourse-marker *like* was that it would be more frequent in clause-medial position. This expectation was met.

6.3 Suggestions for Further Study

Since studies for WSAE regarding pragmatic *like* are scarce and since the current dataset does not cover a wider age range, future studies could investigate the question of age-grading vs real change in progress for DM-*like* and quotative *be like*. For DM-*like*, this study only analysed its structural position regarding clause-initial vs clause-medial *like*. Previous studies have also analysed DM-*like* in terms of its structural position in relation to verb phrases, determiner phrases, adjective phrases and so forth (e.g. Lochner, 2019 for WSAE). A future study could contribute to this finer analysis of DM-*like*'s structural position. In addition, among the international studies for *be like* and DM-*like*, investigations of social class patterns have been scarce and this thus indicates a gap in the study field. Furthermore, continued research into the diffusion and grammaticalisation of *be like* and DM-*like* would be valuable. More recently, the notion that *be like* spreads through geographical diffusion and then advances in its grammatical development (grammaticalisation) the further it spreads in a community, has been questioned (e.g. Gardner et al. 2020:282–283, 311–312). It has been suggested that, rather than the process of grammaticalisation, where it extends to new grammatical contexts each time an older context has been saturated (fully developed), it rather displays a pattern conforming to The Constant Rate Effect, which is defined as a process in which a linguistic innovation rises simultaneously in various English varieties and is already in use in all grammatical contexts from the beginning, it is just that some communities use it less frequently than others. In addition, it would also display

the same grammatical (internal) constraints across all varieties of English (Gardner *et al*., 2020:312). Thus, this new direction offers a rich area of study for quotative *be like*.

REFERENCES

Álvarez-Mosquera, P. & Marín-Gutiérrez, A. 2019. A sociolinguistic approach to implicit language attitudes towards historically white English accents among young L1 South African indigenous language speakers*. International Journal of the Sociology of Language*, 260:131-153.

Andersen, G. 1996. They like wanna see like how we talk and all that: the use of like as a discourse-marker in London Teenage Speech. In: Ljung, M., ed. *Corpus-based studies in English*. Papers from the seventeenth International Conference on English Language Research on Computerized Corpora (ICAME 17). Amsterdam: International Conference on English Language Research on Computerized Corpora 41-52.

Andersen, G. 2000. The role of the pragmatic marker like in utterance interpretation. In: Andersen & Fretheim, T., eds. *Pragmatic markers and propositional attitude*. Amsterdam: John Benjamins Publishing Company. pp. 17-38.

Andersen, G. 2001. *Pragmatic markers and sociolinguistic variation: a relevance-theoretic approach to the language of adolescents*. Amsterdam: J. Benjamins. pp. 209-299.

Baird, S. 2001. How to be like a Kiwi: verbs of quotation in New Zealand English. *New Zealand English Journal*, 15:6-19.

Barbieri, F. 2005. Quotative use in American English: a corpus-based, cross-register comparison. *Journal of English Linguistics*, 33(3):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.

Barnard, A. 1992. *Hunters and herders of southern Africa: a comparative ethnography of the Khoisan peoples*. Cambridge: Cambridge University Press.

Beavon, K. 1970. Factors affecting the growth and form of Port Elizabeth, 1820-1963: a study in historical urban geography. In: Watts, L., ed. *Focus on Cities*. Proceedings of a Conference Organised by The Institute for Social Research, at The University of Natal, Durban, 8-12 July.

Beeching, K. 2016. *Pragmatic Markers in British English*. Cambridge: Cambridge University Press.

Bekker, I. 2012a. The story of South African English: a brief linguistic overview. *International Journal of Language*, *Translation and Intercultural Communication*, 1:139-150.

Bekker, I. 2012b. South African English as a late 19th-century extraterritorial variety. *English World-Wide*, 33(2):127-146.

Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. 2000. *Longman grammar of spoken and written English*. London: Longman.

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.

Boyd, W.C. 1963. Genetics and the human race: definition of race on the basis of gene frequencies supplements definition from morphological characters. *Science*, 140(3):1057-1064.

Branford, W. 1994. English in South Africa. In: de Klerk, V., ed. *Focus on South Africa*. Amsterdam: J. Benjamins. pp. 35-51.

Buchstaller, I. 2015. Exploring linguistic malleability across the life span: age-specific patterns in quotative use. *Language in Society*, 44(4):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.

Chambers, J.K. 2002. *Sociolinguistic theory: linguistic variation and its social significance*. 2nd ed. Oxford: Blackwell Pub.

Chevalier, A. 2016. Globalisation versus internal development: The reverse short front vowel shift in South African English. Cape Town: Cape Town University. (Thesis – PhD)

Chomsky, N. 2015*. Aspects of the theory of syntax*. Cambridge: MIT Press.

Coates, J. 2004. *Women, men and language: a sociolinguistic account of gender differences in language*. London: Routledge.

Coetzee-Van Rooy, S. & Van Rooy, B. 2005. South African English: labels, comprehensibility and status. *World Englishes*, 24(1):1-19.

Corrigan, K.P. 2015. I always think of people here, you know, saying ‘like’ after every sentence. The dynamics of discourse-pragmatic markers in Northern Irish English. In: Amador-Moreno, C., McCafferty, K. & Vaughan, E., eds*. Pragmatic markers in Irish English*. Amsterdam: J. Benjamins. pp.37-64.

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

D'Arcy, A. 2004. Contextualizing St. John's youth English within the Canadian quotative system. *Journal of English Linguistics*, 32(4):323-345.

D'Arcy, A. 2005. Like: syntax and development. Toronto: University of Toronto. (Thesis – PhD)

Da Silva, A.B. 2008. South African English: a sociolinguistic investigation of an emerging variety. Johannesburg: University of the Witwatersrand.

Dailey-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. 2012. The diachrony of quotation: evidence from New Zealand English. *Language Variation and Change*, 24(3):343-369.

D'Arcy, A. 2017. *Discourse-pragmatic variation in context: eight hundred years of like*. Amsterdam: J. Benjamins.

Davydova, J. 2020. The role of sociocognitive salience in the acquisition of structured variation and linguistic diffusion: evidence from quotative be like. *Language in Society*, 50(2):171-196.

De Saussure, F. & Harris, R. 2013. *Course in general linguistics*. Translated from the French by Roy Harris. London: Bloomsbury Academic.

Denis, D., Gardner, M.H., Brook, M. & Tagliamonte, S.A. 2019. Peaks and arrowheads of vernacular reorganization. Language Variation and Change, 31(1):43-67.

Department of Linguistics, U. 2024. Santa Barbara corpus of spoken American English. https://www.linguistics.ucsb.edu/research/santa-barbara-corpus. Date of access: 12 Nov. 2024.

Deuber, D., Hänsel, E.C. & Westphal, M. 2021. Quotative be like in Trinidadian English. *World Englishes*, 40(3):436-458.

Durham, M., Haddican, B., Zweig, E., Johnson, D.E., Baker, Z., Cockeram, D., Tyler, L. 2012. Constant linguistic effects in the diffusion of be like. *Journal of English Linguistics*, 40(4):316-337.

Eckert, P. 1989. *Jocks and burnouts: social categories and identity in the high school*. New York: Teachers College, Columbia University.

Eisikovits, E. 2011. *Language and gender: a reader*. 2nd ed. West Sussex: Wiley-Blackwell.

International Corpus of English. 1999, 2023. ICE-New Zealand. https://www.ice-corpora.uzh.ch/en/joinice/Teams/icenz.html. Date of access: 13 Nov. 2024.

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

Gabrys, M. 2017. Like as a discourse-marker in different varieties of English. Louvain: Université catholique de Louvain. (Thesis – M.A.)

Gardner, M.H. 2017. Grammatical variation and change in Industrial Cape Breton. Toronto: University of Toronto. (Thesis – PhD)

Gardner, M.H., Denis, D., Brook, M. & Tagliamonte, S.A. 2020. *Be like* and the Constant Rate Effect: from the bottom to the top of the S-curve. *English Language & Linguistics*, 25(2):281-324.

Garman, A. 2012. *Troubling White Englishness in South Africa*. Oxford: Inter-Disciplinary Press.

Gelderen, E.v. 2014. *A history of the English language*. Revised edition. Amsterdam: J. Benjamins.

Glaser, D.J. 2001. *Politics and society in South Africa*. London: Sage Publications, Limited.

Goschen, W.S. & Schumann, E.H. 2011. The physical oceanographic processes of Algoa Bay, with emphasis on the western coastal region. South African Environmental Observation Network (SAEON) and the Institute of Maritime Technology (IMT). IMT document number: PO106-10000-730002:1-84. https://www.researchgate.net/publication/275564341_The_physical_oceanographic_processes_of_Algoa_Bay_with_emphasis_on_the_western_coastal_region. Date of access: 13 Nov. 2024.

Gough, D.H. 1996. English in South Africa. In: Silva, P., Muller, C & Wright, M. & Dore, W. A *Dictionary of South African English on Historical Principles*. Oxford: Oxford University Press. Introduction.

Gradín, C. 2021. Occupational gender segregation in post-apartheid South Africa. *Feminist Economics*, 27(3):102-133.

Gries, S.T. 2021. *Statistics for linguistics with R: a practical introduction*. 3rd ed. Boston: de Gruyter.

Herat, M. 2018. You could so easily od like: clause final and other pragmatic functions of like in Liverpool English speech. *International Journal of Language Studies*, 12(2):87-112.

Herbert, R.K. & Bailey, R. 2002. *Language in South Africa*. Edinburgh: Cambridge University Press.

Hill, L. 2020. *Language and Social Space in Nelson Mandela Bay*. Handbook of the Changing World Language Map:1105-1130.

Holmes, J., & Wilson, N. 2017. *An Introduction to Sociolinguistics*. New York: Routledge.

Horn, E. 2022. And then I was like, no way!: A variationist study of be like in young Cape Town speech.

Horvath, B.M. & Horvath, R.J. 2001. A multilocality study of a sound change in progress: the case of /l/ vocalization in New Zealand and Australian English. Language variation and change, 13(1):37-57.

Inggs, J. 1986. Liverpool of the Cape: Port Elizabeth trade 1820-70. *South African Journal of Economic History*, 1(1):77-98.

JMK. 2007. Eastern Frontier, Cape of Good Hope, ca. 1820-1850. https://commons.m.wikimedia.org/wiki/File:Eastern_Frontier,_Cape_of_Good_Hope,_ca_1835.png. Date of access: 13 Nov. 2024.

Kahan, A. 1973. Notes on serfdom in Western and Eastern Europe. *The Journal of Economic History*, 33(1):86-99.

Kallen, J.L. & Kirk, J.M. 2008. ICE-Ireland: A User's Guide. Belfast: Queen's University.

Kruger, H. & Van Rooy, B. 2019. A multifactorial analysis of contact-induced change in speech reporting in written White South African English (WSAfE). *English Language & Linguistics*, 24(1):179-209.

Laberge, S. & Sankoff, G. 1979. Anything you can do. *Syntax and semantics*, 12:419-440.

Labov, W. 1963. The social motivation of a sound change*. Word*, 19(3):273-309.

Labov, W. 1966a (2006). The social stratification of English in New York city. Cambridge: Cambridge University Press.

Labov, W. 1966b. The linguistic variable as a structural unit. *Washington Linguistics Review*, (3): 4-22.

Labov, W. 1972. *Sociolinguistic patterns*. Pennsylvania: University of Pennsylvania press.

Labov, W. 1981. Field methods of the project on linguistic change and variation. Southwest Educational Development Laboratory. pp.29-52.

Labov, W. 2001. *Principles of linguistic change volume 2: social factors*. Massachusetts: Blackwell.

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

Labov, W. 2018. The role of the avant garde in linguistic diffusion. *Language Variation and Change*, 30(1):1-21.

Lambert, J. 2014. English-speaking South Africans: uncertain of their identity. In: Pretorius, F., ed. *A history of South Africa: from the distant past to the present day*. Pretoria: Protea Book House. pp. 588–609.

Lanham, L.W. 1965. South African English pronunciation 2065 AD. *Taalfasette*, 1(1):89-95.

Lanham, L.W. & Macdonald, C.A. 1979. The standard in South African English and its social history.

Lass, R. 2002. South African English. *Language in south Africa*, (104)126:104-126.

Levey, S. 2006. The sociolinguistic distribution of discourse-marker like in preadolescent speech. *Multilingua*, (25):413–441.

Levshina, N. 2020. Conditional inference trees and random forests. In: Paquot, M. & Gries, T., ed. *A practical handbook of corpus linguistics*: Springer. pp. 611-643.

Lochner, N.J. 2019. So we were just like, ok: the discourse markers like and just in the speech of young South Africans. Stellenbosch: Stellenbosch University. (Thesis – M.A.)

Macaulay, R. 2001. You're like why not?: the quotative expressions of Glasgow adolescents. *Journal of Sociolinguistics*, 5(1):3-21.

MacDonald, C.A. & Lanham, L.W. 1979. *The standard in South African English and its social history*. 2nd ed. Heidelberg: Julius Groos Verlag.

Google Maps. 2024. AfriGIS (pty) Ltd, Google. https://www.google.com. Date of access: 16 Nov. 2024.

Marx, K. & Engels, F. 1848. *The communist manifesto*. Routledge.

Meehan, T. 1991. It's like, what's happening in the evolution of like?: a theory of grammaticalization. In: Ichihashi, K. & Linn, M.S., eds. *Kansas working papers in Linguistics, vol. 16*. Kansas: Linguistics Graduate Student Association, University of Kansas.

Mesthrie, R. 2002. *Language in South Africa*. Cambridge: Cambridge University Press.

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., Chevalier, A. & Dunne, T. 2015. A regional and social dialectology of the BATH vowel in South African English. *Language variation and change*, 27(1):1-30.

Mesthrie, R., Swann, J., Deumert, A. & Leap, W. 2009. *Introducing Sociolinguistics*. Edinburgh, Edinburgh University Press.

Meyer, A. 2014. South Africa's primeval past. In: Pretorius, F., ed. *A history of South Africa: from the distant past to the present day*. Pretoria: Protea Book House. pp. 17-40.

Millar, R.M., & Trask, R.L. 2023. *Trask's Historical Linguistics*. London: Routledge.

Miller, J. & Weinert, R. 1995. The function of like in dialogue. *Journal of pragmatics, 23*(4):365-393.

Milroy, L. 1987*. Language and social networks*. Oxford: Wiley-Blackwell.

Morse, J.E. 1971. Roll of the British settlers in South Africa. https://www.amethyst.co.za/FamilyTree/1820SettlersMap.htm. Date of access: 16 Nov. 2024.

Murphy, B. 2015. A corpus-based investigation of pragmatic markers and sociolinguistic variation in Irish English. In: Amador-Moreno, C., McCafferty, K. & Vaughan, E., eds. *Pragmatic markers in Irish English*: John Benjamins. pp. 65-88.

Ogbonnaya, U.I. & Awuah, F.K. 2019. Quintile ranking of schools in South Africa and learners'achievement in probability. *Statistics Education Research Journal*, 18(1):106-119.

Parkington, J. & Hall, S. 2010. The appearance of food production in Southern Africa 1,000 to 2,000 years ago. In: Hamilton, C., Mbenga, B. & Ross, R. eds. The Cambridge history of South Africa: Volume 1, from early times to 1885. Cambridge: Cambridge University Press. pp. 63-111.

Parliamentary Group, 2020. https://pmg.org.za/committee-meeting/30934/. Date of access: 15 Nov. 2024

Pedersen, M. A. H. 2020. *Being and belonging among White English-speaking South Africans*. Cape Town: University of Cape Town. (Thesis – PhD)

Pope. 2007. Forty years of language change on Martha's Vineyard. *First Language*, 83(3), 615-627.

Rickford, J.R. 1975. Carrying the new wave into syntax: the case of black English bín. In: Fasold, R.W. & Shuy, eds. *Analyzing variation in language*. Papers from the second Colloquium on New Ways of Analyzing Variation, pages 162-183.

Rickford, J.R., Wasow, T., Zwicky, A. & Buchstaller, I. 2007. Intensive and quotative all: Something old, something new. *American Speech*, 82(1):3-31.

Roberge, P.T. 2002. Afrikaans: considering origins. In: Mesthrie, R., ed. *Language in South Africa*. Cambridge: Cambridge University Press. pp. 79-103.

Rodríguez Louro, C., Richard, S.L.R. & Bharadwaj, S. 2020. Another story: *be like* across discourse types. *English World-Wide*, 41(3):325-351.

Rodriguez-Louro, C. 2013. Quotatives down under: be like in cross-generational Australian English speech. *English World-Wide*: A Journal of Varieties of English, 34(1):48-76.

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(3):227-279.

RoomsForAfrica. 2024. Suburbs of Port Elizabeth. https://www.roomsforafrica.com/dest/south-africa/eastern-cape/port-elizabeth.jsp?tab=3. Date of access: 14 Nov. 2024.

Ross, R. 2010. Khoesan and immigrants: the emergence of colonial society in the Cape, 1500–1800. In: Hamilton, C., Mbenga, B. & Ross, R., eds. *The Cambridge history of South Africa: Volume 1, from early times to 1885*. Cambridge: Cambridge University Press. pp. 168–210.

Ross, R. 2010. From slave economy to settler capitalism: the cape colony and its extensions, 1800–1854. In: Hamilton, C., Mbenga, B. & Ross, R., eds. *The Cambridge history of South Africa: Volume 1, from early times to 1885*. Cambridge: Cambridge University Press. pp.253-318.

Ross, R. 2017. *Settler colonialism in South Africa*, 1652-1899. New York: Routledge.

Rossouw, R. & Van Rooy, B. 2012. Diachronic changes in modality in South African English. *English World-Wide*, 33(1):1-26.

Salisbury, T. 2003. Discourses of whiteness informing the identity of white English-speaking South Africans. University of Cape Town. (Thesis – PhD)

Sankoff, G. 1973. Above and beyond phonology in variable rules. In: Bailey, C.J.N. & Shuy, R.W., eds. *New ways of analyzing variation in English*. Washington: Georgetown University Press. pp.44-62.

Sankoff, G. & Brown, P. 1976. The origins of syntax in discourse: A case study of Tok Pisin relatives. *Language*, 52(3):631-666.

Sankoff, G., Thibault, P., Nagy, N., Blondeau, H., Fonollosa, M.-O. & Gagnon, L. 1997. Variation in the use of discourse markers in a language contact situation. *Language variation and change*, 9(2):191-217.

Saville-Troike, M. 2003. The ethnography of communication: an introduction. Hoboken: John Wiley & Sons, Inc.

Schilling, N. 2013. *Sociolinguistic Fieldwork*. New York: Cambridge University Press.

Schlemmer, L. 1985. *South African politics and the English speakers*. Durban: University of Natal.

Schourup, L.C. 1982. *Common discourse particles in English conversation*. New York: Routledge.

Schweinberger, M. 2013. Queensland: The University of Queensland. A sociolinguistic analysis of discourse-marker like in Northern Ireland. (Thesis – PhD)

Schweinberger, M. 2015. *The discourse-marker like: A corpus-based analysis of selected varieties of English*. Hamburg: Carl von Ossietzky.

Seekings, J. 2003. Social stratification and inequality in South Africa at the end of apartheid. CSSR Working Paper No. 31. Cape Town: Centre for Social Science Research.

Silva, P. & Dore, W. 1996. A dictionary of South African English on historical principles. Oxford: Oxford University Press.

Singler, J.V. 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(3):257-278.

Stenström, A.-B., Andersen, G. & Hasund, I.K. 2002. Trends in teenage talk: corpus compilation, analysis, and findings. Amsterdam: J. Benjamins.

Stewart, P. & Zaaiman, J. 2020. *Sociology: A comprehensive South African introduction*. Cape Town: Juta (Pty) Limited.

Stone, G. L. 2002. The lexicon and sociolinguistic codes of the working-class Afrikaans-speaking Cape Peninsula coloured community. In: Mesthrie, R., ed. *Language in South Africa*. Cambridge: Cambridge University Press. pp. 381-397.

Tagliamonte, S.A. 2005. Appendix B: Interview Schedule. (Online source removed) http://www.cambridge.org/catalogue/catalogue.asp?isbn=9780521778183&ss=res

Tagliamonte, S. 2005. So who? Like how? Just what?: discourse markers in the conversations of young Canadians. *Journal of pragmatics*, 37(11):1896-1915.

Tagliamonte, S.A. 2006. *Analyzing sociolinguistic variation*. Cambridge: Cambridge University Press.

Tagliamonte, S.A. 2011. *Variationist sociolinguistics: Change, observation, interpretation*. Oxford: John Wiley & Sons.

Tagliamonte, S. & Hudson, R. 1999. Be like *et al.* beyond America: 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.A. & D'Arcy, A. 2007. Frequency and variation in the community grammar: tracking a new change through the generations. *Language Variation and Change*, 19(2):199-217.

Tagliamonte, S.A. & Baayen, R.H. 2012. Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language variation and change*, 24(2):135-178.

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

Tannen, D. 1986. Introducing constructed dialogue in Greek and American conversational and literary narrative. *Direct and indirect speech*, 31:311-332.

Terblanche, H. 1993. Port Elizabeth, 'n Lojale Britse stad, 1902-1937. *Historia*, 38(2):100-111.

Theal, G.M. 1910. History and ethnography of Africa South of the Zambesi. Vol. III: the Cape Colony to 1795, the Koranas, Bantu, and Portuguese in South Africa to 1800. London: George Allen & Unwin Ltd.

Thompson, L. 2014. *A history of South Africa*. Yale: Yale University Press.

Titlestad, P. 1996. English, the constitution and South Africa's language future. In: de Klerk, V., ed. *Focus on South Africa*, 1:63-73.

Toefy, T.L. 2014. Sociophonetics and class differentiation: A study of working-and middle-class English in Cape Town's coloured community. Cape Town: University of Cape Town (Thesis – PhD).

Trudgill, P. 2000. *Sociolinguistics: An introduction to language and society*. Penguin UK.

Trudgill, P. 2003. *A glossary of sociolinguistics*. Oxford: Oxford University Press.

Trudgill, P., Trudgill, S., Anderson, S. & Huddleston, R. 1974. *The social differentiation of English in Norwich*. Cambridge: Cambridge University Press.

Vandelanotte, L.D., K. 2009. The emergence and structure of be like and related quotatives: A constructional account. *Cognitive Linguistics*, 20(4):777-807.

Visagie, J. & Posel, D. 2013. A reconsideration of what and who is middle class in South Africa. *Development Southern Africa*, 30(2):149-167.

Wallerstein, I. 1988. The bourgeoisie as concept and reality. *New Left Review*, 167(1):91-106.

Watermeyer, S. 1996. Afrikaans English. In: Kerk, V., ed. *Focus on South Africa*. Amsterdam: John Benjamins Publishing. pp.99-125

Weinreich, U., Labov, W. & Herzog, M.I. 1969. Empirical foundations for a theory of language change. In: Lehmann, W.P., ed. *Directions for Historical Linguistics*. Austin: University of Texas Press. pp. 97-188.

Winter, J. 2002. Discourse quotatives in Australian English: adolescents performing voices. Australian Journal of Linguistics, 22(1):5-21.

Wolfram, W. 2007. *Ethnic varieties*. New York: Routledge.

APPENDICES

APPENDIX A: PARTICIPANT CONSENT FORM[39]

DATE:

ECLM Authorization

PARTICIPANT INFORMATION LEAFLET AND CONSENT FORM

TITLE OF THE RESEARCH PROJECT: The Culture of the English-speaking Port Elizabeth Community

REFERENCE NUMBER: NWU-00704-21-S7

PRINCIPAL INVESTIGATOR: Yolande Coetzee

ADDRESS: 43 Schilbach Street, Parys

CONTACT NUMBER: 0736447992

You are being invited to take part in a research project that forms part of my master's degree in Linguistics and Literary Theory. Please take some time to read the information presented here, which will explain the details of this project. Please ask the researcher any questions about any part of this project that you do not fully understand. It is very important that you are fully satisfied that you clearly understand what this research is about and how you could be involved. Also, your participation is **entirely voluntary** and you are free to decline to participate. If you say no, this will not affect you negatively in any way whatsoever. You are also free to withdraw from the study at any point, even if you do agree to take part. Prior to publication of the study's results (or the point that publication is in process), you may also withdraw the data you generate.

This study has been approved by the **Ethics Committee for Language Matters (ECLM) of the Faculty of Humanities of the North-West University** and will be conducted according to the ethical guidelines and principles of the international Singapore Statement on Research Integrity (2010) and the ethical guidelines of the National Health Research Ethics Council. It might be necessary for the research ethics committee members or relevant authorities to inspect the research records to make sure that we (the researchers) are conducting research in an ethical manner.

What is this research study all about?

- *This study will be conducted through Zoom and will involve casual conversation.*
- *Approximately 24 participants will be included in this study.*
- *The objective of this research is to find out more about the culture of the home language English-speaking residents of Port Elizabeth.*

Why have you been invited to participate?

- *You have been invited to participate because you are part of the home language English community of Port Elizabeth.*
- *You have also complied with the following inclusion criteria: You were born in P.E. and are a permanent resident (never moved); identify as white, are either male or female and are between the ages of 18–40.*

[39] Original details: (22945830) C:\Users\22945830\Dropbox\Ethics committee\MEETINGS 2020\Informed consent form.docm
31 January 2020

File reference: 9.1.5.1.2

- *You will be excluded if: you do not fall into any of the above-mentioned inclusion criteria.*

What will your responsibilities be?

- *You will be expected to participate in a casual discussion via Zoom.*

Will you benefit from taking part in this research?

- *The direct benefits for you as a participant will be monetary compensation for your time and internet data used.*
- *The indirect benefit will be contributing to the scientific study of South African culture.*

Are there risks involved in your taking part in this research and how will these be managed?

- *The risks in this study, and how these will be managed, are summarised in the table below:*

Probable/possible risks/discomforts	***Strategies to minimize risk/discomfort***
None	None

- *However, we do believe that the benefits to you and to science (as noted in the previous section) outweigh the risk. If you disagree, then please feel free not to participate in this study. We will respect your decision.*

- *Should we learn, in the course of the research, that someone is harming you, or that you are intending to harm someone, then we must tell someone who can help you/warn the person you are intending to harm.*

Who will have access to the data?

- *Anonymity (that is, in no way will your results be linked to your identity) will be achieved by removing names from the transcripts. Confidentiality (that is, I/we assure you that we will protect the information we have about you) will be ensured by the principal investigator. Reporting of findings will be anonymous.*
- *Only the researchers will have access to the data. Data will be kept safe and secure by locking hard copies in locked cupboards in the researcher's office and for electronic data it will be password protected.*
- *Audio-recorded data will be sent to a transcriber who will sign a confidentiality clause (i.e., she will not be allowed to talk to anyone about any aspect of the data). As soon as data has been transcribed it will be deleted from the recorders. The transcripts will be stored on a password-protected computer. All co-coders will sign confidentiality clauses.*
- *Data will be stored permanently for the purpose of future research.*

What will happen to the data?
The data from this study will be reported in the following ways: In the master's thesis of the principal investigator and in possible articles for academic journals. In all of this reporting, you will not be personally identified. This means that the reporting will not include your name or details that will help others to know that you participated (e.g., your address or the name of your school).

Will you be paid/compensated to take part in this study and are there any costs involved?

Yes, you will be compensated to take part in the study with the amount of R175. No travelling is necessary to participate in the study, there will thus be no costs involved.

How will you know about the findings?
- The general findings of the research will be shared with you by the principal investigator.
- If you would like feedback on your personal results, then you can contact the principal researcher.

Is there anything else that you should know or do?
- You can contact the principal researcher at 0736447992 or yc.caly@gmail.com if you have any further queries or encounter any problems. In case the above details change, please contact the following person who knows me well and who does not live with me and who will help you to contact me: Celia Cloete, cell: 0711143911; email: celiacloete1@gmail.com.
- You can contact the chair of the Ethics Committee for Language Matters (Prof S Coetzee-Van Rooy) at 016 910 3442 or susan.coetzeevanrooy@nwu.ac.za if you have any concerns or complaints that have not been adequately addressed by the researcher.
- You will receive a copy of this information and consent form for your own records.

Declaration by participant

By signing below, I …….…… agree to take part in a research study entitled: "The culture of the English-speaking Port Elizabeth community"

I declare that:

- I have read and understood this information and consent form and it is written in a language with which I am fluent and comfortable.
- I have had a chance to ask questions to both the person obtaining consent, as well as the researcher (if this is a different person), and all my questions have been adequately answered.
- I understand that taking part in this study is **voluntary**.
- I understand that the video and audio of the Zoom conversation will be recorded and I give permission for the recording. I also understand that the recording will be deleted after the researcher has transcribed the conversation and that my identity will be removed from the transcription to ensure anonymity.
- I understand that what I contribute (what I report/say/write/draw/produce visually) could be reproduced publicly and/or quoted, but without reference to my personal identity.
- I may choose to leave the study at any time and will not be penalised or prejudiced in any way.
- I may be asked to leave the study before it has finished, if the researcher feels it is in my best interest to do so.

Signed at (*place*) .....…..…….. on (*date*) ……..…….. 20....

....
Signature of participant **Signature of witness**

- You may contact me again ☐ **Yes** ☐ **No**
- I would like a summary of the findings of this research ☐ **Yes** ☐ **No**

Declaration by person obtaining consent

I, Yolande Coetzee declare that:

- I explained the information in this document to
- I encouraged him/her to ask questions and took adequate time to answer them.
- I am satisfied that he/she adequately understands all aspects of the research, as discussed above
- I did not use an interpreter.

Signed at Potchefstroom on:

…… ……
Signature of person obtaining consent **Signature of witness**

Declaration by researcher

I, Yolande Coetzee declare that:

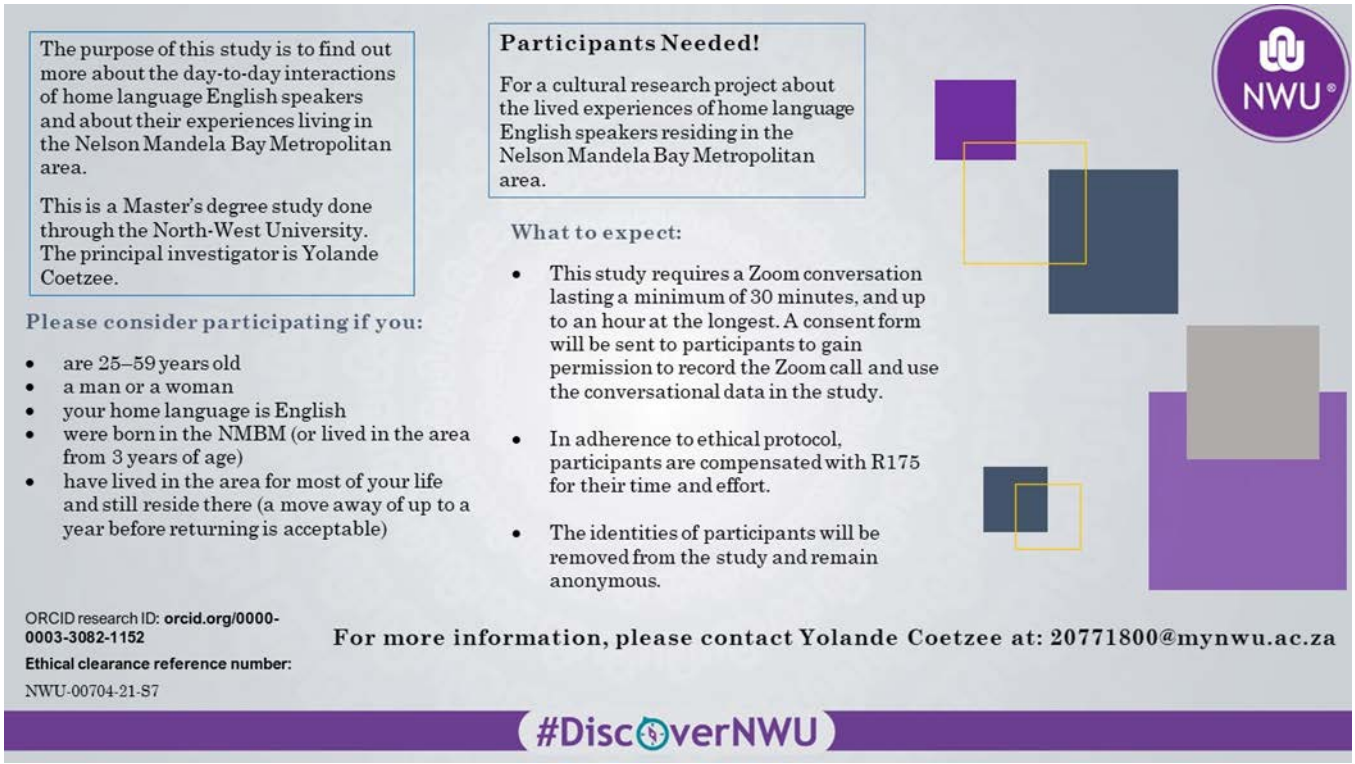
- I explained the information in this document to
- I encouraged him/her to ask questions and took adequate time to answer them.
- I am satisfied that he/she adequately understands all aspects of the research, as discussed above
- I did not use an interpreter.

Signed at Potchefstroom on:

Yolande Coetzee .... **Signature of researcher**

Hercules Crous..... ..... **Signature of witness**

APPENDIX B**: PARTICIPANT RECRUITMENT FLYER**



APPENDIX C: **SOCIAL CLASS CALCULATION**

ID	Pseudonym	Sex	Age	HoH_Occ	Edu	School_type	House_type	No._of_rooms_index_score	House_ownership	Total	
$F	Kiana van der Hoven	F	20	3	4	2	3	2	1	15	LM
$H	Harriet Henderson	F	19	1	4	4	1	1	1	12	LM
$J	Simon Oliver	M	33	3	4	3	2	2	1	15	LM
$K	Greg Harris	M	40	3	4	1	3	3	1	15	LM
$R	Owen Voigt	M	19	1	4	5	1	1	1	13	LM
$E	Michelle Smit	F	20	3	4	1	4	4	2	18	M
$G	Clara Mackle	F	20	3	5	3	3	2	2	18	M
$I	Jeremy Rork	M	18	3	5	2	3	2	1	16	M
$L	Melissa Candor	F	37	3	5	2	3	3	1	17	M
$M	Janet Wallis	F	19	3	4	2	4	4	2	19	M
$N	Katie Paton	F	19	3	5	3	4	2	1	18	M
$O	Natalie Miles	F	19	3	5	3	4	2	1	18	M
$S	Ben G	M	48	3	4	3	3	2	2	17	M
$A	Russel Taylor	M	20	4	5	3	4	3	2	21	UM
$B	Daniel Crouse	M	19	3	5	3	4	4	2	21	UM
$C	Gabrielle Harrington	F	18	4	4	3	4	4	2	21	UM
$D	Alan Harrington	M	20	4	5	5	4	4	2	24	UM
$P	Candace Irving	F	19	4	4	4	4	3	1	20	UM
$Q	Lily Fredericks	F	40	4	5	2	4	3	2	20	UM
										Total	
				5	5	5	4	4	2	25	

APPENDIX D**: LABOVIAN INTERVIEW SCHEDULE**

Introduction

Greeting

Demographics

And your address is?

Were you born in P.E.?

Do you live in a house or an apartment/flat?

What kind of rooms do you have (how many bedrooms, is there a study, swimming pool etc)?

Do you/your parents own or rent the house/apartment?

How long have you lived at that address?

Where were your parents born and raised?

Your grandparents?

Are you working now? Where?

Do you like what you do?

Is there any kind of profession that you would like to do or working towards?

How many years of school did you have a chance to finish?

What was the first job you had when you left school?

What kind of work does your mother do?

What kind of work does your father do?

Do you speak any languages besides English?

Do your parents and grandparents speak any other languages?

Neighbourhood

What kind of people live on your street?

What made your parents [you] move here?

Because of work?

Is it far form where your father [mother] works?

Because of community roots?

How has your neighbourhood changed in your lifetime?

Do you feel that your neighbourhood is as safe as it was when you were growing up? Why or why not?

Is this the kind of neighbourhood where people talk to each other?

Do you know any of your neighbours? [What are they like?]

Is there anyone around here you know well enough, just to walk in?

Who would invite you in for coffee, just talk?

Do people from around here drop by to visit?

Is there any neighbourhood place where people get together?

Are there any shops/cafés etc close to you neighbourhood?

Any hangout places close by?

Are there people you'd like to spend more time with but can't?

Why don't you see them so much anymore?

What do you like best about your neighbourhood? What are the things that make you feel good/bad about your neighbourhood?

Community Events

Did anything really big ever happen around here that you remember?

PARENTS AND FAMILY

What kind of kid were you when you were growing up?

Did you ever get into trouble?

What kinds of things did you do to get into trouble?

How were you punished? By who?

Did you have any rules about when you had to be in at night?

What happened when you stayed out late?

Did you ever get caught sneaking out? Why?

Were you ever grounded?

Did you ever get blamed for something you didn't do?

Were your parents really strict?

What sort of person is your father?

What is your mother like?

Were you close to your siblings growing up or did you fight a lot?

How about now?

**Have you ever been really embarrassed by something your parents/siblings said or did?

What happened?

How did you react?

What can you say from hindsight?

What did your parents want you to do for a living?

What do you do?

What would you like to do?

**Do you ever play tricks on your sister/brother?

What's the worst thing you ever did? Funniest thing you ever did?

Did you ever go on vacations as a family?

Where would you go? How would you get there?

Did you get along while you were on vacation?

SOCIAL PRACTICES

Are there people you spend a lot of time with outside your family?

Do they live nearby? Whereabouts?

What do you do like to do for fun together?

FAMILY MEALS/CRAFTS

Do you like cooking? Baking? What kinds of things do you like to bake/cook?

Do you like doing any crafts; scrapbooking, crochet, knitting etc.

SCHOOL DAYS

Where did you go to primary and high school?

How far is it from your house?

How did you get to school?

What subjects did you take?

What was your favourite subject? Least favourite subject? Why?

Do you remember anything interesting or funny that happened at your school or in class?

Did you have any teachers that are really tough?

Did you ever have a teacher that was a really weird? A real creep? Crazy?

Did you ever have a teacher that was really fair? That you liked?

What makes teachers really mad?

What would a teacher yell at a kid for?

What was the worst thing you ever saw a teacher do to a kid?

Or a kid do to a teacher?

Have you ever pulled an 'all-nighter' for school/studies?

Games

Going back to the time when you were a kid, ten, twelve years old, what were some of the games you used to play?

Did boys and girls play different types of games?

Games at school/home

Teen Life

Do you like to chat with you friends on WhatsAap or any other messaging app?

**Did you ever have an argument with someone on the app? What happened?

Do you like to play video games?

What kind of music do you like? What kind of music are you listening to these days?

Have you ever been to a concert to see one of your favourite groups?

How was it?

What's the best concert you were ever at?

Who's your favourite artist? How come?

What is your favourite movie?

What was it about?

Did you go to your matric farewell? What was it like?

Sport

Do you like any kind of sport?

Did you play any sports in school?

Any now?

Do you like watching sports?

Did any fights ever break out over favourite teams?

HOBBIES/CAMPING

Do you have any special interests?

Anything you are passionate about?

Do you have any hobbies? What?

How did you get into that?

Do you and your family like to go camping?

Any interesting camping stories you can tell?

Did you ever go on a school trip? Where did you go? Did anything interesting ever happen?

BIRTHDAYS

When is your birthday?

**What is the best birthday party you ever had?

What is the best birthday party you ever went to?

Has anyone ever held a surprise birthday party for you?

Who did it?

Were you really surprised or did you pretend?

Has anyone ever forgotten your birthday?

Holidays

Do you do anything for Halloween?

Do you dress up for it / go to dress up parties?

What was your favourite Halloween costume so far?

Does your family celebrate Christmas?

Has anything funny/interesting ever happened at a family gathering/Christmas get together?

What was Christmas like in your family?

Who picked out the tree? Who decorated it?

Did you write to Santa Claus? Did he bring what you wanted?

When did you open your presents? How did you open your presents? Did you have to go in order?

At whose house did you celebrate?

What did you eat at Christmas?

What's your favourite memory of Christmas?

What was your best Christmas?

What do you usually do on New Year's Eve?

Hangouts

Do you ever go to clubs?

What kind of music do they play?

What is your favourite song/artist to dance to?

What do girls wear to clubs like that? Guys?

**Has anything interesting/funny happened at a club you were at? What happened?

FIGHTS/ARGUMENTS

Have you ever witnessed a fight?

Where was it?

What was it about?

Do you ever have fights around here?

How do they start?

Did you ever get into a fight with a girl/guy?

Do you remember getting into an argument with someone?

Who was it with?

What was it about?

How did you resolve your differences? How did it all turn out?

TRAVEL

Have you had the opportunity to travel?

Where did you go? How long? Anything interesting happen?

Many people experience problems when they are at airports, has this ever happened to you?

Has a communication barrier created any funny moments in any of your travels?

Did you ever lose your luggage? Miss a plane? Get stranded?

Where would you like to go that you've never been?

Why?

**What's the funniest/scariest thing that ever happened to you when you were travelling?

DATING PRACTICES

Do you remember going out on dates? Do you go out on dates? Have you ever gone on a date with a boy/girl?

When did you (or your friends) start dating?

What do you like to do on a date?

How did you meet your girlfriend/boyfriend.

How did he ask you out?

What is the best date you've had so far? What happened?

What did you do for Valentine's Day? / What was your fav Valentine's date?

Misc

Have you ever met/seen someone famous?

Who was it? Where was it?

Did you talk to them?

Is there anyone famous you have imagined a meeting with?

Who and how do you imagine it?

Do you play any musical instruments?

If yes, which ones? For how long?

What made you start? e.g. school, parents

If no, is there an instrument you would like to learn to play? Why?

Do you or your family have any pets?

What are they like?

Did you ever have a pet run away? What happened?

Do you have any funny stories about your pets?

would you like to do that you've never done?

Do you have a dream? What is it?

UNCOMMON EXPERIENCES

Was there ever anything that happened when you were growing up that you couldn't explain?

Were there any spooky places you wouldn't go at night?

Does it bother you when people talk about ghosts?

**Do you know anyone who has seen a ghost?

Dreams

Do you dream a lot?

Do you remember any dreams?

Was there ever a dream like that, where you just didn't want to wake up?

Did you ever have a dream that really scared you?

What happened?

Do you think dreams can mean anything?

**Did you ever have a dream that you thought meant something?

COMMON SENSE

People talk a lot about common sense. What is common sense in your opinion?

Did you ever meet anybody that had a lot of common sense?

Did you ever meet anybody that had no common sense?

Do you think you get more common sense as you get older?

LANGUAGE

Have noticed any interesting things about the way people speak English in your area?

A lot of people think that English has changed a lot/is changing a lot, do you think so?

Have you noticed any changes in the way people talk and sound around here?

Can you tell by the way people talk around here that they come from here?

Do people in this neighbourhood sound different?

How about the difference between old and young speakers? Do you sound the same as your parents? Do your parents sound the same as you?

Do you speak the same way as your friends? What kinds of differences to you notice?

Do you think that you try to change how you sound when you are in certain environments? Which ones? Why?

What do you think about the way that young people sound these days? (E.g. how they speak?)

Conclusion

This study was actually about the word like as in “He was like, what?!

And how English South Africans in P.E. use it. The initial information focuses on the social aspect of the study rather than language because this information would influence a person to change how they speak and the aim is to have speech data that is as natural as possible.

You just need to give verbal consent and answer if you allow the speech data to be used in the study now that you know this information.

Thank you for participating!

APPENDIX E: TRANSCRIPTION CODES

The labels that were used for the interviewees are: $A for the first interviewee, $B for the second interviewee and so on. The label $Z was used to identify the interviewer in all the transcriptions. The transcripts were coded using the coding symbols of Nelson (2002). These are listed below.

- <I> This symbol was used at the beginning of each interview to indicate the start of the interview text. Where interview transcripts were divided into two parts because the video recording was in two parts, the symbol only appeared at the start of the part one transcript.
- </I> This symbol was used at the end of a transcription to indicate that the text for a specific interview has ended. In cases where the transcriptions were divided into two parts, this symbol only appeared at the end of part 2 to indicate the conclusion of the interview.
- <#> A text unit marker appears at the start of each speaker turn in order to give the transcription a uniform structure. This symbol is used in the ICE corpora to eventually number the text units automatically for reference (Nelson, 2002:4) but was included in the transcripts for the current study simply to indicate text units.
- <X> and </X> These symbols were used at the start and end of a text unit considered to be extra-corpus text, which instructs a corpus-parsing and tagger programme to ignore text between the symbols. In the transcriptions for this study, these symbols were used to mark the interviewer speech.
- <,> and <,,> These symbols indicate short and long pauses in the interviewees' speech respectively.
- <O> and </O> Anthropophonics – these symbols were used to indicate any non-verbal utterances or sounds such as: <O>cough</O> <O>sneeze</O> <O>laugh</O>. Where relevant, body language such as facial expressions were also indicated between these symbols, for instance:
- <O>nervous facial expression and verbal sound</O>
- <unclear> This symbol was used to indicated unclear speech, where either the Zoom recording lagged or where it was difficult to recognise a specific word in the speech.
- <?> </?> These symbols were used where a good guess could be made at a potentially unclear word, for instance: "This is a typical example of <?>radiation</?> filtering." (Nelson, 2002:8).
- <&> </&> These symbols were used to add editorial comments, such as to provide contextual notes, for example: <&>soft</&> "no"
- <.> </.> These symbols were used to indicate incomplete words uttered by the interviewee, for instance: "Oh <.>pers</.> Like my personal interests?"
- <s n="1"> This code was used to number each interviewee's speaker turn. This code is used in the BNC corpus transcripts (1991–1994) and was obtained from their website

(https://ota.bodleian.ox.ac.uk/repository/xmlui/handle/20.500.12024/2553) and was adopted here for easy reference. For example, wherever the speech of a participant is quoted in this paper, the following code is used to indicate which speaker is being quoted, using the speaker letter which was assigned in the order that interviews were conducted (speaker A [$A] was the first interviewee). In addition, the paragraph or sentence number is included, which indicates the paragraph or sentence number where the quoted section can be found in the transcript.

Like a detached house. (<$A><s n=127>)