“All the people who live in Auckland”: A study of subject and non-subject relative clauses in Auckland English

By

Alexandra Margaret Birchfield

A thesis
submitted to the Victoria University of Wellington
in fulfilment of the requirements for the degree of
Masters of Arts
in Linguistics

Victoria University of Wellington 2018

Abstract

This thesis is a study of the variation in relative marker choice by speakers of Auckland English. The data used in this study was collected as part of "Breaking Babel – Rethinking Language Change in a super-diverse city" (Meyerhoff et al. 2015). The thesis investigates the syntactic and social conditioning on the variation of the complementisers used to introduce relative clauses in the speech of a diverse group of Aucklanders. As a super-diverse city with a rapidly changing sociolinguistic profile, Auckland offers a rich source of data. This research explores how syntactic variation marks speakers of "Auckland English".

This work addresses several key research questions which centre on whether there is evidence of language change for this variable, and if so where has the change been initiated and by whom is it lead. Further, how does the variation in Auckland English compare other communities studied, both in terms of studies of relative clause variation and variation in super-diverse cities.

These questions derive from an exploration of the history of relative clauses in English. In chapter 2, I review how the current variable system of relative markers developed and how they have been treated both by syntacticians and variationists in previous literature. The purpose of a (restrictive) relative clause is to delimit the denotational reference of an antecedent head nominal that it post-modifies (Huddleston & Pullum 2002: 1034–1035). As such, variation in the choice of complementiser that introduces relative clauses tells us a great deal about how speakers specify information. The variability of relative markers is highly circumscribed (Ball 1996, Levey 2014). Nevertheless, the syntactic and social factors governing their distribution vary between speech communities and can offer insight into the linguistic profiles of these communities (Tagliamonte et al 2005, D'Arcy and Tagliamonte 2010).

This study analyses over 2000 tokens of relative clauses, coded for syntactic environment and speaker age, sex and community. Three communities, chosen for their differing demographic profiles, are sampled across Auckland. Significant predictors of relative marker choice are then compared to other studies of relative clause variation. This thesis then explores (i) which factors are universal or common predictors of relativiser choice, (ii) which factors index Auckland English and (iii) which are markers of specific communities within Auckland.

Previous studies of superdiverse cities (cf. Cheshire et al. 2015) have shown that the input of many diverse language varieties into a community can lead to large scale innovation and change. I explore the variation in relative markers in Auckland English in this context. Little evidence of language change taking place is found in this study and in fact, social factors such as age-grading patterns may suggest stable variation. There is some evidence of levelling (Trudgill 2004) in the most diverse of the three communities surveyed. The thesis concludes with a discussion of the significance of these results, both to the study of relative clauses and linguistic variation in general.

Table of contents

Acknowledgements **iii**

List of Figures **iv**

List of Tables **v**

1. Introduction **1**

2. Background **6**

2.1 What is a relative clause? 6

2.2 In the beginning *that* was it 14

2.3 Variationist tradition 21

2.3.1 Syntactic conditioning 23

2.3.2 Social conditioning 26

2.4 Superdiversity and language change 29

3. Method **32**

3.1 Data Collection 32

3.2 Coding 37

3.3 Data Analysis 41

4. Results **43**

4.1 The Dataset 43

4.2 Complimentary distribution of relative clauses 46

4.3 Analysing the variation 49

4.3.1 Non-subject relatives 49

4.3.2 Subject relatives 52

4.4 Constraints on variation 54

4.4.1 Syntactic factors 54

4.4.2 Social factors 58

4.5 Interaction effects 67

4.6 Inter-community variation 72

4.7 Qualitative analysis 77

4.7.1 Ø subject relatives 77
4.7.2 Adverbial relative clauses 80
4.8 Summary of Results 83
5. Discussion 84
5.1 Overview 84
5.2 Review of syntactic constraints 86
5.3 Review of social constraints 89
5.4 Review of qualitative analysis 93
5.4.1 Ø subject relatives 93
5.4.2 Adverbial relative clauses 94
6. Conclusion 96
References 100
Appendices 106
Appendix A - Content of Participant Information Sheet 106
Appendix B - Content of Participant Consent Form 109
Appendix C - Multiple regression analysis of factors contributing to the choice of Ø in non-subject relatives (speaker as random effect). 111
Appendix D - Multiple regression analysis of factors contributing to the choice of *who* in subject relatives with human antecedents (speaker as random effect). 114

Acknowledgments

This research was funded by a Marsden grant awarded to “Breaking Babel – Rethinking Language Change in a super-diverse city” (Meyerhoff et al. 2017). It has also been funded by a Victoria University of Wellington MA by thesis scholarship. I am profoundly grateful for this generous financial support without which I would not have been able to pursue this course of study.

Being part of the Auckland Voices team has been a huge privilege. Special thanks to Miriam Meyerhoff, Helen Charters, Elaine Ballard, Catherine Watson and Richard Arnold for their guidance and support. To all the interviewers, research assistants, transcribers and other champions of the project, those I’ve met and those I haven’t, thank you for creating such a rich and exciting dataset. Thank you especially to my supervisor Helen Charters for her invaluable help with coding. Without you I’d probably still be trudging through ELAN files.

To my colleagues and friends at Victoria, thank you for inspiring conversation and much-needed distractions. You’re all brilliant!

My supervisor, Miriam Meyerhoff has been teaching me from my first syntax tutorial. I don’t know how to express my gratitude to someone whose support and inspiration has shaped my entire experience in linguistics. Thank you, thank you.

Finally, to my family, thank you for your love and understanding. Thank you Dad, for your unwavering enthusiasm and support and all the airport pick-ups. Mum, you’ve very much been on this journey with me, I could not have done it without you. And to my new family, Kalya, my ray of sunshine and Meowzebub, our little ray of pitch black, thank you, there’s no one else I’d rather have by my side.

List of figures

		Page
Figure 2.1	The formation of appositive relatives from main clauses, Emonds (1979: 213) - redacted	8
Figure 2.2	Tree diagram of the position of relative markers in deep structure	10
Figure 2.3	Demonstration of how *that* subject relatives do not violate the anti-locality principle but Ø subject relatives do, Douglas (2017: 6) - redacted	19
Figure 3.1	Map of Auckland showing the three regions data was collected from	33
Figure 3.2	ELAN tier dependency structure	41
Figure 4.1	Percentage of *that, Ø* and *wh-* (*who* and *which*) in subject and non-subject relatives by community and gender.	47
Figure 4.2	Probability of the use of Ø in non-subject relative clauses according to the definiteness of the antecedent	55
Figure 4.3	Probability of *who* as a complementiser in subject relatives with human antecedent according to the definiteness of the antecedent	57
Figure 4.4	Probability of who as a complementiser in subject relatives (human antecedents) by community	59
Figure 4.5	Probability of Ø as a complementiser in non-subject relatives by community	60
Figure 4.6	Percentage of that, Ø and wh- (who and which) in subject and non-subject relatives grouped by community, gender and age (younger group is under 25 and older is over 40).	61
Figure 4.7	Percentage of that, Ø and wh- (who and which) in subject and non-subject relatives by age group.	63
Figure 4.8	Probability of Ø as a complementiser in non-subject relatives by three age groups	65
Figure 4.9	Probability of who as a complementiser in subject relatives with human antecedents by three age groups	66
Figure 4.10	Probability of *who* as a complementiser in subject relatives with human antecedents according to age group and gender.	70
Figure 4.11	Probability of Ø as a complementiser in non-subject relatives according to age group and gender	72

List of Tables

		Page
Table 3.1	Breakdown of corpus by community age and gender	34
Table 3.2	Participant pseudonyms, ethnicity and linguistic background	35
Table 4.1	Distribution of (non-adverbial) relative markers	43
Table 4.2	Distribution of relative clauses according to social factors	44
Table 4.3	Distribution of relative markers by community	45
Table 4.4	Distribution of relative markers according to their role in the relative clause	46
Table 4.5	Distribution of relative markers in subject relatives according to animacy	48
Table 4.6	Distribution of relative markers in non-subject relatives according to animacy	48
Table 4.7a	Multiple regression analysis of factors contributing to the choice of *Ø* in non-subject relatives with speaker as random effect	50
Table 4.7b	Multiple regression analysis of factors contributing to the choice of *Ø* in non-subject relatives without speaker as random effect	51
Table 4.8a	Multiple regression analysis of factors contributing to the choice of *who* in human subject relatives with speaker as random effect	52
Table 4.8b	Multiple regression analysis of factors contributing to the choice of *who* in human subject relatives without speaker as random effect	53
Table 4.9	Distribution of *Ø* and *that* in non-subject relatives when clause length and complexity are combined.	55
Table 4.10	Multiple regression analysis of factors contributing to the choice of *Ø* in non-subject relatives with community and age as an interaction group	68
Table 4.11	Multiple regression analysis of factors contributing to the choice of *who* in human subject relatives with community and age as an interaction group	69
Table 4.12	Multiple regression analysis of factors contributing to the choice of *Ø* in non-subject relatives with gender and age as an interaction group	71
Table 4.13	Three separate variable analyses of factors contributing to the choice of *Ø* in non-subject relatives.	74
Table 4.14	Three separate variable analyses of factors contributing to the choice of *who* in subject relatives with human antecedents.	76
Table 4.15	*Ø* subject relatives	79
Table 5.1	The distribution of relative markers across the dataset for four different studies	84

1. Introduction

This thesis is an investigation of the markers used to introduce restrictive relative clauses in spoken Auckland English. This takes the form of a variationist analysis of the three relativising strategies in English, exemplified in the examples below; the use of complementiser *that* (1), the use of a *Ø* or *null* complementiser (2), and the use of a *wh*-relative pronoun (3 and 4).

1. "there are also places in New Zealand [**that** I do feel like maybe have more sort of personality]" – Caitlyn – (Titirangi, younger)
2. "they stay out at Piha in a caravan [**Ø** they've got permanently out there]" – Charlotte – (Titirangi, older)
3. "now everyone in our class has their own device [**which** they use a lot for writing.]" – Bindi – (Titirangi, younger)
4. "people are like 'oh your internet friends [**who** you've never met]'" – Caitlyn – (Titirangi, younger)

I refer to these variants as relativisers (Sigley 1997) or relative markers (Tagliamonte et al. 2005). The variation in relative markers is of interest due to the diversity of the factors that influence this variation. Some aspects of the variation are highly grammatically circumscribed (Ball 1996, Levey 2014). For instance, when the relative marker takes the role of the subject in the relative clause (such constructions are referred to as subject relatives following eg. Levey (2014)) *Ø* is considered ungrammatical as a relative marker in many varieties of English. This is shown in example 5.

5. a. "So he was the guy [(**who**[1]/that/**Ø*) organised the mutiny on the bounty], and that's how the Pitcairn Islands started." – Chelsea – (Papatoetoe, younger)
   b. "oh your internet friends [(**who**/that/Ø) you've never met]" – Caitlyn - (Titirangi, younger)

[1] In examples taken from the corpus where multiple relative markers are shown for illustrative purposes, the relativiser in **bold** is the one the speaker actually produced.

There are also restrictions on the *wh-* relative pronouns in that they encode information about their referent and therefore must agree with the antecedent NP with which they are coreferential. Consider, for example, the infelicity of *who* as a relativiser in any of the examples in 1, 2, or 3 or the use of *which* in example 4 or 5. This would sound odd, if not actually ungrammatical to native speakers. Referring to an inanimate object such as "a caravan" as *who* would sound unusual, as would "the guy *which* organised the mutiny on the bounty". These agreement rules are not entirely categorical however, and in some cases, speakers have optionality, such as with non-human animate referents as in examples 6 and 7;

6. "And they had horses [**which** they bred and raced.]" – Dot – (Papatoetoe, older)
7. "I don't wanna see that dog [**who** lives in the house go away]" – Aarav – (Mt Roskill, younger)

In both examples 6 and 7 we have a non-human, animate antecedent but in 6 this is coreferential with *which* and in 7 with *who.* (Possibly this is connected to the level of sentience attributed to the animal in question by the speaker but such speculations are beyond the scope of this study – as indeed they are beyond most variationist studies).

However, the syntactic factors governing the variation in relativisers are not universal across varieties of English. The use of *Ø* as a subject relative is acceptable in some varieties (Tagliamonte et al. 2005). In constructions in which there is optional variation, the syntactic and social factors governing the distribution of the variants differ between speech communities and can offer insight into the linguistic profiles of these communities (D'Arcy and Tagliamonte 2010, Levey 2014, Tagliamonte et al. 2005).

This thesis, as with the majority of variationist studies of relative clause variation, focuses specifically on restrictive relative clauses (although this categorisation is not unproblematic as will be seen in chapter 2). In essence, a restrictive relative clause, as its name suggests, restricts the frame of reference of its antecedent to a specific referent (Huddleston & Pullum 2002: 1034–1035). As such, variation in the choice of complementiser that

introduces relative clauses tells us a great deal about how speakers and speech communities conventionalise the encoding of information structure.

This thesis begins with a summary of the treatment of relative clauses by syntacticians, beginning in the late 1970s/early 1980s (Emonds 1979, Jackendoff 1977, McCawley 1981), then moving through to an explanation of how relative clauses are generally conceptualised by contemporary grammarians (Carnie 2013, Haegeman 1994). An important part of this discussion is the question of restrictiveness and whether this concept is binary or not. This is of crucial importance to the study of relative marker variation because non-restrictive relatives are generally excluded from studies of relative marker variation (Ball 1996, D'Arcy and Tagliamonte 2010, Levey 2014, Tagliamonte et al. 2005). Canonically, non-restrictive relative clauses differ from restrictive relative clauses both in discourse function as well as in regards to which relativisers they admit. However, it does not appear that the distinction between restrictive and non-restrictive relative clauses is clear cut and unambiguous. I will defer further discussion of this issue to chapter 2.

From defining the context in which the variation will be studied, I move on to an exploration of how English developed its current system of relative markers and discuss the different distributions and restrictions on these. Chapter 2 concludes with a review of previous studies of relative marker variation in English and their findings. This provides context to this study and a basis for four key research questions. First, is there evidence of language change in the relative marker system of Auckland English? Second, if there is evidence of change, does this change emanate from Auckland's more diverse communities? If so, is it being led by younger speakers? Third, how do the grammatical constraints on the variation of relative markers in Auckland English compare to those found in other studies of more homogeneous communities? Finally, is there evidence of levelling, both in terms of one variant becoming more dominant and in terms of a loss of significant grammatical constraints on variation.

Chapter 3 describes the dataset and methodology of the study. The data used in this study was collected as part of "Breaking Babel – Rethinking Language Change in a super-diverse city" (Meyerhoff et al. 2015). This study analyses over 2000 tokens of relative clauses, coded for syntactic environment and speaker age, sex and community. Three communities,

chosen for their differing demographic profiles, are sampled across Auckland. As a super-diverse city (Blommaert 2013, Vertovec 2014) with a rapidly changing sociolinguistic profile (Cheshire et al. 2011), Auckland offers a rich source of data. The three communities sampled are as follows:

- Titirangi - monocultural and monolingual (predominantly English speaking Pākehā/European), and has been so for a long time;
- Papatoetoe - ethnically and linguistically diverse since at least the 1970s;
- Mt Roskill – formerly predominantly Pākehā but demographic change since approximately 2000 resulting in considerable synchronic diversity and no ethnic majority today.

Within the communities, participants were selected from two age groups, under 25 (younger) and over 40 (older). The corpus was designed in this way in order to test the effects of demographic change and community diversity on language change. Following the seminal work of Cheshire et al. (2011) and Cheshire (2013), the study was designed to explore the linguistic impact of significant in-migration to established communities. The hypothesis guiding the study as a whole (Meyerhoff 2017) is that speakers from communities in which rapid and significant levels of immigration mean there is no longer any ethnolinguistic majority will be the leaders of innovation and change. In the case of this study, that means that I expect the younger cohort of speakers in Mt Roskill to be most likely to exhibit linguistic innovations or innovative patterns of levelling (Trudgill 2004), much as Cheshire et al. (2011) have found for the adolescents they recorded in inner London.

The data is analysed using multiple regression analysis in Rbrul (Johnson 2009), results of which are presented in chapter 4. Factors found to be significant predictors of relative marker choice are analysed and discussed. I look both at the data as a whole to understand the broader patterns of Auckland English and at the communities individually to explore whether there might be features that (a) index belonging to a smaller speech community of a particular area of Auckland, and (b) suggest innovations are diffusing from communities of greatest ethnolinguistic diversity.

As part of the presentation of results, I look at some under-explored aspects of relative clause variation. The corpus contained a small number of examples of Ø subject relatives, a construction – show in example 8a. – generally considered ungrammatical in New Zealand English. Although there are too few tokens to perform any quantitative analysis, in section 4.7 I make some qualitative observations about these tokens and compare them with the analysis of Ø subject relatives from a variety in which these constructions are acceptable (Tagliamonte et al. 2005).

8. a.“but these are guys [Ø wouldn't know what was way was up].” – June – (Paptoetoe, older)
   b. “There'd be days [**where** I'd have one lecture]” – Chloe – (Titirangi, younger)

Adverbial relative clauses (as seen in example 8b) are generally excluded from analyses of relative clause variation on the grounds that they “admit a different range of variants” (Levey 2014: 26). However, as will be seen, the canonically studied relative markers exhibit similarly complementary distribution. Through an examination of some of the adverbial relatives in this corpus, I explore whether this further variation in the relative clause system should in fact be excluded from variationist analyses.

Chapter 5 discusses the implications of the results, particularly in reference to the research questions. The factors found to condition relativiser choice are predominantly syntactic and these are discussed and compared with other studies of relativiser variation. There are some social dimensions to the variation, notably gender for non-subject relatives and community for subject relatives. There is no evidence found for change taking place in the system but some evidence for age-grading. In the case of the most diverse community there is evidence that the relativiser system may have already been through a process of levelling.

2. Background

This chapter explores the history of relative clauses, both their development and the progression of their study. I begin with a chronological explanation of the syntactic understanding of relative clauses in section 2.1 and from here move on to discuss their development in the English language in section 2.2. Section 2.3 then looks at previous variationist studies of relative clauses to provide some idea of the usual syntactic and social factors found to condition their variation which will suggest the factors that should be looked for in this study.

2.1 What is a relative clause?

In order to investigate the variation in relativising strategies in Auckland English, it is necessary to have a clear definition of which constructions I am looking at. Huddleston and Pullum (2002) use the more general term "relative construction" to cover the whole range of constructions containing a relativising element. For instance, in addition to examples 1-7 in the previous chapter, Huddleston and Pullam also discuss constructions such as examples 9 and 10. However in contrast with examples 1-7, in example 9, the relativiser *that* introduces an adverbial relative (manner) and in example 10 the relativising element introduces parenthetical information (non-restrictive relative).

9. "I'm just used to the way [**that** people behave here especially when you're driving]." – Mishti – (Mt Roskill, younger)
10. "she always worked at like Manukau or like Hunter's Plaza [**which** is just around the corner]." – Steven – (Papatoetoe, younger)

Examples 8 and 9 are canonically excluded from variationist analyses of relative clauses which focus exclusively on the type of relative clauses shown in examples 1-7. As will become clear, the variationist tradition has seen these kinds of constructions as qualitatively different – both in their semantics and in terms of the interchangeability of the relativiser introducing the clause.

Huddleston and Pullum (2002) discuss the “traditional distinction” between restrictive and non-restrictive relative clauses where restrictiveness is defined in terms of the relationship of the clause to its antecedent. Restrictive relatives are defined as “'restricting' the reference of the antecedent to a subclass of itself” (Sigley 1997: 106) while non-restrictive relatives "simply make an additional assertion” (Lakoff 1968: 42). One of the key distinctions made is whether the information provided by the relative clause is essential to the meaning of the matrix clause (as in restrictive relatives) or not (as is the case of non-restrictive relative) (Hodges 1951). This distinction is used in most variationist studies of relative markers. For the most part, studies of variation in relative clauses in English treat this distinction as fundamentally unproblematic and focus exclusively on restrictive relatives. However, there’s a long tradition in the syntactic literature which problematises the distinction between restrictive relatives and non-restrictive relatives.

Earlier syntactic accounts from the 1970s and 80s have a tendency to seek an analysis of relative clauses that can include and account for all relative-like constructions. Even when they disagree fundamentally on how this should work, there seems to be a desire in this literature for an explanation of relative clauses that accounts for both non-restrictive and restrictive relatives.

Jackendoff (1977) proposes that non-restrictive, or appositive, relatives are underlyingly subordinate clauses and similar in deep structure to restrictive relatives. However this analysis then requires extra, specific operations for deriving the surface form of appositive relatives. He argues that the reason appositive relatives must follow restrictive relatives is because appositive relatives are N’’’ complements while restrictive relatives are N’’ complements. He then proposes different rules for appositive versus restrictive relatives in order to explain the difference in behaviour of these two types of relative clauses. One such set of rules explains why restrictive relatives can generate recursively but appositives cannot. Example 11 (taken from Jackendoff 1977: 172) shows two phrase structure rules, the first for restrictive relatives and the second for appositive relatives. 11b lacks the star present in 11a which permits multiple clauses so “multiple appositives are possible only through conjunction under the single S̄” (Jackendoff 1977: 172).

11. a. N’’ - > … N’ … ($\bar{S}$)*

b. N’’’ - > … N’’ … ($\bar{S}$) (Jackendoff 1977: 172)

Emonds (1979) on the other hand, favours an interpretation where appositive relatives are main clauses in deep structure. Underlyingly, they are the second in a set of two propositions, and are given their status as parentheticals by the rightwards movement of a constituent part of the first proposition (see figure 2.1). Their resemblance to restrictive relatives is, under this interpretation essentially superficial. However, they are able to be derived using the same rules as for restrictive relatives and do not need the extra operations as required, for example by Jackendoff, when they are treated as subordinate clauses.

Figure 2.1 – The formation of appositive relatives from main clauses, Emonds (1979: 213)

McCawley (1981) largely agrees with Emonds to the extent that he argues that non-restrictive relatives are underlyingly separate propositions whereas restrictive relatives “have an underlying structure in which they are conjoined with something” (McCawley 1981:123). He then develops an analysis for the construction of restrictive relatives whereby they are derived using “core” grammatical rules augmented by “patches,” in such cases where the “core” rules fail to provide full or correct derivations.

More contemporary accounts tend to be less concerned with finding an explanation for every possible kind of relative clause. They focus instead on providing a simpler account of unambiguous, unproblematic, or to use terminology borrowed from Ross (1995) “goody-goody-two-shoes” relative clauses. Carnie (2013) describes a relative clause as “a CP with a *wh*-element in it [that] modifies a noun” Carnie (2013:369). He makes brief mention of non-restrictive relatives, using the same definition as Lakoff (1968), namely that restrictive relatives restrict a frame of reference where non-restrictive relatives add parenthetical information. He does not go into any in depth discussion of this other than to state that “Demirdache (1991) suggests that restrictive relatives are adjoined to the N’, and non-restrictive relatives are right-adjoined higher up to the D’” (Carnie 2013:373). He does state that restrictive relatives must be closer to the head noun than non-restrictive relatives of their semantic scope but does not make clear, as Emonds (1979) does, that only restrictive relatives can generate recursively.

Carnie’s syntactic description is limited to restrictive relatives and is typical of most introductory syntactic accounts from the nineties onwards (cf. Adger 2003, Haegeman 1994). The surface structure of a relative clause is derived by the same *wh*-movement that forms *wh*-questions, that is, a *wh*-element moves up into the specifier position of the CP. For relative clauses that do not have an overt *wh*-element, the theory posits that this movement still takes place but it is an empty operator (OP) that now moves up into the Spec CP.

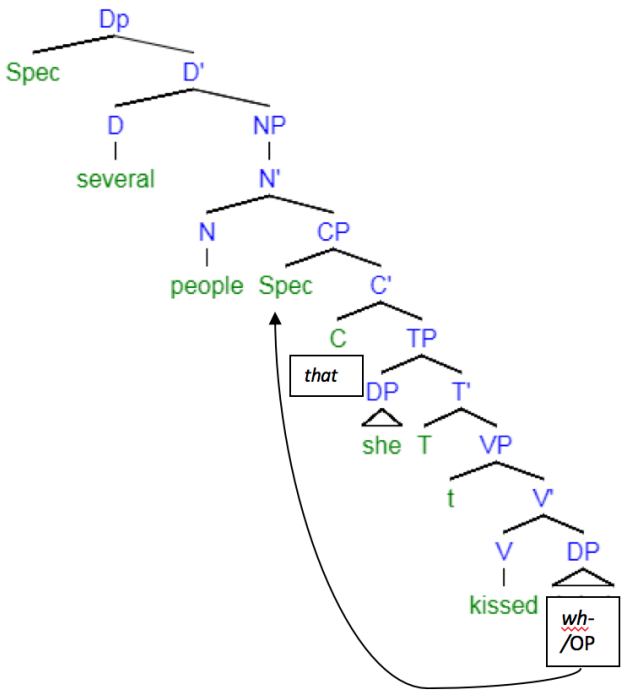

Figure 2.2 – Tree diagram illustrating *wh*- movement and position of relative markers in deep structure

The necessity for the empty operator movement is explained (Carnie 2013, Haegeman 1994) in terms of the theta criterion; the NP in the matrix clause cannot be assigned two theta roles (one from the matrix clause and one from the relative clause) so there must be some kind of element, whether overt or non-overt, for the predicate in the relative clause to assign its theta role to. This non-overt OP is also present in relative clauses that utilise *that* as a relative marker. According to Carnie (2013), in the underlying structure *that* is in the head of the CP and OP moves up into the Spec position. The absence of *that* in some cases does not seem to be problematic as "We know that complementiser *that* can delete in English" (Haegeman 1994). The conclusion that *that* and *wh*-elements occupy different positions in deep structure synchronically is motivated by evidence from earlier varieties of English where relative clauses can be introduced by a *wh*-word followed by *that* as seen in example 12.

12. thy freend [**which that** thou hast lorn]

"your friend that you have lost" (Carnie 2013: 373)

To summarise, relative clauses are formed by movement either of an overt *wh*-element or OP into Spec CP. In the cases where there is an OP rather than an overt *wh*-element, the CP is headed by *that* which can then optionally delete. One difference between restrictive and non-restrictive relatives is that non-restrictive relatives tend to favour the use of *wh*-relativisers while restrictive relatives are more variable (Ball 1996). Indeed, McCawley (1981) claims restrictive relatives cannot be introduced by *that,* although this claim has been shown to be false (Sigley 1997). Regardless, using the choice of relativiser to define whether a relative clause is restrictive in a study of (restrictive) relative variation is, as pointed out by Sigley (1997) methodologically suspect as it is inherently circular. Given this, providing a clear definition of non-restrictive relatives would appear to be crucial to any variationist study of restrictive relatives yet many studies restrict their commentary on restrictiveness to brief side notes: "Non-restrictive relative clauses differ on a number of counts from restrictive relative clauses, and thus cannot be considered in the same analysis" (Tagliamonte et al. 2005: 85). See also, for example, D'Arcy and Tagliamonte (2010) and Levey (2014).

In practice, making a binary distinction between restrictive relatives and non-restrictive relatives is problematic (Sigley 1997). Levey (2014) follows Huddleston and Pullum's (2002) definition of restrictive and non-restrictive relatives and states that clauses that simply "add parenthetical information about the antecedent in the matrix clause" (Levey 2014: 26) are excluded.

Ball is similarly black and white in her definition of restrictiveness, using examples from Sells (1985) to illustrate the distinction, but not addressing the fact that naturally occurring linguistic data may not always provide such clear-cut examples. This may be due in part to the fact that much, though not all of her data is from written rather than spoken sources. In written sources, non-restrictive relatives are more readily identifiable in that they are marked off by commas. This is a more objective criterion than the comma intonation criterion for spoken English whereby spoken non-restrictive relatives are supposed to be "characteristically marked off prosodically" (Huddleston and Pullum 2002: 1035). Nevertheless, studies using spoken data, such as Tagliamonte et al. (2005: 85) still use these

intonational cues to help determine restrictiveness. Further to this, Tagliamonte et al. (2005) created a strict set of criteria for exclusion:

> "We included only those relative clauses that "serve to identify their antecedent" (Denison, 1998: 278), thereby excluding a broad range of other types, most of which were easily identified by specific syntactic characteristics. Among these nonrestrictives were tokens in which an entire sentence served as antecedent, […] antecedents that were a full NP, […] and for which the relative clause added ancillary and or additional information" (Tagliamonte et al. 2005: 85)

D'Arcy and Tagliamonte (2010) also follow this set of criteria to isolate restrictive relatives in their data. It is perhaps worth noting that Tagliamonte et al. (2005) and subsequent studies following her work seem to assign an idiosyncratic meaning to "full NPs". Given the examples provided ("John-Docherty" and "the late Emrys-Hughes" (Tagliamonte et al. 2005: 85)), it appears that what is intended is antecedents that are proper nouns; since other syntactically full NPs are clearly included, as seen in example 13.

13. "And there were a wee alarm clock [**Ø** sat on the window.]" (Tagliamonte et al. 2005: 87)

Sigley (1997) discusses the possibility of viewing restrictiveness as a continuum, either "arbitrarily divided into categories […] or left undivided". While he expresses sympathy for this theoretical approach as perhaps more accurately representative of the grammar of English relative clauses, he acknowledges that in practice it is necessary to make a categorising distinction for the purposes of quantitative analysis. Sigley constructs a binary which he refers to as restrictive and non-restrictive but states "in my usage, these terms will in fact be equivalent to "nonparenthetic" and "parenthetic", respectively" (Sigley 1997: 133). Sigley relies in part on prosodic, intonation information to determine whether an utterance is parenthetical. He uses this (non)parenthetical dichotomy as one part of a factor group with which to encode relatives. The other factors are the level of specificity of the antecedent and the dichotomy illustrated by the terms "(non)restrictive, and (non)contrastive, (and possibly also "(non)defining") [which] divide clauses according to an objective, 'logical' function of (not) marking a proper subset of reference." (Sigley 1997: 129).

This factor group created four groups of relatives, namely “clauses ambiguous in context (removed from further analysis); objectively restrictive, nonparenthetic clauses; intonationally parenthetic, nonrestrictive clauses; and the intermediate nonrestrictive, nonparenthetic cases labelled "aspective"” (Sigley 1997: 130).

Intonation is clearly an important factor in determining restrictiveness and is thus a key factor in the decisions made in my own study of Auckland English. However, this criterion cannot be used in isolation, as Fox and Thompson (1990) argue that intonation does not always reliably disambiguate restrictive from non-restrictive relatives. Consequently, Tagliamonte et al.’s (2005) criteria were used and constructions were excluded that had full sentence antecedents (14a), proper noun antecedents (14b), and/or added ancillary information (14c).

14. a. “And they spoke to my parents and that makes you want to come over and you're comfortable with each other's parents, you know **which** always makes a difference” – Chloe – (Titirangi, younger)

    b. “a guy called Gilbert Myles, **who** ended up in New Zealand First” – Peter – (Mt Roskill, older)

    c. “we could have a meeting and exclusively talk in English **which** is fine, but to actually bench both Pacific- ah English and the Pacific language” – Catarina – (Papatoetoe, older)

Aside from non-restrictive relatives, there are some other constructions that are often explicitly mentioned as being excluded from studies of relative marker variation (D’Arcy and Tagliamonte 2010, Levey 2014, Tagliamonte et al. 2005). These include “headless” or “free” relatives as in example 15a and constructions known as “whiz deletion” (cf. Huckin et al. 1986), as in example 15b.

15. a. “we know **what** we might need” – Catarina – (Papatoetoe, older)

    b. “one of the Kauri trees **Ø** at the back of the farm” – Henry – (Papatoetoe, older)

As is clear from the above examples, with free relatives the relative clause lacks an (overt) head and with whiz deletion both the relativiser and its associated verb are deleted. As these are more unambiguously identifiable than non-restrictive relatives and are excluded from this analysis in keeping with the literature, it is not necessary to discuss them further.

Now that we have established an understanding of what relative clauses are, it is important to explore why they are worth studying. An answer lies in the textbook definition of restrictiveness. A (restrictive) relative clause gives specificity to the information a speaker is imparting. They allow the speaker to narrow the frame of reference of a particular NP and direct the listener to exactly what they are talking about. Examples like 16a and 16b illustrate this.

16. a. "further across was the primary school [**that** I went to]." – June - (Papatoetoe, older)
    b. "I know the guy [**who** works there]" – Jemaine - (Papatoetoe, younger)

June could be referring to any number of primary schools and Jemaine could be referring to many possible individuals but the use of a relative clause allows them to tell the listener that it is the particular primary school they went to or guy that works in a particular shop. Observing even subtle differences in how speakers define the parameters of the information they are supplying may add to our understanding of how speakers of New Zealand, or more specifically, Auckland English express specificity more generally. Moreover, as will be seen in the next two sections, the synchronic variation in relative clause marking continues to shed light on the historical grammaticalisation of relativisation strategies in English, and it affords us an excellent site for investigating whether there is evidence of grammatical levelling in the most diverse suburbs of Auckland.

2.2 In the beginning *that* was it

The Oxford English Dictionary lists the first attestations of *that* as a relative complementiser as early as c825 (OED 2018*a*). In fact, Gisborne and Truswell (2017) tell us that English has

used *that* – originally its Old English antecedent *þe* - "throughout its written history" (Gisborne and Truswell 2017:7). *Wh*- dependent relatives entered the language later and rather gradually. The OED lists the earliest restrictive relative use of the Middle English forms of *wh*- relative pronouns as c1297 for *who* and c1320 for *which* (OED 2018*b*).

Many previous studies of variation in relative clauses (Ball 1996, Nevalainen & Raumolin-Brunberg 2002, Romaine 1982, Tagliamonte et al. 2005) assume that *who* was the last relative to develop, around the early 15th century. Romaine (1982) argues that *wh*- relatives developed as a response to the disappearance of demonstrative relatives. But Gisborne and Truswell (2017) argue that, despite one form emerging as the other disappeared, simple replacement is not the sole factor motivating the emergence of *wh*- relatives. Instead, they provide data in support of a different semantic constraint on the new relativisers. A key aspect of their account of how *wh*- relatives entered the language involves Keenan and Comrie's (1977) accessibility hierarchy (AH):

SU > DO > IO > OBL > GEN > OCOMP (Keenan and Comrie 1977: 66)
(SU: subject; DO: direct object; IO: indirect object; GEN: genitive; OCOMP: object of comparison)

In brief, the relevance of the AH to relative clauses is that if a given language can relativise on any given position in the hierarchy it can relativise on all higher positions and relativisation on higher positions is more common cross-linguistically than on lower ones. Gisborne and Truswell (2017) demonstrate that *wh*- relatives seemed to enter the lexicon "by the back door" in that they initially relativised low-accessibility functions, leaving high-accessibility functions to be relativised by *þe,* then *wh*- relativisers gradually diffused up the hierarchy. Part of the evidence cited for this is that *where* relatives were the earliest to develop, occurring in adverbial like constructions from the 14th century (Gisborne and Truswell 2017:11).

If *wh*- relatives entered the language in complementary syntactic distribution to *þe,* this means we cannot assume, as D'Arcy and Tagliamonte (2010) did, that *wh*- relatives entered the language as prestige variants. The reason this is important for an analysis of relativiser variation is that D'Arcy and Tagliamonte argue the early status of *wh*- relatives as prestige forms continues to influence the variation today. They suggest that, in speech communities where there is any social stratification of the variation, *wh*- relatives have retained this legacy of prestige. It seems evident however that if *wh*- relatives entered the language in a different environment to *þe/that,* the distribution of these variants was initially syntactically not socially conditioned. As *wh*- relatives diffused up the accessibility hierarchy and came to be used in the same environments as *þe/that,* they could have developed connotations of prestige and this might subsequently influence their distribution. However until we get more historical sociolinguistic data on the diffusion of *wh*- relatives, we should be cautious about motivating diachronic change in relation to notions like prestige.

Regardless of disagreements over where and how *wh*- relatives entered the lexicon, they are indisputably the new kid on the block in terms of the history of English. Romaine (1982: 208) states that "WH entered a system where *that* and *Ø* existed in alternation". Tagliamonte et al.'s (2005) overview of the diachronic development of relativising strategies in English follows this assertion. They illustrate that *Ø*, like *þe/that,* has been used since Old English and, by the Middle English period, the main relativising strategies used were *that* and *Ø.* By Early Modern English, a distinction emerges between subject relative and non-subject relative clauses with different relativiser variants associated with the two types of relative clause.

Subject relatives are clauses where the relativised element is in subject position in the relative clause. *Ø* is less frequently used as a subject relative although it is still used as a relative in non-subject positions (Tagliamonte et al. 2005: 77). For example, while the construction in example 17 (a non-subject relative) would, to most speakers of New Zealand English, be perfectly grammatical with a *Ø* relative marker, example 18 (a subject relative) might not be. However, it is clearly not entirely unacceptable for all New Zealand English speakers as it was produced with a *Ø* relative marker, this is discussed in greater detail below.

17. “And that was another thing [(**Ø**/that/which) I like about the area]” – Mary – (Mt Roskill, older)
18. “and there’s a wooden bridge [(?**Ø**/that/which) goes across]” – Ian – (Papatoetoe, older)

The ungrammaticality of Ø subject relatives is referred to as the “anti-*that*-trace effect” (Bresnan 1972). The *that*-trace effect refers to constructions where the complementiser must be deleted, the anti-*that*-trace effect therefore refers to constructions where it must be retained. Explanations for the *that*-trace effect and the anti-*that*-trace effect are still debated amongst syntacticians (Carnie 2013). Douglas (2017) provides an analysis of both of these based on the “anti-locality effect”. I will focus on the details of this account only in so far as it relates to the anti-*that*-trace because this is where Douglas deals with the kinds of construction of interest to this study (as in examples such as example 18).

In brief, Douglas argues that Ø relative clauses contain less syntactic structure in their C-domains than *that* and *wh*-relative clauses do. His evidence for this comes primarily from the ability of relative clauses with overt relativisers to contain fronted adverbials while this is not possible for Ø relatives. Looking at the examples below, it is clear that adverbial material can be fronted only if the complementiser is present, as in examples 19b and 19b. However, with a Ø relative, as in examples 19c and 20c, the fronted adverbial does not have the same reading.

19. a. “the runner [**that** I was aiming for with the front wheel]” – Ian – (Papatoetoe, older)
    b. the runner [**that** [with the same wheel] I was aiming for]
    c. * the runner [**Ø** [with the same wheel] I was aiming for]
20. a. “Yeah I think that would be some problems […] [**that** I was aware of as a child]” – Chloe – (Titirangi, younger)
    b. Yeah I think that would be some problems […] [**that** [as a child] I was aware of]

c. * Yeah I think that would be some problems […] [**Ø** [as a child] I was aware of]

According to Douglas's analysis, the C-domain of *that* relative clauses contains what Douglas refers to as a ForceP and a FinP. He uses these terms, "for expository convenience only" (Douglas 2017: 4), to refer to the higher and lower C heads of *that* clauses. The ForceP, similar to the CP in Carnie's (2013) account, is either headed by *that* or has a *wh* element in Spec position. In a departure from the traditional account as summarised by Carnie (2013), in Douglas's (2017) view Ø relative clauses, in contrast, have no such complex C-domain structure. They are simply CPs with a single head which contains the Ø relativiser. This means that, in the case of subject relatives, a Ø relative clause would violate the anti-locality principle (Erlewine 2016: 445). This is because, in the SpecTP to SpecCP movement, the subject only crosses the maximal projection of TP, so the position the subject is base generated in, in the underlying structure, is within the locality domain of the final position of the relativiser it is coreferential with. In *that* and *wh* relative clauses however, movement from SpecTP to SpecForceP crosses both TP and FinP, is no longer local to the base subject position and thus acceptable. This is only an issue in subject relatives because other relativised elements would be moving from a lower projection subordinate to the TP and hence will be crossing a sufficient number of maximal projections. Figure 2.3 (taken from Douglas 2017: 6) illustrates this.

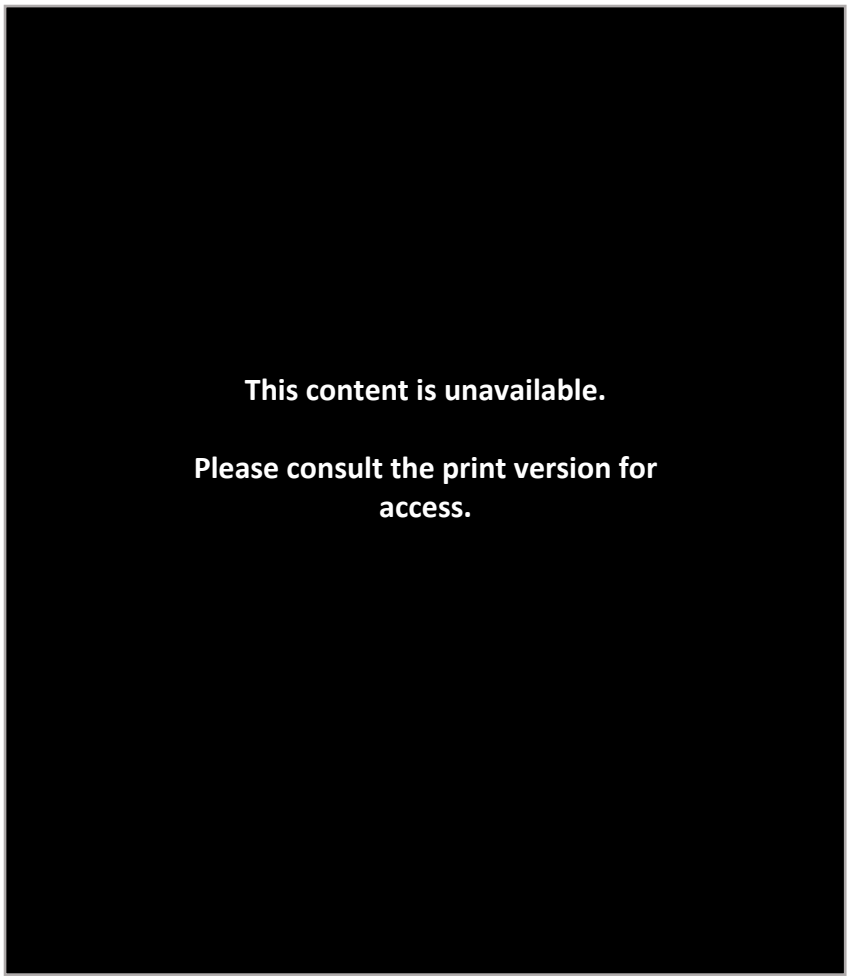

Figure 2.3 – Demonstration of how *that* subject relatives do not violate the anti-locality principle but Ø subject relatives do, Douglas (2017: 6)

This is all very well as an account of why Ø subject relatives are ungrammatical in standard English but Ø non-subject relatives are grammatical in standard English. However, it is important to remember that Ø subject relatives are not ungrammatical in all varieties and their acceptability (or not) tends to "align with the standard/nonstandard dimension of language use." (D'Arcy and Tagliamonte 2010: 385). What then can be said of varieties that do permit Ø subject relative clauses such as African American English (Kautzsch 2002) and non-standard varieties of British English (cf. Cheshire 1982)? Kandybowicz (2014) proposes an account of the *that*-trace effect (referred to as the Comp-trace effect) that is based on prosodic factors. He does not explicitly extend this to the anti-*that*-trace effect. Nevertheless, the argument that *that*-trace (and by extension anti-*that*-trace) effects "fall under the domain of the syntax-phonology interface" (Kandybowicz 2014: 227) might help to explain why Ø subject relatives are in some cases acceptable. If the proscription against Ø

subject relatives is not a purely syntactic rule but rather a prosodically motivated effect, it is less surprising to find these constructions appearing in some varieties. If, in some cases, prosodic well-formedness allows constructions that would normally violate the *that*-trace effect, or at least render an utterance comprehensible and/or acceptable to a listener, prosodic factors may also licence constructions which violate the anti-*that*-trace effect. Kandybowicz (2014) shows that this prosodic well-formedness is only possible in some, not all, constructions.

There are not only restrictions on *Ø* as a relativiser. *Wh*- relatives also have distributional restrictions but these derive from their status as pronouns which encode semantic information about their referent. Consider the below examples;

21 a. "Y'know and that's like the society [(which/**that**/Ø/*who) we're like living in]." – Aarush – (Mt Roskill, younger)

b. "We've got people [(who/that/**Ø**/*which) we go camping with each year]" – Charlotte – (Titirangi, older)

c. "There were, there were moments [(**when**/that/Ø/*who/*which) you thought, this, this, this could go badly]." – Nicholas – (Mt Roskill, older)

d. "It was just a place [(**where**/that/Ø/*who/*which) I could afford to go]" – Mary – (Mt Roskill, older)

Consider what this means for an analysis of English relativising strategies. If I were to follow the practice of excluding adverbial relatives because they do not admit the same range of variants (Levey 2014), the data in examples 21 a and b would oblige restricting the analysis to only subject relatives or only non-subject relatives with an antecedent in the main clause that has the same animacy (or the same adverbial meaning). Clearly this would undermine the use of a variationist approach to explore the constraints on relative markers in English. Taken along with Gisborne and Truswell's (2017) compelling picture of when and how *wh*-relatives entered the English relative system – led by adverbial relatives – this observation suggests that future variationist work on English relativisation might want to reconsider the practice of excluding adverbial relatives from the study of relative clause variation.

2.3 Variationist analysis of relative clauses

There is a fairly strong tradition of variationist research into relative clauses. Many of these studies focus exclusively on written data (cf. Hinrichs, Szmrecsanyi and Bohmann 2015, Huber 2012, Romaine 1982). However there are a number that draw either partially (Ball 1996, Guy and Bayley 1995, Sigley 1997) or exclusively on spoken data (D'Arcy and Tagliamonte 2010, Levey 2014, Tagliamonte et al. 2005). In this section I will first explore the syntactic features that have been found to condition the choice of relative marker before moving on to look at potential social aspects to the variation.

Romaine (1982) and Ball (1996) conduct diachronic investigations of the development of the English relative marker system. Although used as a touchstone in many contemporary accounts of relativisation (D'Arcy and Tagliamonte 2010, Levey 2014, Tagliamonte et al. 2005), Romaine (1982) is criticised by Sigley (1997) for her use of implicational scaling which "seems to offer too much scope for the analyst to force compliance with a model, with consequent reduction in the explanatory usefulness of the method" (Sigley 1997: 217).

Sigley's (1997) PhD thesis explores the variation in relative markers in New Zealand English, although his data comes from Wellington speakers rather than Auckland speakers. His is one of the few studies that includes an analysis of non-restrictive relatives as well as restrictive relatives. As previously discussed, his coding of restrictiveness is more complex and idiosyncratic than the coding systems used in other variationist studies of relative clauses.

Tagliamonte et al. (2005) and D'Arcy and Tagliamonte (2010) are both interested in Romaine's (1982: 212) argument that while *wh*- relativisers have completely infiltrated written English they have made little impact on the spoken language, especially in vernacular varieties. Ball (1996) also attests that it is in written domains that *wh*- relative markers made the earliest and strongest inroads. Tagliamonte et al. investigate this claim by exploring three vernacular varieties in the UK where they discover that *wh*- relatives have indeed made very few inroads into the spoken language. D'Arcy and Tagliamonte (2010)

look at urban speakers in Canada and discover that where *wh*- relatives are used in speech they do correlate with standardness and prestige.

Despite *wh*- relative markers having their strongest foothold in written domains, Hinrichs, Szmrecsanyi and Bohmann (2015), show a marked shift from *which* to *that* in written corpora of standard English from 1961 to 1992. They attribute this change to the editorial practice of "*which* hunting" as part of a prescription against the passive voice in many genres of writing. As part of an argument that linguistic prescriptivism can influence wider changes in language, they illustrate that avoidance of *which,* as part of an avoidance of passive constructions, led to a corresponding rise in use of *that* bringing patterns of relative marker variation in writing closer to the patterns of variation in speech.

Levey (2014) uses comparative variationist methods to investigate how children acquire the constraints of relative marker use. He looks at two different age groups of children; 8 to 9 year olds and 11 to 12 year olds as well as a group of 20 to 39 year old adults. With this he investigates which constraints on relativiser variation are acquired early and which later, whether children show differences in the relative markers they use, and the factors that influence their distribution, and what this can tell us about the underlying structural properties of relative clauses.

Huber (2012) compares relative marker variation in Ghanaian and British English corpora. He argues that the complexity of the English relativiser system makes it ideal for the kind of reanalysis and restructuring that take place as part of the nativisation of an emerging L2 variety. He shows that the Ghanaian English relative system is organised differently to the relative system in British English with differences in the factors that condition the variation.

Both Tottie and Harvie (2000) and Kautzsch (2002) look at the variation in relative markers in African American English (AAE). Kautzsch discusses the use of non-standard variants such as *what* as a headed relative marker – as previously discussed, *what* is only used in standard varieties as a relativiser for headless relatives – and the use of Ø as a relative marker in subject relatives. Kautzsch also calls for further study into non-standard varieties of English to provide fodder for comparison with his data (Kautzsch 2002:211). He argues that most

extensive studies of relative clause variation focus on standard varieties which therefore do not include the same range of variants as AAE (cf. Guy and Bayley's 1995 study of relative marker variation in standard written and spoken American English). Kautzsch also argues that studies of non-standard varieties either do not investigate the full range of constraints or, as is Kautzsch's criticism of Tottie and Harvie's (2000) study, they only look at one relative marker.

The range of constraints, both syntactic and social, that are explored in these previous studies of relative marker variation are discussed below.

2.3.1 Syntactic conditioning

One of the strongest factors that condition relativiser choice is the syntactic function of the relativiser in the relative clause. This could also be referred to as the grammatical role of the gap (Ball 1996). This refers to the syntactic position (subject or object) of the NP in the relative clause which the relativiser stands for. I will refer to this from now on as NP-Rel following terminology used in Keenan (1985). As demonstrated in the discussion in the last section, this is so significant a constraint that it is common, and often necessary for subject and non-subject relatives to be analysed separately (Ball 1996, Levey 2014). A large part of this is due to the grammaticality of Ø in subject relatives. If you are looking at a variety in which it is generally considered ungrammatical to have Ø as a relative marker in a subject relative it seems self-evident that you will find vanishingly few tokens of this (D'Arcy and Tagliamonte 2010, Levey 2014). Indeed, Levey (2014) illustrates that this particular constraint seems to be well established in children's use of relative clauses even though they are still acquiring some of the other constraints that appear in the adult speech in his data. Even in those varieties where Ø is attested as a subject relative in informal speech Ø is still used less often than it is in non-subject relatives (Tagliamonte et al. 2005:101). Conversely, *who* is often found with much greater frequency in subject than in non-subject relatives (Ball 1996, D'Arcy and Tagliamonte 2010, Levey 2014, Tagliamonte et al. 2005). Length and complexity of the relative clause has also been shown to condition relativiser choice, with longer and more complex clauses disfavouring Ø relativisers (Hinrichs, Szmrecsanyi and Bohmann 2015, Tagliamonte et al. 2005). The present study follows

Tagliamonte et al.'s (2005) coding of relative clauses consisting of three words or less (excluding the relativiser) as short (22a) and those consisting of four words or longer as long (22b and 22c). Relative clauses that contain just the verb and its nominal arguments are coded as simple (22a and 22b) while those containing another finite or non-finite clause are complex (22c).

22. a. "I don't tend to eat the stuff [**Ø** I bake]" – Abigail – (Titirangi, younger)
b. "[…] for the sake of upholding something [**that** no longer serves its purpose]" – Mark - (Papatoetoe, younger)
c. "all those sorts of things [**that** you do before you leave on your OE]" – Charlotte - (Titirangi, older)

Various features of the antecedent have also been found to condition relativiser choice (Ball 1996, Levey 2014, Tagliamonte et al. 2005). This is the NP in the matrix clause that the relative clause modifies, which I will refer to as NP-matrix. Animacy of the NP-matrix has a strong effect (D'Arcy and Tagiamonte 2010, Guy and Bayley 1995, Levey 2014, Tagliamonte et al. 2005). Given that *who* has the semantic feature [+human], it is unsurprising that human antecedents have been shown to favour *who* as a relativiser. Studies vary, however, in how they categorise the animacy of NP-matrix. Some opt for a binary distinction such as personal and non-personal (Ball 1996) or human and inanimate (Levey 2014). Tagliamonte et al. 2005 and D'Arcy and Tagiamonte 2010 have a more complex system that classifies the NP-matrix as human, people, collective, things and animals. Regardless of the categorisation system used, all these studies show a favouring effect of human antecedents for relative *who*, and once again this appears to be a constraint that is acquired relatively early (Levey 2014). This study is constrained to a three-way distinction between human (23a), other animate (23b) and inanimate (23c).

23 a. "I had children in classes who were kids of parents [**Ø** I taught the first time around]" – Kirsten - (Titirangi, older)
b. "My dog is the laziest dog [**Ø** you'll ever see in like the history of (life)]" – Mishti – (Mt Roskill, younger)

c. “Like the foundation [**which** this society's built on]” – Mishti – (Mt Roskill, younger)

Other features of the antecedent show less consistency in how they pattern between studies. For instance, there seems to be disagreement over whether indefinite antecedents favour overt relativisers (Huddleston 1971) or Ø relativisers (Tagliamonte et al. 2005: 104). In Levey’s data, only the youngest group of children showed any significant effect for animacy, favouring *who* with indefinite antecedents.

Also of possible interest is the grammatical role of the NP-matrix, that is, whether it is a subject (24a) or direct object (24b) or other object (24c) in the main clause, or the complement in an existential construction (24d).

24 a. “<u>all the oil</u> [**that** I put in] fell all over the driveway and stained the drive.” – Max - (Titirangi younger)

b. “we had <u>friends</u> [**that** had lived here all their lives]”- Nicholas - (Mt Roskill, older)

c. “So every time I drive past I just think of <u>the poor guy</u> [**that** fell in].” – Dwayne-Johnson – (Mt Roskill, younger)

d. “there are <u>things</u> [**that** we would change now] if we could go back” – Rose Petal – (Titirangi, younger)

According to Fox and Thompson (1990) there may be a discourse basis for the interaction between the grammatical role of the NP-Matrix and the grammatical role of the NP-Rel. Fox and Thompson argue that four possible combinations of NP-matrix and NP-Rel role combinations are favoured or disfavoured for discourse reasons. These are:

I. Non-human subject NP-matrix and object NP-Rel (S-O) is a favoured combination. They argue that this is favoured because the object relative helps to “ground” the subject NP and tie it to the wider discourse.

II. Non-human object NP-matrix and object NP-Rel (O-O) is a disfavoured combination. In this case the NP-matrix is often already grounded.
III. Where the NP-matrix is Existential, subject NP-Rel is favoured. Fox and Thompson show that Existential matrix NPs tend to be indefinite and human. For non-human referents, NPs tend to be grounded in relation to the humans that interact with them, but for human referents, NPs are grounded in relation to their own behaviour and actions, hence are the subject of the relative clause.
IV. Where the NP-matrix is Existential, object NP-Rel is disfavoured. See explanation for (III.).

Another interaction between NP-matrix and NP-rel that may condition relativiser choice is that the "distinctness" of the relative clause from the matrix clause may disfavour *Ø* relatives. Thompson and Mulac (1991) argue that the optional deletion of complementiser *that* is conditioned by how distinct the complement clause is from the "main" clause. They argue that some of these constructions are being "reanalyzed as unitary epistemic phrases" (Thompson and Mulac 1991: 249) and are thus more likely to have a *Ø* relativiser. Their data looks at clauses which are the complements of verbs and verbs such as "think" and "guess" which are used productively to express epistemicity occur more frequently without *that*. Whether this theory could be extended to account for the variation between *that* and *Ø* as relativisers is a possibility that will be explored in this study.

2.3.2 Social conditioning

Moving now to the extra-linguistic factors that may affect the distribution of relative markers, we have already seen from Levey's (2014) work that there is some effect of speaker age on relativiser use, although Levey is focused more on the acquisition of variation than he is with age related stylistic differences.

Class and style seem to be the most consistently significant social factors in relativiser choice. Both Tagliamonte et al. (2005) and D'Arcy and Tagiamonte (2010) show associations of *wh*- relativisers with higher levels of education and other correlates of prestige or "standardness". They argue that this may be an artefact of the way in which these relativisers entered the language as formal forms, namely by association with French and

Latin relative constructions (D'Arcy and Tagiamonte 2010). However, as noted in section 2.2, this does not tally with Gisborne and Truswell's (2017) reconstruction of how the *wh*-relative markers emerged in English. Tagliamonte et al. (2005) show that the vernacular varieties in their study remain fairly resistant to the use of *wh*- relativisers and argue that "*that* never shifted away from personal subjects in these varieties in the first place. Instead, it is still holding its own as the universal relative marker" (Tagliamonte et al. 2005: 106). Tagliamonte et al. (2005) argue that variation in relative markers in their study is stable and age graded. The competition between *that* and *Ø* appears fairly robust in non-subject relatives across age groups, and the fact that the distribution of the two variants hardly changes across age groups conforms to Labov's (1994) definition of stability in which variants are used at a similar rate relative to each other across time.

In subject relatives with human antecedents, *who* is used most frequently by speakers in the "middle-age" group and less by both older and younger speakers. However using this as an argument for either stability or prestige should be done with caution. A peak in usage of a variant in middle-aged speakers does not only occur in stable variation. While older studies (Downes 1998, Trudgill 1974) as well as more contemporary ones (Cheshire 2005) show stable prestige variants may be used more frequently by middle-aged speakers, other studies argue that middle-aged speakers may also shy away from innovative variables in situations of linguistic change (Buchstaller 2015, Sankoff 2006).

However with inanimate antecedents, Hinrichs, Szmrecsanyi and Bohmann (2015) show a marked shift from *which* to *that* in written corpora of standard English from 1961 to 1992. This is particularly interesting as this is where *wh*- relativisers made the earliest and strongest inroads. Ball (1996) states that *wh*- relativisers were taken up more quickly in written than spoken English. This adds a new dimension to the question of the stability of the variation. If *wh*- is losing ground in written environments, might it also be doing the same in other traditionally formal contexts? This would require future studies of variation in relative clauses to consider different genres of spoken English.

Gender is a social factor that its fairly ubiquitous in studies of sociolinguistic variation. It is possible that there is a tendency to overemphasise or give too much weight to apparent

gender differences when there is ample evidence that differences are often minimal (Hyde 2005) and variation between members of the same gender is as extensive, if not more so, than those between the groups (Chambers 1995, Eckert 1989). That said, studies of relative clause variation do find some significant effects for gender which it is worth while to note for reference and comparison with this study.

Levey (2014) sees some gender effects with the youngest children in his study, but this is much less pronounced in the older children and adults. The effect observed in his youngest group of children is a favouring of *who* in girls' speech. D'Arcy and Tagliamonte (2010) do not find a significant gender effect overall in their data, but they do see an interaction between gender and level of education. Participants with post-secondary education used more *who* in human subject relatives than those without post-secondary education. While there was no difference between men and women who had no post-secondary education, woman who did have post-secondary education used significantly more *who* than their male peers.

Looking at the only other New Zealand English study of relative clause variation, Sigley (1997) finds some rather different results in Wellington English. In his data, there is an interaction effect for gender and ethnicity. Pākehā females favoured *that* at least for personal subjects while all Māori participants slightly disfavoured it and Pākehā males disfavoured it even more. This large difference for Pākehā men and women is interesting given that it is the reverse of the effect seen in the previously mentioned studies. Sigley also found that Pākehā males used a higher rate of Ø than either Pākehā females or Māori males and females. He argues that this is because overt complementisers signal a speaker's intention to keep talking and allow them to maintain control of the floor. As the group with the highest social privilege, his argument goes, white men are the least likely to be interrupted and therefore do not need to use overt complementisers in order to maintain their turn in the conversation. This is an interesting notion although without further evidence to support this it remains largely speculation.

A common thread in all these previous studies that show a gender dimension to relative clause variation is that it is always only a sub-group of the corpus where this variation is

found, that is one age group (Levey 2014), one ethnicity (Sigley 1997) or one education level (D'Arcy and Tagliamonte 2010). This suggests that gender is not of primary importance in this variation.

2.4 Superdiversity and language change

The wider study (Meyerhoff 2017) of which this thesis is a part was designed to investigate the effect of young, superdiverse communities on language change. Cheshire et al. (2011) demonstrate that in London, the interaction of several ethnolinguistic groups where no single group constitutes a majority has led to widespread linguistic innovation and change. Kerswill and Williams (2000) show how young speakers can spearhead and diffuse change, carrying innovative variations that have emerged in communities where young people of different backgrounds interact regularly and taking these innovations out to neighbouring communities that may be more ethnolinguistically homogeneous.

One of the frequently observed effects of such linguistic diversity is levelling (Trudgill 2004). When a large number of new variables enter the feature pool (Mufwene 2001) linguistic change can occur such that a single, unmarked variant is favoured, and less common, marked variants undergo a process of attrition. As part of this, variables may lose some of their more complex grammatical constraints as one variant is generalised to all syntactic environments.

Auckland offers a particularly interesting demographic profile for exploring the impact of superdiversity on language change. The migration trends that have led to Auckland's increased diversity have occurred more recently (since the 1990s-2000s) and involve a greater number of distinct linguistic communities all being represented in some suburbs (Statistics New Zealand n.d.*a*). The demographics of the three communities sampled for this study are discussed in more detail in chapter 3. Crucially, Auckland currently contains both communities that are extremely heterogeneous and communities that have (so far) remained fairly homogeneous. In addition, due to the recency of the demographic change, Auckland also has communities that have become superdiverse recently enough to create a generational split. Older members of the community are still mostly from the same

ethnolinguistic background, while the younger members have grown up and been socialised in a community with much more diverse social networks and where Pākehā speakers of New Zealand English did not make up the majority of their speech community (at school or in the neighbourhood).

Conventional sociolinguistic analyses in general tend to look at fairly homogeneous language communities (Labov 2001, Meyerhoff 2010). This is no less true of studies of relative clause variation. Even those studies that investigate non-standard varieties of English are still looking at communities that have clear ethnolinguistic majorities (Kautzsch 2002, Tagliamonte et al. 2005). This study can provide new insight into relative clause variation in more diverse communities that are catalysts for change. We also know that language change is, to an extent systematic (Meyerhoff 2010). By studying the variation in relative markers in Auckland English at this point in time, it may be possible to offer some insight into how this variable may continue to develop.

There are a number of things I might expect to see in the data. First, if there is change happening, that it is led by younger speakers from more diverse areas, possibly being picked up by younger speakers from the more homogeneous community (Titirangi). If these younger Titirangi speakers are picking up on a change, however, I would expect them to be behind younger speakers from the other communities. Second if there is change emanating from the more diverse communities, it might be expected to be in the direction of a levelled system (Trudgill 2004) with a less complex system of syntactic constraints on the variation.

If such a levelled system is emerging in Auckland English, the relativiser *that* would seem the obvious candidate to emerge as the preferred, unmarked variant. There are several arguments in favour of this. As we have seen, *that* is the only relative marker that can be used in all syntactic environments with all antecedents and, as stated by Siegel (1997), regularity – that is a lack of exceptions – is an important factor in selecting a form from the feature pool (Mufwene 2001). Previous studies of relative clause variation (D'Arcy and Tagliamonte 2010, Levey 2014, Sigley 1997, Tagliamonte et al. 2005) also show us that *that* is the most frequently used relativiser overall in most varieties of spoken English. As

Cheshire et al. (2011) point out, frequency is another important determining factor in which variant will be selected from a heterogeneous feature pool.

This review should provide context and background to the four key research questions mentioned in chapter 1 and repeated here for clarity:

1.) Is there evidence of language change in the relative marker system of Auckland English?
2.) Does this change emanate from Auckland's more diverse communities and is it being led by younger speakers?
3.) How do the grammatical constraints on the variation of relative markers in Auckland English compare to those found in other studies of more homogeneous communities?
4.) Is there evidence of levelling, both in terms of one variant becoming more dominant and in terms of a loss of significant grammatical constraints on variation?

3. Method

3.1 Data collection

The data for this study was collected as part of "The Auckland Voices Project" (Meyerhoff 2017). Most of the sociolinguistic research carried out on New Zealand English to date has been focused in the Wellington (Holmes, Bell and Boyce 1991, Sigley 1997, Warren 2006) and Canterbury regions (Hay, Drager and Thomas 2013, Maclagan and Gordon 1999). However, as the most populous and also most diverse city in the country, Auckland provides a golden opportunity to carry out the kind of investigations into variation and change seen in other culturally and linguistically heterogeneous environments elsewhere in the world, for example London (Cheshire et al. 2011) and Toronto (Nagy 2011). Auckland represents roughly a third of the population of New Zealand; 1,415,550 of a national total of 4,242,051 (Statistics New Zealand n.d.*b*). According to the 2013 Census nearly 2 in 5 Auckland residents were born overseas – 39.1% - dramatically higher than the national average of 25.2%. Although the data for the more recent 2018 census was not yet available at time of writing, migration trends would suggest this proportion is more likely to have increased than dropped.

Data collection was conducted by means of traditional sociolinguistic interviews (Labov 1972, Meyerhoff, Schleef and Mackenzie 2015, Milroy and Gordon 2008) conducted in three different areas of Auckland between 2016 and 2018. These three areas were chosen for their contrasting demographic makeup which is as follows (see figure 3.1 for a map showing where each community is located):

- Papatoetoe which is very ethnically and linguistically diverse, and has been so for a long time
- Titirangi which has been and continues to be quite monocultural and monolingual
- Mt Roskill which has become much more multicultural and diverse in recent times

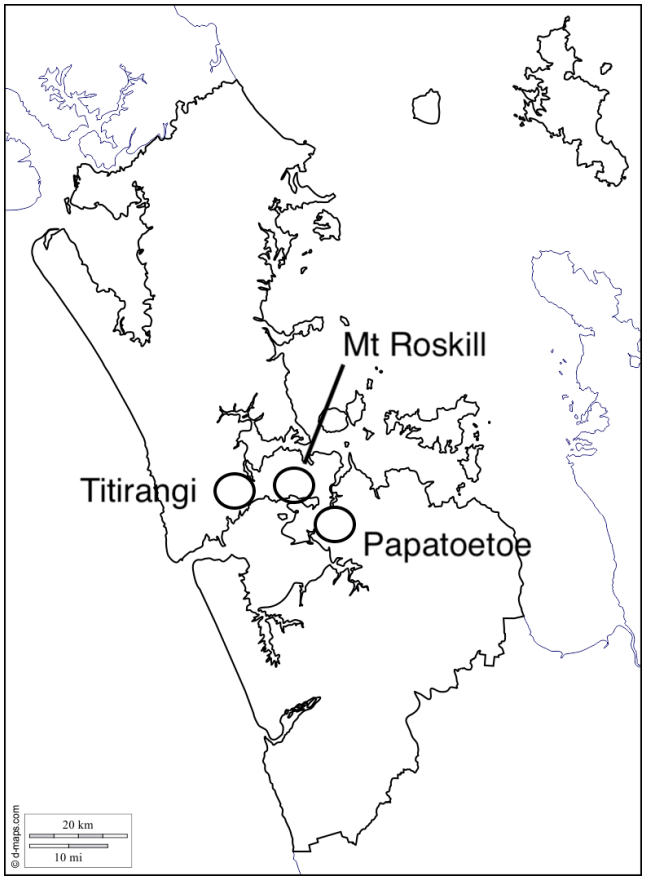

Figure 3.1 – Map of Auckland showing the three regions data was collected from

It is important to note that the boundaries of these areas were extended somewhat in the interests of finding suitable participants. This is especially true in the case of Papatoetoe which became a catchment of Papatoetoe/Manurewa/Ōtara. This is not problematic as care was taken not to stray into areas with widely different demographics. As someone who spent most of the first twenty years of my life in Auckland, I can attest to the fact that South Auckland, although loosely defined geographically, is quite a distinct "community" to Aucklanders. Titirangi was also extended to include participants from surrounding areas but also keeping to areas of the same demographic. For clarity, the communities will continue to be referred to as Papatoetoe, Titirangi and Mt Roskill throughout the study.

The 2013 Census can give us some further information about the demographics of these three areas. The Auckland local board areas correspond fairly closely to the areas chosen for the study. First is the Ōtara-Papatoetoe local board area. As mentioned, this is a diverse area in which 45.7% of the population are Pasifika, 15.6% Māori, 30.9% Asian and 20.7%

European. This local board also represents our youngest demographic with a median age of 29.3 years. Many people in this area also work locally suggesting dense social networks (Milroy and Milroy 1992). In stark contrast to this is the Waitakere Ranges area which contains Titirangi. Here there is a clear European majority (78.8%) being greater even than the national average of 74%. The median age is higher (36.8 years) and this is also the only community of the three where the median income for individuals over 15 is higher than the national average. Third is Puketāpapa board (Puketāpapa is the Māori word for Mt Roskill). Though Puketāpapa has a larger proportion of Europeans than Ōtara-Papatoetoe, at 38% it is still well below both Titirangi and the national average. This area has a large Asian population at 44.2%, 15.9% Pasifika and 5.7% Māori, but crucially, this area has no single ethnic majority. The median age is 34.3 years and while the median income is higher than that of Ōtara-Papatoetoe, it is still below the national average.

From each of the three communities, recordings were conducted in two age groups, under 25 and over 40. Table 3.1 shows the number of participants in each community by age group and gender at the time of this study (data collection continued into 2019).

	Papatoetoe		**Titirangi**		**Mt Roskill**	
	Female	Male	Female	Male	Female	Male
Under 25	6	7	5	5	8	6
Over 40	4	3	8	2	5	6

Table 3.1 – Breakdown of corpus by community age and gender

All participants are native speakers of New Zealand English. Table 3.2 gives an overview of each participant's linguistic repertoire and cultural background. This also highlights the demographic differences between the three communities. Despite the heterogeneity of the participants' linguistic backgrounds, all acquired New Zealand English as their L1 or as children in the critical period, either exclusively or alongside another language. In the case of some older South Auckland participants, the criteria had to be relaxed a little in order not to exclude older Pasifika members of the community so some of these speakers arrived in Auckland in their teens and acquired New Zealand English to high levels of proficiency then.

File	Pseudonym	Ethnicity	Language(s) used at home
Mt Roskill			
AK-M-R01	Aarav	Indian	
AK-M-R02	Kristen	Taiwanese/Malaysian	English
AK-M-R03	Mishti	Indian	Malayalam/English
AK-M-R04	Freya	New Zealand Kiwi	English
AK-M-A05	Aarush	Fiji Indian (Gujurati)	English
AK-M-A06	Mayumi	Japanese Malaysian	
AK-M-R07	Eoin (Chosen Pseud = Bobby)	Hong Kong Chinese/Kiwi	
AK-M-R08	Samaira	Indian	
AK-M-R09	Vihaan	Indian	
AK-M-R10	Noor	Somali	Arabic/English
AK-MY-R11	Flora	Pākehā/Chinese	
AK-M-M12	Siale	Tongan	Tongan/English
AK-MY-M13	Leletike	Tongan	English/Tongan
AK-MY-M14	Dwayne-Johnson	Tongan	Tongan/English
AK-M-M01	Chub	European	English
AK-M-M02	Ellen	European	English
AK-M-M03	Mary	European	English
AK-M-M04	Peter	European	English
AK-M-M05	Annie	European	English
AK-M-M06	Nicholas	Pākehā	English
AK-M-M07	Luke	Pākehā	English
AK-M-M08	Sharon	Pākehā	English
AK-MO-M09	Roslyn	Pākehā	English
AK-M-M10	John	Kiwi	English
AK-MOB-M11	Darryl	New Zealander	English
Papatoetoe			
AK-P-R01	Vernon	Samoan/Cook-Island/ Māori	English
AK-P-R02	Jemaine	Māori/Tongan/Niuean/Pākehā	English
AK-P-R03	Eleanor	Samoan/Pākehā	English
AK-P-R04	Jenny	Samoan/Cook-Island/Māori	English
AK-P-R05	Andrew	Samoan/Pākehā	English
AK-P-R06	Mark	Samoan	English/Samoan
AK-P-R07	Teagan	South African	English
AK-P-R08	Shannon	Samoan/Pākehā	English/Samoan
AK-P-R09	Chelsea	Tongan/Filipino	English
AK-P-R10	Adrian	Samoan	English
AK-P-R11	Vivian	Pākehā	English

AK-P-R12	Ioana	Samoan/Māori/Pākehā	English
AK-P-R13	Steven	Pākehā	English
AK-PB-M01	Ian	Pākehā	English
AK-P-M02	Muriel	Pākehā	English
AK-P-M03	June	Pākehā	English
AK-P-M04	Dot	Pākehā	English
AK-P-M05	Henry	Pākehā	English
AK-P-F06	Catarina	Samoan	Samoan
AK-P-M07	Terry	Chinese	Cantonese
Titirangi			
AK-T-V01	Max Power	Pākehā	English
AK-T-V02	Rose Petal	Pākehā	English
AK-T-V03	Jack	Pākehā	English
AK-T-V04	Ronaldo	Pākehā	English
AK-T-V05	Messy	Pākehā	English
AK-T-V06	Chloe	Pākehā	English
AK-T-V07	Caitlyn	Pākehā	English
AK-T-V08	Bindi	Pākehā	English
AK-T-B09	Billy	Pākehā	English
AK-T-B11	Abigail	Pākehā	English
AK-T-B12	Edward	Pākehā	English
AK-T-I01	Gina	Pākehā (Father part Māori)	English
AK-T-I02	Kate	Pākehā	English
AK-TB-I03	Valerie	Pākehā	English
AK-T-I04	Kirsten	Pākehā	English
AK-T-I05	Deb	Pākehā	English
AK-T-I06	Charlotte	Pākehā	English
AK-T-I07	Rachael	Pākehā	English
AK-T-I08	Jim	Pākehā	English
AK-T-I09	Prue	Pākehā	English
AK-T-I10	Matt	Pākehā	English

Table 3.2 – Participant pseudonyms, ethnicity and linguistic background

All participants were provided with an information sheet about the study (appendix A) and asked to sign a consent form (appendix B). They were given the option of selecting a pseudonym or were assigned one. The Auckland Voices project as a whole was granted ethics approval (reference number 22841) and all researchers and transcribers signed confidentiality agreements.

3.2 Coding

Interviews were transcribed and coded using the ELAN software program (Wittenburg et al. 2006). Each interview was checked twice for relative clauses including reliability checks with an independent trained linguist and coder (Helen Charters, University of Auckland). As previously mentioned, including or discarding potential relative clauses involved applying a fairly complex set of criteria. The opaque nature of restrictiveness (Sigley 1997) called for cross referencing of Tagliamonte et al.'s (2005: 85) criteria (excluding relative clauses that have full sentence antecedents, proper noun antecedents and/or added ancillary information) with (i) whether it would be possible to generate recursive relatives with this construction as per Emonds (1979) and (ii) my own judgements based on the context and prosodic cues of the utterance (this latter criterion was much assisted of course by working from ELAN files). Even with this relatively extensive set of criteria, there are still tokens that appear very ambiguous between restrictive and non-restrictive functions. For instance, in example 25;

25. "my brother [**who** lives at home at the moment]" – Jenny - (Papatoetoe, younger)

This could either be specifying a particular brother out of several, or simply providing extra information about a brother whose identity is already clear. In the case of this token, context and intonation suggested a restrictive interpretation was more likely, the speaker had recently been talking about other family members and there was a lack of the intonation pattern that indicates parenthetical information (Sigley 1997). Given this, this token was included.

Another issue is the inclusion of adverbial relatives. However, as previously mentioned, adverbial relatives are generally excluded from other studies of relative clause variation, I would argue that there are reasons for including them. While these constructions are introduced by a different range of *wh*- relativisers (Levey 2014) this is not in kind different from the complementary distribution of *who* and *which.* Moreover, they all have the same optional variation of some *wh*- form (26a), *that* (26b) and Ø (26c).

26. a. “we try and like head up to areas [**where** it snows and things like that]” – Mayumi – (Mt Roskill, younger)
b. “it's an Island thing [**that** everyone's just kind of mucking around]” – Vivian – (Papatoetoe, younger)
c. “if we ever got to the point [**Ø** we'd do them]” – Charlotte – (Titirangi, older)

It is worth noting too that some of these constructions do not seem unequivocally adverbial. That is, the antecedent they are relativising does not seem to refer to place, time or manner. This is more the case with *where* relatives than with *when* relatives as seen in example 27. Adverbials were therefore included in the initial coding but are excluded from some later analyses on the grounds that they will be explained in the quantitative analysis of results. The issues of whether the more ambiguous tokens should be viewed as adverbials is also addressed in the results and discussion section.

27. “No, only the one [superstition] [**where** they say not to sleep under the moon.]” – Shannon – (Papatoetoe, younger)

The initial level of coding specifies which relativiser is used; *that, Ø* or *wh-* (*which*, *who, where, what* etc.). The second level of coding involves the role of the relativised element, both in the relative clause and in the matrix clause. Following Keenan (1985) the grammatical role of the element in the relative clause is referred to as NP-Rel. NP-Rel is coded as either Subject (28a), Direct Object (28b), Other Object (28c), or Adverbial (as seen in 26a above). Other Object covers constructions where the relativiser is an indirect object and was also the coding used for the adverbial relative constructions discussed previously.

28. a. “I got an older sister [**who's** police woman].” – Aarush - (Mt Roskill, younger)
b. “And now he's in the new campus [**that** they've got]” – Vivian – (Papatoetoe, younger)

c. “She showed us the 60 Minutes episode [**Ø** she was on]” – Billy – (Titirangi, younger)

The grammatical role of the antecedent in the matrix clause is referred to as NP-matrix and is coded as Subject (29a) Direct Object (29b) Other Object (29c) or Existential (29d).

29. a. “all the oil [**that** I put in] fell all over the driveway and stained the drive.” - Max – (Titirangi, younger)
b. “we had friends [**that** had lived here all their lives]” – Nicholas – (Mt Roskill, older)
c. “So every time I drive past I just think of the poor guy [**that** fell in].” – Dwayne-Johnson – (Mt Roskill, younger)
d. “there are things [**that** we would change now if we could go back]” – Kate – (Titirangi, older)

NP-Rel is also coded for length and complexity, following the criteria in Tagliamonte et al. (2005). Relative clauses of three words or less, excluding the relativiser, are coded as short (29a) and those of four words or more are coded as long (30b and c). Contractions such as let’s, it’s and can’t are counted as two words, so example 30b is counted as long.

Relative clauses can also be distinguished not just on the basis of string length but also on the basis of their internal syntactic structure. Relative clauses that contain just the verb and its nominal arguments are coded as simple (30a and 30b) while those containing another finite or non-finite clause are complex (30c).

30. a. “There's a lot of focus on just the main subjects [**that** are career driven]” – Vihaan – (Mt Roskill, younger)
b. “my sister [**that's** older than me] stays with him.” – Dwayne-Johnson – (Mt Roskill, younger)
c. “I have like, Samoan co-workers [**who** always ask ((pause)) if I can speak it]” – Jenny – (Papatoetoe, younger)

NP-matrix is coded for animacy and definiteness. In terms of animacy, NPs are coded as human, Other Animate or Inanimate. Collectives that involve humans are coded as human as seen in 31. Definiteness is not just in relation to whether the NP is introduced by a determiner that is definite (32a) or indefinite (32b) but also in relation to the semantic sense of the NP. Indefinite pronouns (32c) are coded separately following the literature (Ball 1996, Levey 2014, Tagliamonte et al. 2005) but in this study, in addition to their coding system, codes negative indefinite (32d) separately because these are "absolute negators" (Huddleston and Pullum 2002).

31. a. "I don't know of many people [**that** have moved out of Roskill]" – Samaira – (Mt Roskill, younger)
    b. "there was sort of government agencies [**which** would take on juniors and train them]." – June - (Papatoetoe, older)
32. a. "this one particular lady [**that** always was on my back about everything]" – Steven – (Papatoetoe, younger)
    b. "its going to other community members [**that** can afford it]" – Noor – (Mt Roskill, younger)
    c. "Someone [**who** knows how the community works there]" – Teagan – (Papatoetoe, younger)
    d. "I mean y'know stupid stuff but nothing [**that** was actually probably really life threatening]" – Mary – (Mt Roskill, older)

The structure of the coding in ELAN allows for maximal information in the subsequent quantitative analysis. The tier dependencies are shown in Figure 3.2 so readers can better understand the structure of the dataset.

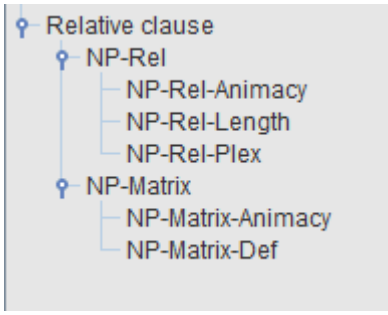

Figure 3.2 – ELAN tier dependency structure

3.3 Data analysis

Rbrul (Johnson 2009) is used to perform multiple regression analysis on the coded data. Multiple regression analysis is particularly important with data like this in which tokens are distributed unevenly across all the independent factors coded for, both linguistic (internal) and social (external) (Tagliamonte 2006). Stepwise regression allows us to model the probability of a particular outcome in the dependent variable (in this case, which relative marker is used) against multiple independent/explanatory variables (antecedent animacy, clause length, speaker gender, age etc.).

Rbrul allows us to test not only multiple independent variables but also to model random effects, such as speaker, that are not controlled for in the data. This does not mean ignoring individual variation but rather that it is accounted for in the model (Johnson 2009). As Johnson points out, if individual speaker differences are ignored entirely, the significance of external factors such as gender and age may be overestimated. However, if speaker is included as a separate factor group, multiple regression analysis might find individual variation accounts for so much of the variation that truly significant external factors may be discounted, for these reasons, Johnson recommends building models with speaker as a random effect.

In this study I therefore opt to include speaker as a random effect – an advantage of Rbrul over other regression software – to mitigate both the under and over estimation of the significance of external factors. This is done with caution, however, due to the fact that there is a very uneven distribution of relative clauses across individual speakers. In some

cases, especially where individual speakers have categorical behaviour (100% use of one variant), speaker is not included as a random effect as random effects analysis may not be appropriate for such unevenly distributed data (Roy and Levey 2014).

4. Results

4.1 The Dataset

A total of 2482 tokens of relative clauses were coded across the entire data set. Adverbial tokens are excluded from this first stage of analysis in order to keep results comparable with those of previous studies (D'Arcy and Tagliamonte 2010, Levey 2014, Tagliamonte et al. 2005). These are discussed separately in section 4.7.2. This leaves us with a total of 2261 tokens distributed as shown in table 4.1.

Relative marker	**%**	**N**
That	60.7	1373
Ø	20.6	466
Who	13.7	309
Which	5.0	113

Table 4.1 – Distribution of (non-adverbial) relative markers

The mean number of tokens per speaker was 34.3 and the median 30. There was a lot of inter-speaker variation in the number of relative clauses produced. The fewest tokens (5) were produced by a 10-year-old male from Titirangi and the most (114) by a 22 year old female from Mt Roskill.

Table 4.2 provides a breakdown of the relative clause distribution across the sample according to the three key social factors the data set was designed to test, namely community, gender and age.

	Community			**Gender**		**Age group**	
	Papatoetoe	Mt Roskill	Titirangi	Female	Male	Older	Younger
Mean	34.6	33.68	34.6	37	30.9	34.89	33.79
Median	31.5	26	34	31.5	27.5	30	30
Total N	727	842	692	1332	929	977	1284

Table 4.2 – Distribution of relative clauses according to social factors

The fact that Mt Roskill accounts for the highest number of relative clauses in the data is not surprising given that there are 25 speakers from Mt Roskill and only 20 each from Papatoetoe and Titirangi.

From this overview of the data, let us move on to the key issue for variationist analysis; which relative markers are used, where, and by whom. Table 4.3 shows the distribution of relative markers across the data by community. There are several things to note from this table. Relativiser *that* has a clear majority, accounting for over half of the total number of relative markers produced, a total of 1373 which equates to 61% of the data.

Although not included in table 4.1, there are 2 tokens of *what* as a relativiser, as shown in example 33.

33. a. "You want to always glorify an experience [**what** comes on TV]" – Aarav – (Mt Roskill, younger)
    b. "I was nosey one day and went and at looked at mum's old houses old houses [**what** she had]" – Edward – (Titirangi, younger)

Although acceptable in some varieties, *what* is not generally part of the New Zealand English repertoire for headed relative clauses, although it is perfectly acceptable for headless relatives as seen in example 14a in section 2.1. Not much more can be said based on two tokens so they are necessarily excluded from the rest of the analysis. I note their occurrence in the interests of accountability.

	Ø		**that**		**which**		**who**		**total**
	N	**%**	**N**	**%**	**N**	**%**	**N**	**%**	
Mt Roskill	221	26%	470	56%	39	5%	112	13%	842
Papatoetoe	119	17%	457	66%	28	4%	88	13%	692
Titirangi	126	17%	446	61%	46	6%	109	15%	727
total	466	21%	1373	61%	113	5%	309	14%	2261

Table 4.3 – Distribution of relative markers by community [due to rounding, percentages may not equal 100]

Which and *who,* the two *wh*- relative markers are used at a slightly higher rate in Titirangi than in the other two communities. As discussed earlier, the Titirangi population generally represents a higher socio-economic group than the other two communities. Sigley (1997) and Tagliamonte et al. (2005) show that *wh*- relatives are generally used to a greater degree in more formal styles of speech and/or by speakers with a higher level of education or socio-economically prestigious jobs. The distribution shown in this study would be consistent with these prestige/standardness associations of *wh*- as a relative marker.

Mt Roskill speakers meanwhile use more Ø relatives than the speakers in other communities. I have seen little evidence in the literature to suggest that there is any particular social index associated with a preference for Ø relatives, except in varieties where Ø subject relatives are possible (D'Arcy and Tagliamonte 2010). Further analysis will tell us whether this difference is indeed significant, and if it is, whether it is truly a community effect or whether it is attributable to some other factor that is interacting with community.

Table 4.4 shows the distribution of relative markers depending on the grammatical role of the NP-Rel, i.e. the relativised element in the relative clause. Subject relatives are the most frequently occurring form in the data at 50.4% (N=1139), followed by direct objects (901). This is to be expected given that these are the top two positions on the Accessibility Hierarchy (Keenan and Comrie 1977) and thus the most available linguistically for relativisation.

	Ø		that		which		who		total
	N	%	N	%	N	%	N	%	
Subj	14	1%	754	66%	80	7%	291	26%	1139
DO	366	41%	512	57%	13	1%	10	1%	901
Other Obj	86	39%	107	48%	20	9%	8	4%	221
total	466	21%	1373	61%	113	5%	309	14%	2261

Table 4.4 Distribution of relative markers according to their role in the relative clause [due to rounding, percentages may not equal 100]

As previously discussed, Ø subject relatives are generally considered ungrammatical in New Zealand English but our data included 14 tokens of these. While this is too small a number to do any meaningful quantitative analysis, section 4.7.1 looks at these 14 tokens and their distribution qualitatively. I will explore these examples in greater detail because, while it may be possible to explain them away as simply errors in either production or coding, they could also be the embryonic stage of syntactic variation (cf. Cheshire 2013). It is relatively rare that we see the earlier stages of linguistic variation and change. Perhaps this is because most linguistic innovations look like mistakes or idiosyncratic quirks until they are reasonably established. So, while 14 tokens of a construction generally considered to be ungrammatical may be insignificant errors, it is important to at least make note of them, for posterity and documentary completeness.

Who is the most frequent *wh*- relative and it predominantly occurs in subject relatives. There were vanishingly few tokens of its object and genitive forms *whom* and *whose* (2 and 5 respectively). I excluded the genitives and included the 2 tokens of *whom* with the object relative tokens of *who* as these appeared to be structurally and semantically equivalent.

4.2 Complementary distribution of Relative Clauses

Figure 4.1 shows the rate of use of *that*, *Ø,* and *wh*- relativisers by gender and community.

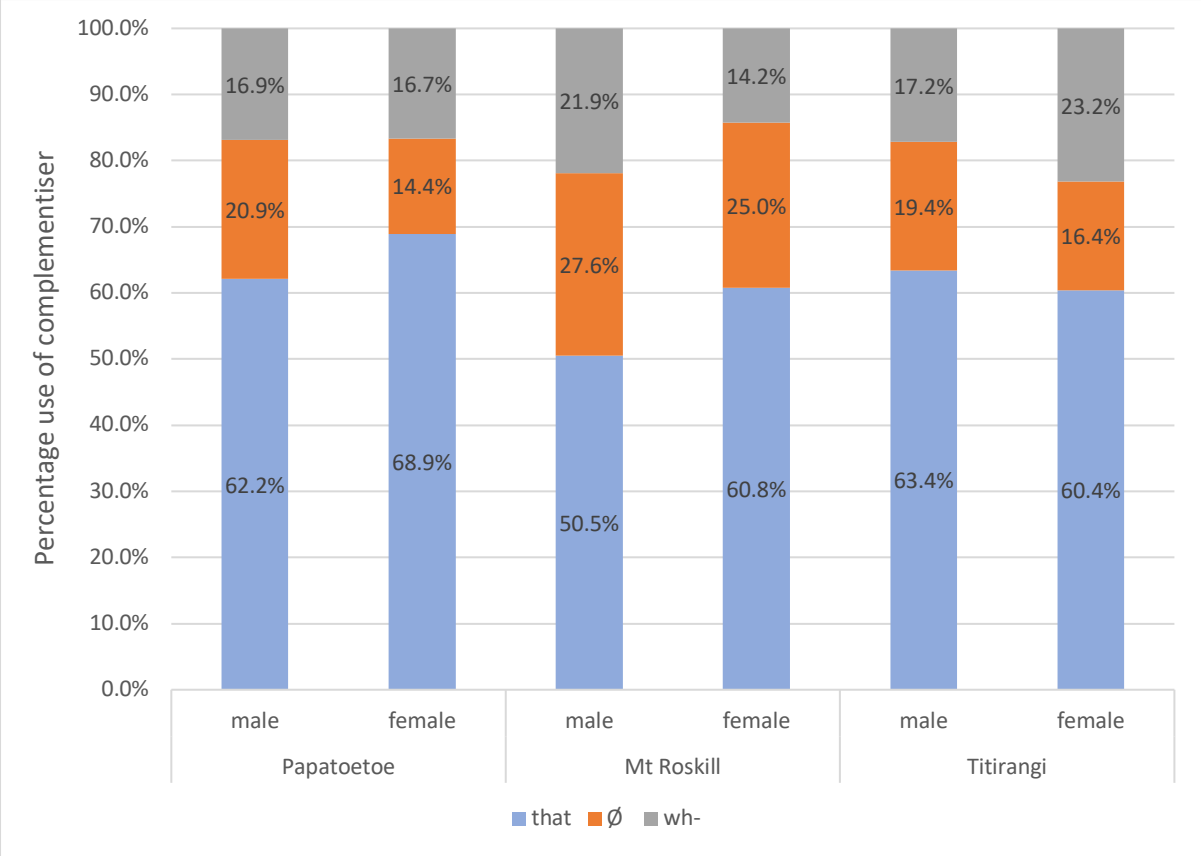

Figure 4.1 – Percentage of *that, Ø* and *wh-* (*who* and *which*) in subject and non-subject relatives by community and gender.

Figure 4.1 confirms the impression gained when looking at the raw numbers, namely that speakers in Mt Roskill uses more *Ø* relatives. Not only this, but in each community the men use a higher rate of *Ø* relatives than the women. Mt Roskill men not only have a high rate of *Ø* relatives, they also use a lot of *wh-*, second only to the Titirangi women. Mt Roskill women, by contrast, have the lowest rate of *wh-* of any of the groups. Given that this is a combination of the counts for both *who* and *which,* it is necessary to look in more detail at the different constructions these relativisers appear in to see which syntactic and social factors are influencing their distribution.

It is clear from table 4.4 that the distribution of relativisers is not even across all constructions. The second most commonly used relativiser, *Ø,* is in fairly clear competition with *that* in non-subject position. A *Ø* is used 40% of the time in all non-subject relatives compared to 55% for *that*. In subject relatives, the strongest competitor with *that* is *who.* This asymmetric distribution means that it is necessary to model the variation in subject and

non-subject relatives separately which is exactly the approach taken by other variationist studies of relative clauses (Ball 1996, Levey 2014, Tagliamonte et al. 2005). One other distinction that is often made in these studies is based on the animacy of the antecedent. The variation between *who* and *that* is generally modelled exclusively for human subject relatives. Table 4.5 shows clearly that tokens of who are concentrated overwhelmingly in subject relatives with human antecedents. Table 4.5 also shows that subject relatives have substantially more human antecedents than non-subject relatives as seen in table 4.6.

	Human	**Other Animate**	**Inanimate**
null	11	1	2
that	382	8	364
which	10	1	69
who	285	5	1
Total	688	15	436

Table 4.5 – Distribution of relative markers in subject relatives according to animacy

	Human	**Other Animate**	**Inanimate**
null	73	11	368
that	114	6	499
which	1	1	31
who	17	1	0
Total	205	19	898

Table 4.6 – Distribution of relative markers in non-subject relatives according to animacy

To summarise, in non-subject relatives the main alternation is between *that* and *Ø,* regardless of the animacy of the antecedent. In subject relatives, where Ø is not (generally) an option, *who* alternates with *that*, principally where the antecedent is human and sometimes also with other animates as well (but there are too few tokens of other animates (N=5) to make a strong generalisation). *Which* is most frequently used to introduce inanimate subject relatives, although this is still a comparatively small number of tokens (N=69).

There is one token of *who* used with an inanimate antecedent, shown below in example 34;

34. “So you get the websites [**who**'ve got the stock]” – Luke – (Mt Roskill, older)

Collective nouns that represent groups of humans such as companies and organisations were usually coded as human and, in this case, “website” seems to be behaving like one of these. This token was therefore recoded as human.

4.3 Analysing the variation - Multiple Regression Analysis

I now turn to modelling the variation of the most frequently occurring relative markers. First, I model the probability of the use of *Ø* over *that* in non-subject relatives and briefly outline the constraints that appear as significant. I then model the probability of *who* over *that* in subject relatives with human antecedents. I will only give a brief overview of significant constraints found in the models as they will be discussed in detail in section 4.4.

4.3.1 Non-subject relatives

It appears that *that* is the “default” relative marker, given that it is not only by far the most frequently used but it is also the only relativiser used unproblematically in all constructions and with all antecedents. I will now model the variation of the two most frequent alternatives to *that,* namely *Ø* and *who*. As these variants have complementary distributions, I look specifically at them separately, starting with the *that/Ø* alternation in non-subject relatives. This results in a data set of 1064 tokens for analysis.

All the linguistic and social factors identified as possible predictors were modelled as independent factors, with speaker as a random effect. *Ø* relatives were modelled as the application value of the rule. Table 4.7a shows the significant predictors from the variable rule analysis of the alternation between *that* and *Ø* in non-subject relatives.

Non-subject Relatives Ø vs Ø + *that* **R2 =** 0.253 **input probability =** 0.528 **N=** 1064			
	FW	**%**	**Total N**
Definiteness of the antecedent	p.<0.0000		
Indefinite Pronoun	0.67	61.5	104
Definite	0.46	46.2	556
Indefinite	0.36	32.2	404
Gender	p.<0.05		
Male	0.58	49.3	420
Female	0.42	37.9	644
Length of relative clause	p.<0.05		
Short	0.54	48.4	512
Long	0.46	36.8	552

Table 4.7a - Multiple regression analysis of factors contributing to the choice of Ø in non-subject relatives. (Random effect: Speaker. Not selected as significant: Role of antecedent in matrix clause, community, age group, complexity of relative clause, animacy of antecedent.)[2]

The most significant predictor of the choice of Ø as a relative marker is the definiteness of the antecedent with indefinite pronouns as the antecedent (NP-Matrix) favouring Ø. Negative pronouns are excluded from this analysis due to their very low frequency (N=7) and the fact that they are semantically distinct from other pronouns (Huddleston and Pullam 2002). Earlier analyses suggested that it was correct to keep these separate from the other indefinite pronouns because what few tokens there were had a very different distribution from indefinite pronouns, with negative pronouns strongly disfavouring Ø relatives (1/7) while indefinite pronouns favoured Ø relatives (64/104).

The only significant social factor in table 4.7a is speaker gender. When I ran this same model without speaker as a random effect (table 4.7b), community did emerge as a highly significant constraint (p<0.00), and this suggests that the variation between individuals is what prevents community showing up as a significant predictor. Even with speaker as a random effect, community still leans toward significance (p=0.0916) suggesting there may be something further to explore here.

[2] For full results of this multiple regression analysis with speaker as a random effect, including all factors not found to be significant, see Appendix C.

Non-subject Relatives Ø vs Ø + *that* **R2 =** 0.105 **input probability =** 0.523 **N=** 1064			
	FW	**%**	**Total N**
Definiteness of the antecedent	p.<0.0000		
Indefinite Pronoun	0.67	61.5	104
Definite	0.46	46.2	556
Indefinite	0.36	32.2	404
Community	p.<0.001		
Mt Roskill	0.59	51.9	414
Titirangi	0.46	35.9	345
Papatoetoe	0.45	36.7	305
Gender	p.<0.01		
Male	0.58	49.3	420
Female	0.42	37.9	644
Role of the antecedent NP in the matrix clause	p.<0.05		
Subject	0.57	51.8	276
Existential	0.49	34.1	88
Other Object	0.47	40.9	274
Direct Object	0.46	39.0	426
Length of relative clause	p.<0.05		
Short	0.54	48.4	512
Long	0.46	36.8	552

Table 4.7b - Multiple regression analysis of factors contributing to the choice of Ø in non-subject relatives. (Not selected as significant: age group, complexity of relative clause, animacy of antecedent.)

In sum, the apparent preference for Ø relatives in Mt Roskill, seen in tables 4.2 and 4.7b is either an effect of individual speaker variation or something to do with the gendered sample of that community. At this point, there is no reason to think that Ø relativisers in non-subject relative clauses is a characteristic of different community grammars.

4.3.2 Subject relatives

I now turn to subject relatives, where the salient alternation is between *that* and *wh-* forms (mainly *who*). As seen in table 4.5, *who* mainly occurs with human antecedents so I restrict my analysis to these. This dataset consists of 668 tokens.

Tables 4.8a and b show the results of the variable rule analysis of *who* versus *that* in subject relatives with a human antecedent. When run with speaker as a random effect (table 4.8a), the model found only one of the independent variables tested to be a significant predictor of the use of *who,* definiteness of the antecedent.

Subject Relatives *who* vs *who* + *that* **R2 =** 0.43 **input probability =** 0.446 **N=** 668 (human antecedents only)			
	FW	**%**	**N**
Definiteness of the antecedent	p.<0.01		
Indefinite Pronoun	0.65	57.1	42
Indefinite	0.49	46.8	365
Definite	0.36	34.9	261

Table 4.8a - Multiple regression analysis of factors contributing to the choice of *who* in human subject relatives. (Random effect: Speaker. Excluded from model as not significant: Role of the antecedent NP in the matrix clause, community, gender, age group, length of relative clause, complexity of relative clause.)[3]

Despite the fact that it is now axiomatic to model speaker as a random effect in order to minimise overstating social constraints, in this case, I feel warranted in presenting the results without speaker as a random effect because 15 of the individual speakers have categorical behaviour (3 use 100% *who* and 12 use 0% *who*). Roy and Levey (2014) that such unevenly distributed data is not suited to the statistical assumptions (normal distribution of variables across speakers) underpinning a random effects analysis.

[3] For full results of this multiple regression analysis with speaker as a random effect, including all factors not found to be significant, see Appendix D.

Subject Relatives *who* vs *who* + *that* **R2 =** 0.086 **input probability =** 0.463 **N** = 668 (human antecedents only)			
	FW	**%**	**N**
Community	p.<0.00		
Titirangi	0.62	55.4	168
Mt Roskill	0.47	43.0	249
Papatoetoe	0.40	34.3	251
Definiteness of the antecedent	p.<0.01		
Indefinite Pronoun	0.61	57.1	42
Indefinite	0.51	46.8	365
Definite	0.38	34.9	261
Role of the antecedent NP in the matrix clause	p.<0.01		
Other Object	0.59	48.5	163
Direct Object	0.55	45.6	272
Subject	0.50	38.6	145
Existential	0.36	30.7	88

Table 4.8b - Multiple regression analysis of factors contributing to the choice of *who* in human subject relatives. (Excluded from model as not significant: gender, age group, length of relative clause, complexity of relative clause.)

The most significant predictor for this alternation is community (p<0.001), *who* being favoured most in Titirangi, although note that this is only significant without speaker as a random effect. As with non-subject relatives, this community effect seems to be an artefact of individual variation. Only one factor remains significant whether speaker is modelled as a random effect or not. This is definiteness of the antecedent (NP-Matrix). As table 4.7 showed, this was also the strongest predictor for the alternation between Ø and *that* in non-subject relative clauses. Definite antecedents disfavour *who*, indefinites are almost neutral and indefinite pronoun antecedents favour *who.* The final significant predictor of *who* in subject relatives is the role of the antecedent in the matrix clause. When speaker is run as a random effect, the role of NP-Matrix stops being significant but remains on the edge of significance (p=0.054).

4.4 Constraints on Variation

In this section I consider the effects of different constraints on the variation in more detail. I will start with an exploration of the syntactic factors that appear to condition the alternation of *that* and *Ø* in non-subject relatives and the alternation of *who* and *that* in subject relatives with human antecedents. The purpose of this is to explore possible reasons for the patterns of variation, that is, why particular syntactic properties of the relative clause or its antecedent (NP-Matrix) might favour one relativiser over another. Further this allows us to compare our data with that from other studies of relative clause variation. I then move on to an examination of the social constraints on the variation, once again to compare this data with other studies as well as to explore what, if any, socially motivated variation and change is present for relative clauses in Auckland English.

4.4.1 Syntactic Factors

For non-subject relatives, Table 4.7 showed clause length is a significant factor in the choice of relative marker. This is not surprising when the two forms in alternation are *that* and *Ø.* In this case, the speaker is making a choice between using an overt or a non-overt relativiser. Hence one would expect longer clauses to favour overt relativisers "to offset processing difficulties" (Levey 2014: 27). An overt relativiser explicitly sets apart the relative clause from the NP-Matrix. Cofer (1975) claims this clarifies the structure, making the information easier for the listener to process.

Although clause complexity is not a significant predictor in this model, complexity is linked to clause length and may also need to be investigated. Indeed, other studies (Tagliamonte et al. 2005) code complexity and length together, i.e. simple-short, simple-long, complex-long. Table 4.9 shows how *Ø* and *that* relative markers are distributed in non-subject relatives.

	Ø		that	
	N	%	N	%
simple short	248	48.5%	263	51.5%
simple long	164	38.2%	265	61.8%
complex long	39	31.5%	85	68.5%

Table 4.9 – Distribution of Ø and *that* in non-subject relatives when clause length and complexity are combined.

Simple, short clauses have the highest percentage of Ø relatives, and complex long clauses have the lowest percentage of Ø relatives. Given that our results show clause length as a significant predictor of relativiser choice but clause complexity as not significant, it would suggest that it is the length of the clauses that are more responsible for this distribution.

Moving on to definiteness, figure 4.2 shows the proportion of Ø versus *that* in non- subject relatives.

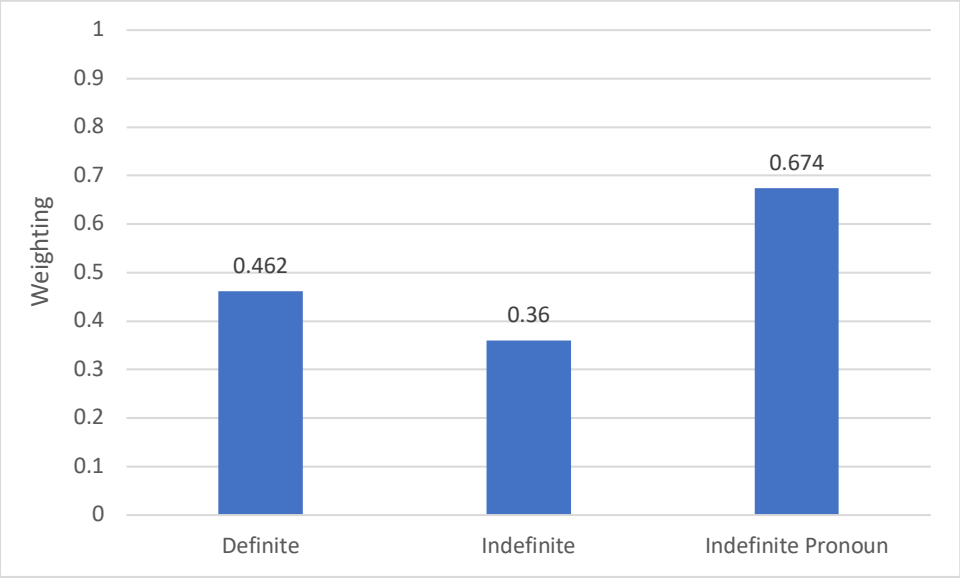

Figure 4.2 – Probability of the use of Ø in non-subject relative clauses according to the definiteness of the antecedent

Focussing first on the definite and indefinite NPs, it is clear that both constructions favour *that* but indefinite NPs favour it more. This conforms to previous literature showing a correlation between overt relativisers and indefinite antecedents (Huddleston 1971). However, Tagliamonte et al.'s (2005) UK English data found no significant effect for

definiteness for the choice of Ø in non-subject relatives. Levey's (2014) Canadian data (for the adult speakers), showed a very slight favouring effect for Ø with pronominal antecedents. Levey's data show both definite and indefinite antecedents disfavouring Ø but he has a fourth category, not usually coded, of unique antecedents which strongly favour Ø. Tottie & Harvie's (2000:214) African American English data show pronominal antecedents most favouring Ø relatives, as is the case in our data.

Just as I accounted for the effect of clause length in terms of processing factors, I will suggest that the dispreference for Ø relativisers with indefinite NPs is due to cognitive constraints. Indefinite NPs are, by definition, introducing new and less specific referents; "an apple" versus "the apple". The job of the relative clause is to define an antecedent, and in the case of indefinites, this defining load is higher, thus the need for an overt marker.

What then can be said of indefinite pronouns favouring Ø? One possible explanation comes from Thompson and Mulac (1991) who argue that the optional deletion of complementiser *that* is conditioned by how distinct the complement clause is from the "main" clause. They argue that some of these constructions are being "reanalyzed as unitary epistemic phrases" (Thompson and Mulac 1991: 249) and are thus more likely to have a Ø relativiser. This argument can be extended to the indefinite pronouns in the corpus. Indefinite pronouns seem to be more likely to be part of stock phrases such as "somebody I met" or "anyone you like". These would be prime targets for reanalysis as unitary epistemic phrases and thus less likely to have an overt relativiser. While it is beyond the scope of this thesis to undertake a quantitative analysis of these (more or less) fixed phrases, future research might want to consider analysing constructions like these separately.

Focusing still on definiteness but moving to subject relatives with human antecedents, figure 4.3 shows the rate of use of *who* in human subject relatives. Here the probability of use of *who* increases as the definiteness of the antecedent decreases. Remembering that these are exclusively human antecedents, in the left-hand column are constructions like "the man" or "my sister". In the middle, "a woman" or "some guy" and on the right, "someone" or "anybody".

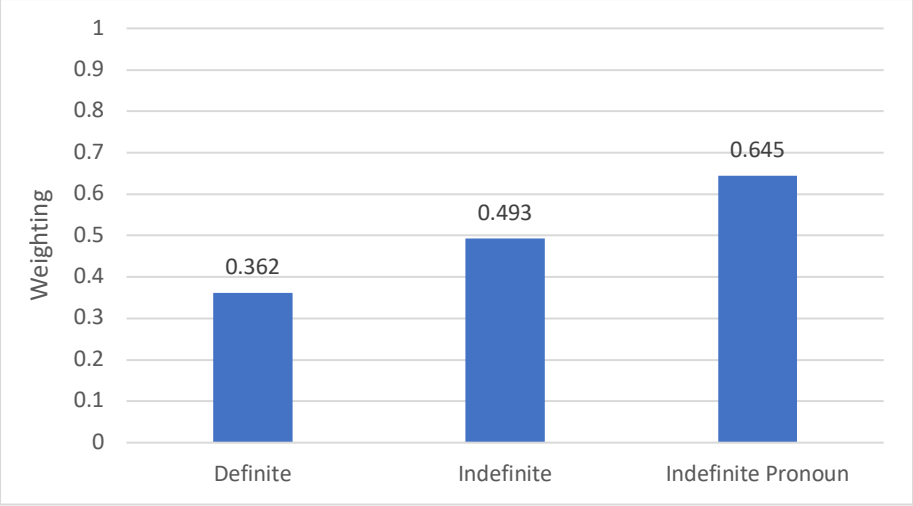

Figure 4.3 – Probability of *who* as a complementiser in subject relatives with human antecedent according to the definiteness of the antecedent

These results do not tally with many previous studies. Those studies that do look at the alternation between *that* and *wh* with regards to the definiteness of the antecedent (rather than just the alternation of overt and non-overt relativisers) seem to find extremely heterogeneous patterns across their sub corpora (Kautzsch 2002, Levey 2014). The only significant results Levey finds are in his youngest group of speakers who favour *who* with definite NPs and disfavour it with indefinites. The adults in his study have an almost neutral effect for both definite and indefinite antecedents.

There are, however, some parallels with Sigley's (1997) study of New Zealand English. Sigley does not code for definiteness in the same way as this study but does code the antecedent premodifier (determiners and quantifiers). He shows a result that is comparable to what this study finds for *who* in subject relatives, where more explicit (or definite) premodifiers favour less definite relative markers (Sigley 1997:305). Sigley argues that relative markers are on something of a heirarchy of explicitness; "(least explicit) Ø < *that* < *which* < case-marked *wh*-relatives (most explicit)" (Sigley 1997: 352). Determiners and other markers of definiteness (i.e. pronouns) also correlate with explicitness. "My book" is more explicit than "a book". According to Sigley, we should expect *who* which is more explicit than *that* to disfavour more definite antecedents which is precisely what can be observed in figure 4.3.

The final syntactic factor that is shown to be a significant factor in the choice of relative markers in our data is the role of the relativised element within the matrix clause. For both subject and non-subject relatives, the role of the NP-Matrix is only significant in models run without speaker as a random effect. The coding for the structure of the matrix clause varies across different studies. Guy and Bayley (1995) code this as subject, direct object and other but find no significant effect for this factor. Tagliamonte et al. (2005) code NP-Matrix as existentials, possessives, clefts and other. They do not model the variation of *who* in subject relatives. For Ø in non-subject relatives, Tagliamonte et al. (2005) find that only two of the three UK communities they studied show a significant effect for matrix clause structure. For these communities, existential and cleft constructions tend to favour Ø relatives while possessive constructions disfavour Ø relatives. Levey (2014) finds no significant effect for matrix clause structure for the adults in his data.

As there is little from previous studies to help us explain the effects seen for the role of the relativised NP in the matrix clause, alternative explanations for this are needed. One possibility here is to look at these effects in light of Fox and Thompson's (1990) discourse-based account of the relationship between the grammatical role of the NP-matrix and the NP-Rel as reviewed in section 3.2.1. I will defer the discussion of this to chapter 5.

4.4.2 Social Factors

Gender is the only social factor that comes through as significant for non-subject relatives. The favouring effect here is small with men slightly favouring Ø and women slightly disfavouring it. This is reminiscent of the effects seen in Sigley (1997) where Pākehā men used more Ø than other speakers (see section 2.3.2). Both Levey (2014) and D'Arcy and Tagliamonte (2010) find no gender effect for the use of Ø relativisers. However they both find a small correlation between *who* (in human subject relatives) and female speech.

D'Arcy and Tagliamonte (2010) link this preference to the perceived formality of the *wh*-forms. They support this explanation by showing that *who* has the n-shaped age distribution pattern of a prestige variant of a stable variable and thus could be expected to be used more by women (Labov 1990, 1994, 2001). Gender and education interact in a way that is

also consistent with the inference that *wh-* forms have some prestige association. Although there is no gender effect for their corpus as a whole, when divided by education, speakers with post-secondary education have the highest rates of *who* as a relative marker in subject relatives. Among them, women lead men.

If I remove speaker as a random effect, community is a significant predictor for both subject and non-subject relatives (p<0.00 for both). As discussed in section 4.3.2 there is good reason for doing this when there is so much inter-speaker variation, especially in the case of subject relatives where there are a number of categorical users of one or other of the variants.

Figure 4.4 shows the rate of use of *who* in subject relatives by community and we can clearly see that *who* is favoured by speakers in Titirangi, the community with the higher socio-economic demographic. In the ethnically and economically mixed Mt Roskill there is no clear effect one way or another on the use of *who* and *who* is disfavoured in the lower socio-economic area of Papatoetoe. This conforms to what might be expected for a variant often associated with standardness (D'Arcy and Tagliamonte 2010, Tagliamonte et al. 2005).

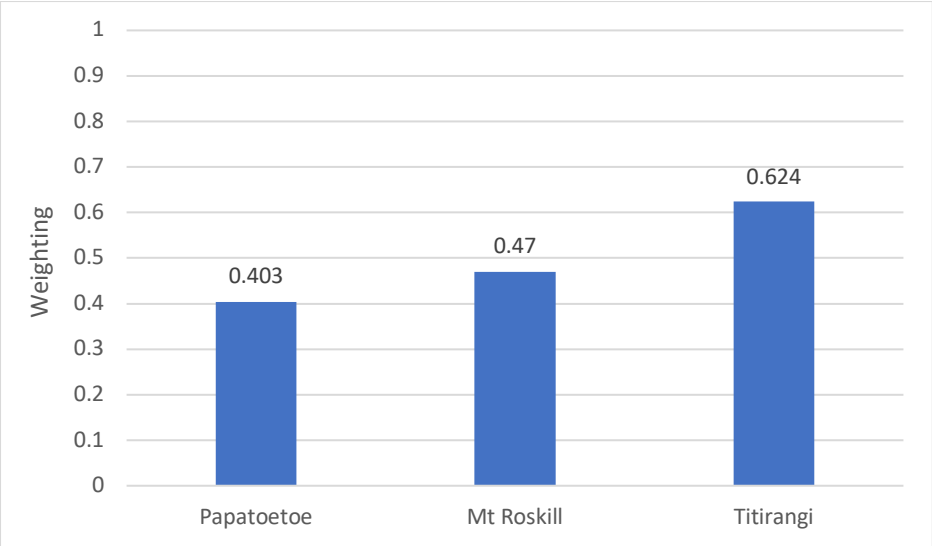

Figure 4.4– Probability of *who* as a complementiser in subject relatives (human antecedents) by community

For non-subject relatives there were only 2 categorical users (one each for *that* and Ø), but among the remaining speakers, the distribution of *that* versus Ø also does not approximate a normal distribution. This warrants an analysis of the data in which speaker is not modelled as a random effect. In this event, community does emerge as a significant constraint as shown in table 4.7b. Figure 4.5 shows the rate of use of Ø in non-subject relatives by community. Speakers in Mt Roskill favour Ø while speakers in Titirangi and Papatoetoe slightly disfavour it.

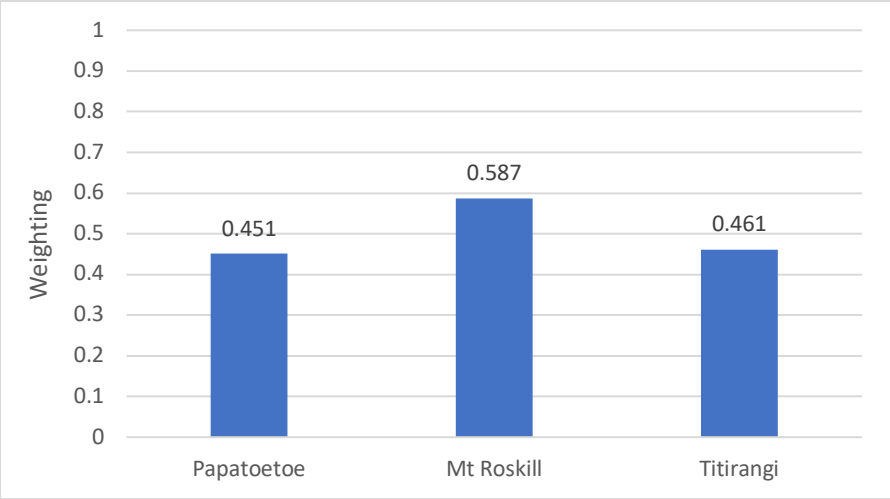

Figure 4.5 – Probability of Ø as a complementiser in non-subject relatives by community

It is still unclear why speakers in Mt Roskill are showing this preference for *Ø.* There is no obvious prestige variant in the alternation of *that* versus Ø in non-subject relatives. According to our hypothesis, based on the theory that ethnolinguistically diverse communities level towards a single, unmarked variant (Cheshire et al. 2011), one would expect communities like Papatoetoe and Mt Roskill to favour *that.* This is because *that* can be used in all relative clause types while Ø is grammatically constrained by the prescription against Ø subject relatives. I will leave the discussion of possible explanations for this unexpected result until chapter 5.

The final social factor coded for was age group. This did not approach significance in any of the models run. On the face of it this would suggest stable variation with no age grading, however further investigation of the data suggests that this is not the end of the story.

Figure 4.6 shows the distribution of *Ø, that,* and *wh-* relative markers in subject and non-subject relatives grouped by age, gender and community.

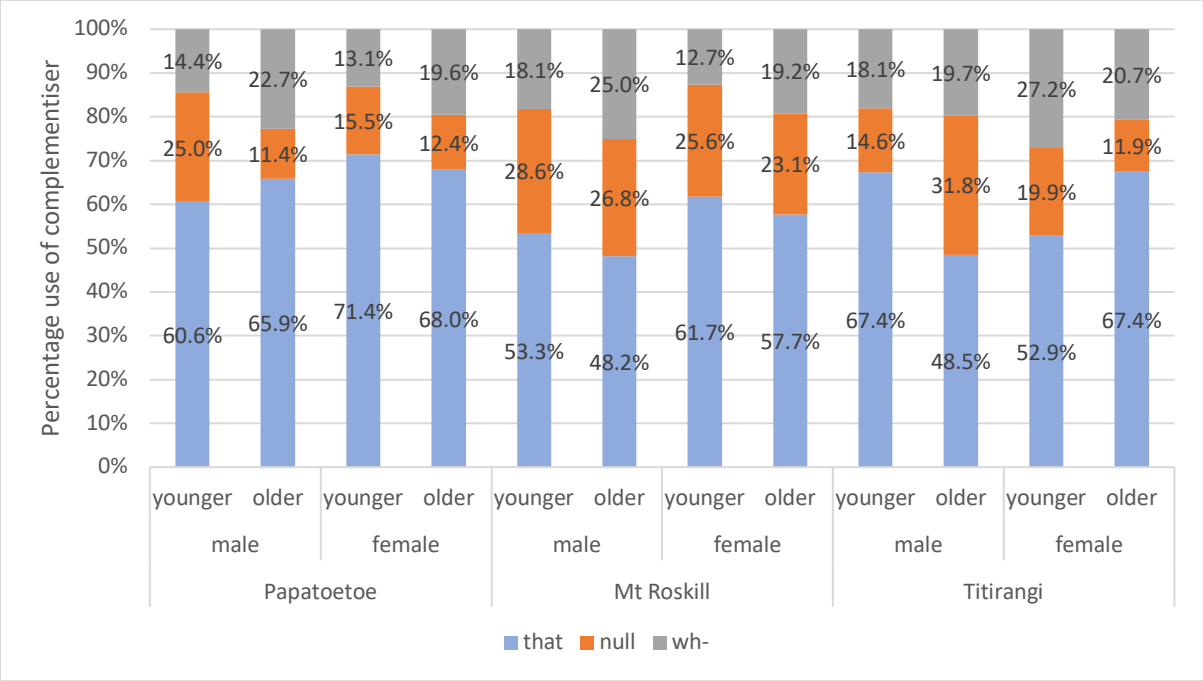

Figure 4.6 – Percentage of *that, Ø* and *wh-* (*who* and *which*) in subject and non-subject relatives grouped by community, gender and age (younger group is under 25 and older is over 40).

The first notable feature in figure 4.6 is that in every group (except for the Titirangi women), younger speakers use *wh-* relativisers less than their older counterparts. This might suggest an across time decline in the use of *wh-* relative markers. Hinrichs, Szmrecsanyi and Bohmann (2015), show a marked shift from *which* to *that* in written corpora of standard English from 1961 to 1992. However, few of the spoken English studies show evidence of this. Levey (2014) explores acquisition of variation and therefore does not consider possible apparent time language change in his data. Tagliamonte et al. (2005) find such low rates of *wh-* in their data that they focus primarily on *that* and *Ø.* D'Arcy and Tagliamonte (2010) argue that their data shows *wh-* relatives to be stable, age-graded variation.

The second point to note is that older and younger Titirangi males have dramatically different rates of use of *Ø* relativisers. The rate of use of *Ø* by the older Titirangi men is the highest of the whole data set, higher even than that of the Mt Roskill men, while younger

Titirangi men use Ø relativisers infrequently. Any generalising about these groups must be done with great caution however, as some contain too few participants to constitute a representative sample.

The issue with breaking the speaker data down into smaller and smaller subgroups is that individual speaker differences have an increasing effect on the results for each group. A prime example of this is the Titirangi older male group. Unfortunately, the data available for analysis at the start of 2019 only included two speakers in this group. Of them, one speaker, Matt, produced 21 Ø relatives, much higher than the average number of Ø relatives produced across all speakers in the corpus (7.9, median 6). He also unusually (although not uniquely) produced more tokens of Ø as a relativiser than *that*, (N=17). This being the case, and given that there is only have one other older Titirangi male in the sample at the time of writing, it is hard to make any general comment about the use of Ø by older Titirangi men.

There is a similar issue when exploring the higher rates of *wh*- use by younger Titirangi women. The sample size for the Titirangi women is more balanced (4 younger, 8 older), but there is an outlier in the younger speakers. Caitlyn produces 27 tokens of *who*, the highest number by far for the whole dataset and well above the average (4.8, median 4). This exceptionally high use of *wh*- suggested it might be informative to exclude this speaker from the model that looked at the use of *who* and *that* in human subject relatives in case this outlier was affecting the result. However, running the model again without Caitlyn made no difference to which variables were significant predictors and very little difference to the factor weights.

One frequently observed age effect is the “v-shaped” formality curve (Downes 1998). This shows middle aged speakers using less of a non-standard or vernacular variant than both older and younger speakers. The explanation for this relates largely to societal pressure. Younger speakers, especially in adolescence may use more of a vernacular feature to conform to their peer group, while middle aged speakers are under societal pressure in the job market to be more linguistically conservative. This pressure then decreases after retirement allowing older speakers to return to the vernacular. Depending on which variant is used as the application value of the rule, this v-shaped curve can appear as an n-shaped

curve as shown by D'Arcy and Tagliamonte (2010: 393). If the variant has prestige or standardness associations the curve will be n-shaped but if the variant is non-standard the curve will be v-shaped.

A problem with the structure of the data collected in the Auckland study is that it is partitioned in two very broad age groups: 25 and under and 40 and over. This will not reveal any possible age grading for which at least three age groups are needed. If I redistribute the speakers for this study into three rather than two age groups, I can explore whether this effect occurs for relative markers in Auckland English.

To determine how best to redistribute the age groups, it is necessary to look more closely at the composition of the age groups in this study. The younger age group is made up of under 25 year olds, many of whom are still in education. Very few of those who are in work are in what would be termed a "white-collar" workplace environment. For this reason, it seems justifiable to keep the younger group as one unit and split the older group into middle (45-59) – those who are likely to still be active in the workplace, and older (over 65) – those who are above the retirement age. The three speakers who were aged between 60 and 65 were excluded[4].

Figure 4.7 shows the rate of use of each of the main relative markers (*wh-* is both *who* and *which*) in both subject and non-subject relative clauses after undertaking this regrouping of speakers. This gives us a rough overview of the data in much the same way as figure 4.1 did for gender and community.

[4] These three speakers comprised two women from Mt Roskill and one from Titirangi.

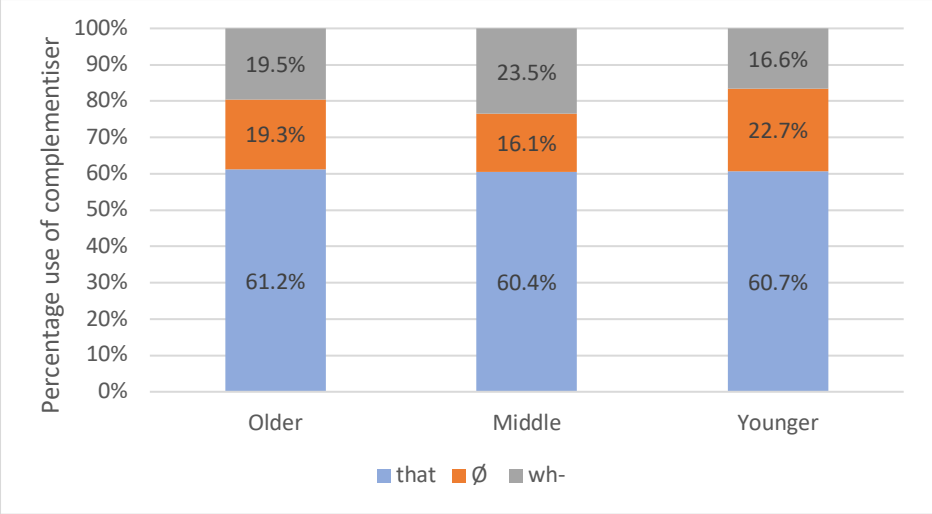

Figure 4.7 - Percentage of *that, Ø* and *wh-* (*who* and *which*) in subject and non-subject relatives by age group.

While all three age groups have similar rates of use of *that,* there are differences in their use of *wh-* and *Ø.* We know already that *Ø* and *wh-* have complementary distributions, with *Ø* used predominantly in non-subject relatives and *wh-* more frequently in subject relatives. If, for example, the speakers in the middle age group are producing more subject relative clauses and the younger speakers more non-subject relative clauses, that might influence the different rates of use seen in figure 4.9. In actual fact, however, there is very little difference between the middle and younger groups of speakers in the types of relative clause they produce. The younger speakers produced 47% (N=603) subject relative clauses to 53% (N=681) non-subject relative clauses while the middle aged speakers produced 48.5% (N=184) subject relative clauses and 51.5% (195) non-subject relative clauses. The outliers here were the older speakers who produced 58.9% (308) subject relative clauses and 41.1% (215) non-subject relative clauses.

To investigate whether this distribution across three age groups is in fact significant, I need to run the variable rule analyses again to see if age emerges as significant. Due to the previously discussed non-normal distribution of tokens across speakers, these runs were performed without speaker as a random effect.

For the use of *Ø* over *that* in non-subject relative clauses, this reassignment of speakers into three age groups proves to be extremely productive. When I analysed the data with two age

groups, age did not come close to significance (p=0.552), but when I model the variation with speakers divided into 3 age groups, age comes very close to being a significant predictor (p=0.0551).

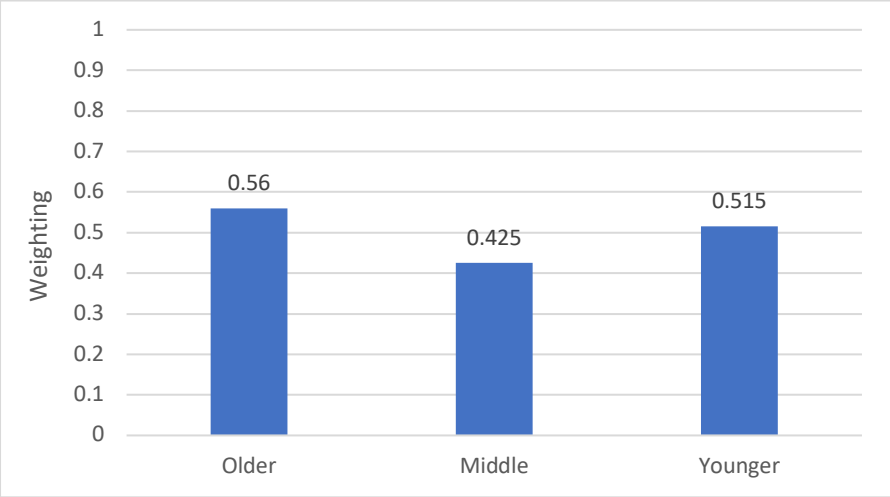

Figure 4.8 – Probability of Ø as a complementiser in non-subject relatives by three age groups

Figure 4.8 shows the factor weights for the use of Ø in non-subject relative clauses. Older speakers slightly favour Ø while middle aged speakers disfavour Ø. Younger speakers have a neutral effect for this variable. This does suggest a weak form of age grading with speakers in the 45-59 age group disfavouring Ø relatives.

For the use of *who* over *that* in subject relative clauses with human antecedents, reassigning speakers into three age groups again proves to be productive. When I analysed the data with two age groups, age did not come close to significance (p=0.766), but when I model the variation with speakers divided into 3 age groups, age comes very close to being a significant predictor (p=0.0654)

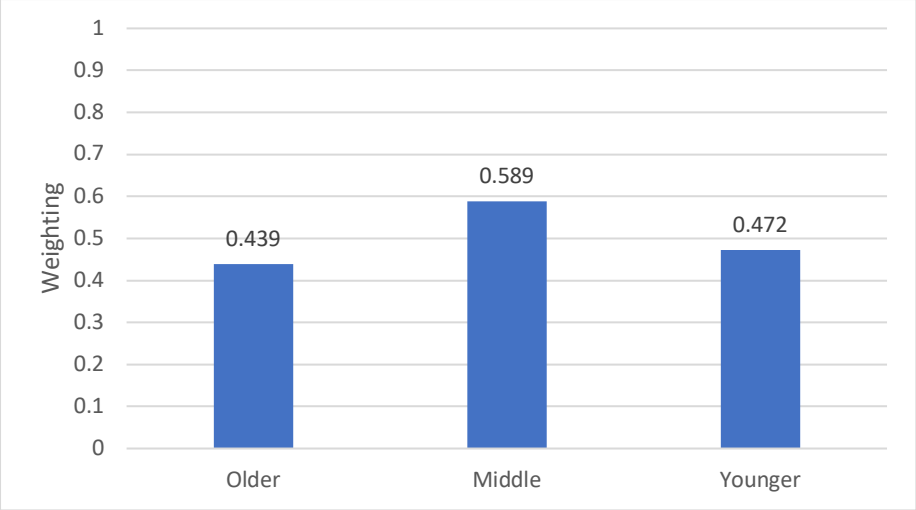

Figure 4.9 – Probability of *who* as a complementiser in subject relatives with human antecedents by three age groups

Figure 4.9 shows the probability of speakers choosing *who* in human subject relatives. Younger speakers, slightly disfavouring who. Older speakers disfavour *who* more while middle aged speakers favour *who*.

In both subject and non-subject relative clauses, there is a pattern that resembles the previously mentioned v-shaped (or the reversed n-shaped) curve typical of age-grading (D'Arcy and Tagliamonte 2010) with either a peak or a dip in usage of a particular variant around middle age. The peak in use of *who* around middle age conforms to the proposition that it is perceived to, in some way, be more standard or prestigious. (D'Arcy and Tagliamonte 2010, Levey 2014, Tagliamonte et al. 2005). In contrast, the dip in use of Ø around middle age suggests it is negatively correlated with standardness. Tagliamonte et al. (2005) make this observation about Ø relativisers in subject relatives but this analysis suggests the argument can be extended to non-subject relative clauses in Auckland English.

Sigley (1997: 436) found a similar age-grading pattern in his data. For his speakers he found that the middle aged group used more *who* in personal subject relatives and more overt relativisers in non-subject relatives than the younger and older groups. This is directly in line with what is found for Auckland speakers. This would appear to be evidence that the New Zealand English relative marker system is not only stable and age graded but has been so for some time, given that Sigley's data is over two decades older than this study.

4.5 Interaction effects

This corpus was designed with a particular interest in the interaction between community and age. Given this, it was important to consider possible interaction effects as part of our models. However, when modelling interaction effects between community and age, either with two or three age groups, these interaction groups never emerge as having a significant effect for Ø relativisers in non-subject relatives. Neither does community and gender when run as an interaction group.

For *who* in subject relatives, community and gender does not emerge as significant when run as an interaction group. Nor does community and three age groups when run as an interaction group. However, when I run community and two age groups as an interaction group, a significant effect does emerge for *who* in subject relatives.

Below, I present the results of the variable rule analyses with community and two age groups as an interaction effect for both Ø in non-subject relatives (table 4.10) and *who* in subject relatives with human antecedents (table 4.11).

Non-subject Relatives Ø vs Ø + *that* **R2 =** 0.105 **input probability =** 0.523 **N=** 1064			
	FW	**%**	**Total N**
Definiteness of the antecedent	p.<0.0000		
Indefinite Pronoun	0.67	61.5	104
Definite	0.46	46.2	556
Indefinite	0.36	32.2	404
Gender	p.<0.01		
Male	0.58	49.3	420
Female	0.42	37.9	644
Role of the antecedent NP in the matrix clause	p.<0.05		
Subject	0.57	51.8	276
Existential	0.49	34.1	88
Other Object	0.47	40.9	274
Direct Object	0.46	39.0	426
Length of relative clause	p.<0.05		
Short	0.54	48.4	512
Long	0.46	36.8	552
Community and age interaction group	p.=0.109		
Papatoetoe younger	[0.55]	41.7	192
Mt Roskill older	[0.54]	56.3	142
Titirangi older	[0.51]	33.5	158
Titirangi younger	[0.49]	38.0	187
Mt Roskill younger	[0.46]	49.6	272
Papatoetoe older	[0.45]	28.3	113

Table 4.10 - Multiple regression analysis of factors contributing to the choice of Ø in non-subject relatives. (Interaction effect: community and age group. Not selected as significant: complexity of relative clause, animacy of antecedent.)

Subject Relatives *who* vs *who* + *that* **R2 =** 0.104 **input probability =** 0.481 **N=** 668 (human antecedents only)			
	FW	**%**	**N**
Definiteness of the antecedent	p.<0.01		
Indefinite Pronoun	0.62	57.1	42
Indefinite	0.51	46.8	365
Definite	0.37	34.9	261
Community and age interaction group	p.<0.01		
Titirangi younger	0.60	64.5	76
Mt Roskill older	0.57	51.5	103
Papatoetoe older	0.53	35.7	115
Papatoetoe younger	0.47	33.1	136
Mt Roskill younger	0.43	37.0	146
Titirangi older	0.40	47.8	92
Role of the antecedent NP in the matrix clause	p.<0.01		
Other Object	0.59	48.5	163
Direct Object	0.55	45.6	272
Subject	0.50	38.6	145
Existential	0.36	30.7	88

Table 4.11 - Multiple regression analysis of factors contributing to the choice of *who* in human subject relatives. (Interaction effect: community and age group. Excluded from model as not significant: gender, length of relative clause, complexity of relative clause.)

In these models, the factor weights returned are inconsistent with the percentage rates of the variables. This suggest that there is another interaction at play here with something that I have not coded for. There are a number of social factors that have not been coded for in this corpus. Factors such as level of education, social network and profession could be interacting with community and age and influencing the distribution of relativisers. In the present study, community was used as a proxy for level of education/profession, however clearly these would be useful directions for future research. Despite the inconsistencies between factor weights and percentages, the significance of the community and age interaction group in subject relatives warrants further examination.

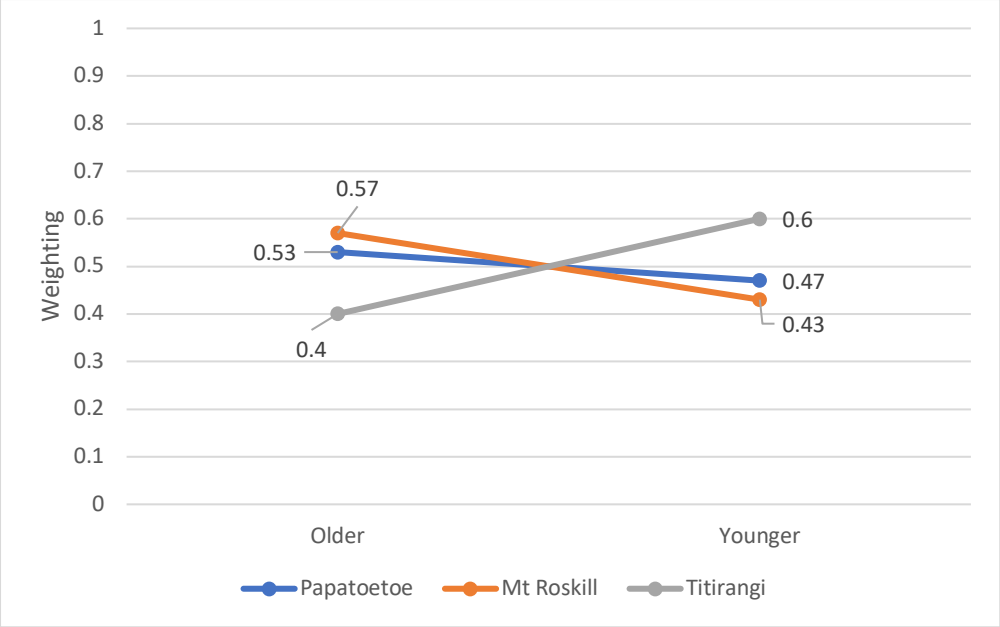

Figure 4.10 – Probability of *who* as a complementiser in subject relatives with human antecedents according to age group and gender.

Figure 4.10 shows that the greatest difference in probability of a speaker using a *who* as a relativiser is between the older and younger Titirangi speakers. Titirangi speakers also show the opposite age effect to Mt Roskill and Papatoetoe speakers. The direction of effect for Mt Roskill and Papatoetoe where older speakers favour *who* and younger speakers disfavour *who* is what I might expect to see given the hypothesis that language change in Auckland would move away from marked, lower frequency variants like *who.* There is no immediately obvious explanation for why older speakers in Titirangi should disfavour *who* and younger speakers favour *who.* Given the inconsistency between the factor weight and percentage rates of the variant it is necessary to be cautious about extrapolating from these results.

The only interaction group that emerged as a significant predictor for Ø in non-subject relatives was one comprising the three age groups and gender (table 4.12). This interaction group did not emerge as significant for *who* in subject relatives. While there is still a discrepancy between the factor weights and the percentages suggesting some other factor is at play here, it is nevertheless worth examining how these two factors interrelate in order to better understand the grammar of relative clauses in Auckland English.

Non-subject Relatives Ø vs Ø + *that* **R2 =** 0.125 **input probability =** 0.506 **N=** 1035			
	FW	**%**	**Total N**
Definiteness of the antecedent	p.<0.0000		
Indefinite Pronoun	0.68	62.7	102
Definite	0.45	45.9	538
Indefinite	0.37	32.4	395
Community	p.<0.001		
Mt Roskill	0.60	52.2	395
Titirangi	0.45	36.1	335
Papatoetoe	0.44	36.7	305
Gender and age interaction group	p.<0.01		
Female middle	0.61	34.8	135
Male older	0.61	59.6	109
Male younger	0.50	47.1	261
Female younger	0.50	41.8	390
Female older	0.39	28.6	98
Male middle	0.39	31	42
Role of the antecedent NP in the matrix clause	p.<0.05		
Subject	0.58	51.9	266
Existential	0.49	34.1	85
Other Object	0.48	40.5	269
Direct Object	0.45	39.3	415
Length of relative clause	p.<0.05		
Short	0.54	48.3	499
Long	0.46	36.9	536

Table 4.12 - Multiple regression analysis of factors contributing to the choice of Ø in non-subject relatives. (Interaction effect: gender and age group. Not selected as significant: complexity of relative clause, animacy of antecedent.)

Figure 4.11 shows the probability of Ø in non-subject relatives. The v-shaped distribution seen earlier for Ø appears here only for the men. This suggests that the effect seen in figure 4.8 is influenced mainly by the male speakers' behaviour. What is apparent is that there is a shift from more gendered variation in relativiser choice for older speakers to a much less varied system for younger speakers.

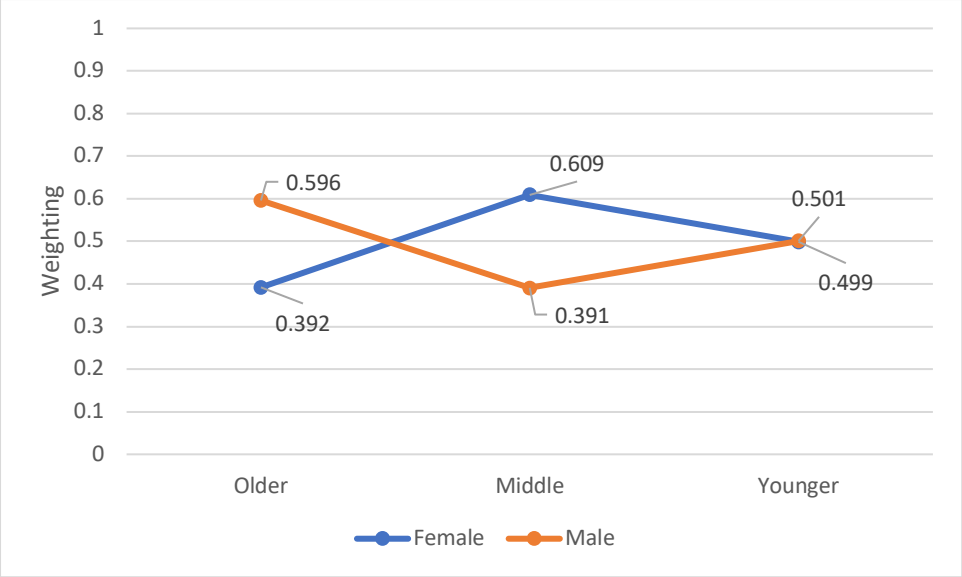

Figure 4.11 – Probability of Ø as a complementiser in non-subject relatives according to age group and gender

There appears to be little evidence of a change in progress for Ø in non-subject relatives in Auckland English. There may be some change occurring for *who* in subject relatives. These results should be treated with caution however, given the discrepancy between factor weights and percentages and the likelihood that some other social factor is interacting in some way.

4.6 Inter-community variation

As previously discussed, community is a highly significant predictor of relativiser choice ($p<0.00$) for both subject and non-subject relatives in models where speaker is not included as a random effect. This warrants looking at the intra-community variation to see if there are any constraints on the distribution of relative markers that are specific to any of the communities.

Table 4.13 shows the variable rule analysis for the choice of Ø in non-subject relatives. These runs were all conducted with speaker as a random effect. One of the first things points to notice is that definiteness of the antecedent, the most significant conditioning factor for this variable in the corpus as a whole, is also significant for two of the three communities. Only in Papatoetoe is it not significant, although it is approaching significance (p=0.09). The direction of the effect is also the same in all three communities. This suggests that how you orient to the definiteness of the NP in the matrix in determining whether to use Ø or *that* in a non-subject relative clause is what marks you as a speaker of "Auckland English".

Ø* vs *Ø* + *that	**Papatoetoe**			**Mt Roskill**			**Titirangi**		
Total N	305			414			345		
R2	0.24			0.31			0.208		
Input probability	0.443			0.505			0.407		
	FW	%	total N	FW	%	total N	FW	%	total N
Definiteness of the antecedent	p=0.09			p<0.01			p<0.05		
Indefinite Pronoun	[0.67]	63.6	22	0.71	71.4	42	0.63	50	40
Definite	[0.43]	35.8	151	0.45	53.8	253	0.51	44.1	152
Indefinite	[0.40]	33.3	132	0.33	41.2	119	0.36	24.2	153
Length of relative clause	p=0.147			p=0.177			p=0.412		
Short	[0.55]	42.3	137	[0.54]	57.8	230	[0.53]	39.3	145
Long	[0.45]	32.1	168	[0.46]	44.6	184	[0.47]	33.5	200
Gender	p=0.592			p<0.01			p=0.603		
Male	[0.53]	40.1	142	0.64	61.4	176	[0.53]	41.2	102
Female	[0.47]	33.7	163	0.36	45	238	[0.47]	33.7	243
Role of the antecedent NP in the matrix clause	p=0.381			p=0.979			p<0.01		
Subject	[0.55]	43.5	85	[0.52]	55.3	123	0.69	55.9	68
Direct object	[0.41]	28.8	118	[0.50]	53.3	150	0.45	32.9	158
Existential	[0.55]	44	25	[0.50]	40.9	22	0.43	24.4	41
Other object	[0.49]	39	77	[0.49]	48.7	119	0.43	30.8	78
Complexity of the relative clause	p=0.39			p<0.05			p=0.975		
Simple	[0.45]	37.2	269	0.62	54.6	368	[0.50]	36.6	303
Complex	[0.55]	33.3	36	0.39	30.4	46	[0.50]	31	42
Age group	p=0.0654			p=0.376			p=0.732		
Younger	[0.61]	41.7	192	[0.45]	49.6	272	[0.52]	38	187
Older	[0.39]	28.3	113	[0.55]	56.3	142	[0.48]	33.5	158
Animacy of the antecedent	p=0.912			p=0.971			p=0.868		
Inanimate	[0.50]	35.8	218	[0.50]	52.7	357	[0.50]	35.3	286
Human + Animate*	[0.50]	39.1	87	[0.50]	47.4	57	[0.50]	39	59

Table 4.13 – Three separate variable analyses of factors contributing to the choice of *Ø* in non-subject relatives. Results in square brackets are not significant. *Human and other animate have been recoded together due to low number of tokens of other animate (N=2, 7 and 8 respectively)

Papatoetoe lacks any significant constraints on relativiser choice. This could be because the variation is due to some other factor that has not been investigated or the amount of data is insufficient to get a clear signal from the noise. Although given that there are a similar number of tokens for Papatoetoe and Titirangi, this seems unlikely. For all three factors that are significant in the data overall – definiteness of the antecedent, gender and clause length – Papatoetoe has the same direction of effect as the other two communities. In other words, it seems that Papatoetoe contributes to the significance of definiteness, gender and clause length in the overall dataset. The power of combined data is most clear in the case of clause length for which no individual community shows a significant effect but overall in the aggregated data this is significant.

In Mt Roskill, while clause length isn't significant, clause complexity is. Since length and complexity are linked – all complex clauses are long, all short clauses are simple – it is possible that some communities, in this case Mt Roskill, make use of complexity more than length to determine relativiser choice. Mt Roskill is also the only community that shows a significant gender effect. This makes Mt Roskill the community with the largest number of constraints on the choice of Ø non-subject relatives. Mt Roskill is also the community with the highest rate of use of Ø in the corpus overall, so it may just be that we learn more when we have more data points to work with.

The role of the antecedent NP in the matrix clause is a significant constraint only for Titirangi. Subject position NPs prefer Ø relatives while all other constructions disprefer *Ø.* Although not significant, Mt Roskill has the same direction of effect, Papatoetoe however does not. As previously discussed, the structure of the matrix clause shows inconsistent effects across different studies of relative clause variation.

Table 4.14 shows the variable rule analyses for the choice of *who* in subject relatives. The only community that shows any significant constraints for this variable is Titirangi. Both the definiteness of the antecedent NP and its role in the matrix clause are significant in this community as they were in non-subject relatives for Titirangi.

who* vs *who* + *that	**Papatoetoe**			**Mt Roskill**			**Titirangi**		
R2	0.453			0.464			0.352		
Input probability	0.289			0.417			0.608		
Total N	251			248			168		
	FW	%	total N	FW	%	total N	FW	%	total N
Definiteness of the antecedent	p=0.569			p=0.122			p<0.05		
Indefinite Pronoun	[0.60]	57.1	14	[0.64]	45.5	22	*	*	*
Indefinite	[0.48]	35.6	118	[0.50]	45.6	147	0.63	64.2	106
Definite	[0.42]	30.3	119	[0.36]	36.7	76	0.37	40.3	62
Length of relative clause	p=0.326			p=0.543			p=0.839		
Long	[0.55]	35.5	141	[0.43]	47.3	146	[0.51]	57.3	110
Short	[0.45]	32.7	110	[0.47]	36.3	102	[0.50]	51.7	58
Gender	p=0.635			p=0.646			p=0.339		
Male	[0.55]	35.6	104	[0.54]	47.4	135	[0.42]	44.2	43
Female	[0.45]	33.3	147	[0.46]	37.2	113	[0.58]	59.2	125
Role of the antecedent NP in the matrix clause	p=0.746			p=0.245			p<0.05		
Other	[0.51]	37.7	69	[0.61]	45.8	59	0.74	74.3	35
DO	[0.51]	32.3	93	[0.59]	47.8	115	0.52	60.3	63
Subj	[0.58]	36.2	58	[0.44]	35	40	0.5	44.7	47
Existential	[0.40]	29	31	[0.37]	29.4	34	0.25	34.8	23
Complexity of the relative clause	p=0.477			p=0.613			p<0.05		
Complex	[0.46]	31.6	57	[0.53]	50	40	0.68	75	24
Simple	[0.54]	35.1	194	[0.47]	41.3	208	0.32	52.1	144
Age group	p=0.756			p=0.636			p=0.115		
Older	[0.53]	35.7	115	[0.55]	51	102	[0.38]	47.8	92
Younger	[0.47]	33.1	136	[0.46]	37	146	[0.62]	64.5	76

Table 4.14 – Three separate variable analyses of factors contributing to the choice of *who* in subject relatives with human antecedents. Results in square brackets are not significant. *Indefinite pronouns have been included with indefinite NPs in Titirangi for this analysis because they occurred categorically with *who.*

The role of the NP matrix is a significant constraint in the data overall, but table 4.14 shows that the direction of the effect is quite different in Papatoetoe than the other two communities even though the grammatical role of the NP-Matrix is not a significant predictor on its own in Papatoetoe. One other factor emerges as significant in Titirangi and that is clause complexity. This is not significant in the data overall and neither is the related factor of clause length.

Titirangi is the community with the highest rate of *who* in the data overall and the only community showing significant constraints on the use of *who*. This is similar to the case of non-subject relatives where the community with the highest rate of Ø, Mt Roskill, also had the greatest number of significant constraints on its use.

4.7 Qualitative Analysis

There are some relative clause constructions for which there are too few tokens on which to perform any quantitative analysis. However, it is still worthwhile exploring these to see what might be learnt from a qualitative analysis. Below, I discuss the small number of subject relative clauses introduced by a Ø relative marker. Following this, I look at adverbial relative clauses which are canonically excluded from variationist studies of relative clauses. As part of this qualitative discussion I interrogate the reasons for their exclusion and whether they should in fact be included in future quantitative studies of relativiser variation.

4.7.1 Ø Subject Relatives

It was noted earlier that the corpus includes 14 examples of Ø subject relatives (even though these are generally considered ungrammatical in New Zealand English). In the case of these 14 Ø subject relatives for instance, I explore whether all can reasonably be interpreted and thus explained away as mere production errors. Disfluent or unclear speech may produce something that sounds like a Ø subject but may not be. This is illustrated in example 35:

35. “Once again because (of) this guy (**Ø**?) had been in America” – Chub – (Mt Roskill, older)

There is more than one possible interpretation of this utterance. The participant's speech at this point is slightly mumbled, making it difficult to determine whether the word *of* is present. If it is not, the phrase contains no relative clause at all. However, if the *of* is present, the utterance could either be a true Ø relative, or it could be the case that the speaker has changed his mind halfway through the sentence. He may initially have intended to say "because of this guy" and then changed to "because this guy had been to America" and the two utterances have been combined.

Not every example in the data is this ambiguous however. Tokens such as example 36 seem to be fairly unequivocally intended as Ø subject relative clauses.

> 36. "So, there's a few of us [**Ø** go in and listen to them read]" – June – (Papatoetoe, older)

Table 4.15 shows all fourteen tokens from the data, along with some demographic information about the speakers who produced them.

Speaker	Community	Age group	Gender	Context
Chub	Mt Roskill	Older	Male	Once again because of this guy Ø had been in America
Luke	Mt Roskill	Older*	Male	We- we had a lady at work Ø was ah, phoning up defence department round the country trying to get, um, material for props.
John	Mt Roskill	Older	Male	there's one there Ø just, keeps following the lawn mower (to get the worms).
Noor	Mt Roskill	Younger	Female	I'm not saying only Pākehās and people from Asian communities can afford it but there are Somalis Ø can too
Leletike	Mt Roskill	Younger	Male	you- on- the person Ø christen you gives you your name.
June	Papatoetoe	Older	Female	but these are guys Ø wouldn't know what was way was up.
June	Papatoetoe	Older	Female	So, there's a few of us Ø go in and listen to them read
Henry	Papatoetoe	Older	Male	so they called out all the children Ø were there and
Henry	Papatoetoe	Older	Male	There was one Ø was a (mantz) and it went back to
Andrew	Papatoetoe	Younger	Male	I got a few friends there Ø say it's alright but (I mean) they're single so I guess it's different.
Ian	Papatoetoe	Older	Male	heaps of people Ø came to the funeral had never seen it
Ian	Papatoetoe	Older	Male	and there's a wooden bridge Ø goes across which is w- and then there was metal the other side but
Max	Titirangi	Younger	Male	…even the people Ø live out there
Jack	Titirangi	Younger	Male	I have family Ø live there.

Table 4.15 - Ø subject relatives. * Note that the age group classifications in this table are listed as they were before the division into three age groups. After the redistribution, Luke was recoded as “middle”.

Five speakers each from Papatoetoe and Mt Roskill produce Ø subject relatives, with some speakers producing more than one. Only two speakers from Titirangi produce them, both young men. Aside from the previously mentioned token produced by Luke, there are two other tokens that are particularly ambiguous, namely Leletike and Ian. In the former, the speaker is being particularly disfluent at the time the utterance is produced, with multiple

pauses and repairs and the resulting phrase is unclear in the recording. It is possible that the word *who* is stumbled over rather than deliberately excluded before the word *christen*. In the case of the latter, the speaker pauses, both after *people* and *funeral*. The ambiguity here is partly because a complementiser could be in either of these positions, as demonstrated below;

37. a. heaps of people [**that** came to the funeral] had never seen it.
    b. heaps of people came to the funeral [**that** had never seen it].

Both of these interpretations make sense in the wider spoken context.

Even setting aside these more ambiguous tokens, that still leaves us with 11 tokens of Ø subject relatives. Papatoetoe speakers produce the greatest number of tokens and all but one of these are produced by older speakers. The appearance of Ø subject relative clauses most in Papatoetoe and least in Titirangi would tend to agree with D'Arcy and Tagliamonte's (2010: 385) assertion that there is a non-standardness to this construction. There are other indications that Titirangi speakers converge on more standard-like forms (eg. using the highest rate of *who* in human subject relatives.)

There is little evidence of language contact as the motivation for these forms. The majority of tokens are produced by monolingual, Pākehā New Zealanders so it is unlikely that the use of Ø subject relative clauses in English is the result of grammatical interference from another language. Many of the speakers producing them do not have another language in their repertoire or their home background.

4.7.2 Adverbial relative clauses

Relative clauses introduced by *when* and *where*, and indeed those introduced by *that* and Ø where the *wh*- alternation would be with *when* or *where* rather than *who* or *which,* are generally excluded from variationist studies of relative clauses. As previously discussed, most relative markers have complementary restrictions on their use. As discussed in chapter 2, *wh*- relatives, being pronouns, encode information about their referents which influences

this complementary distribution by restricting the antecedents *wh-* relatives can relativise. To illustrate this, and for convenience I repeat the examples from chapter 2 here:

38. a. “Y'know and that's like the society [(which/**that**/Ø/*who) we're like living in].” – Aarush – (Mt Roskill, younger)
b. “We’ve got people [(who/that/**Ø**/*which) we go camping with each year]” – Charlotte – (Titirangi, older)
c. “There were, there were moments [(**when**/that/Ø/*who/*which) you thought, this, this, this could go badly].” – Nicholas – (Mt Roskill, older)
d. “It was just a place [(**where**/that/Ø/*who/*which) I could afford to go]” – Mary – (Mt Roskill, older)

Examples 38 c and d illustrate the canonical associations for the two adverbial pronouns, that is, use of *when* with a temporal antecedent and *where* with a locative antecedent. There are also some cases in which *where* is used to locate something in time as well as place and thus is used in contexts where it could be interchangeable with *when.* However, this alternation is unidirectional. *When* seems to retain a strictly temporal frame of reference and all examples of it used as relative marker specify a point or period of time. *Where* seems to be much more productive as a relative marker in general.

39. a. “There's normally a day a week [(**when**/that/Ø/where) you didn't have lectures] so you worked at home.” – Darryl – (Mt Roskill, older)
b. “There'd be days [(**where**/that/Ø/when) I'd have one lecture]” – Chloe – (Titirangi, younger)
c. “we try and like head up to areas [(**where**/that/Ø/*when) it snows and things like that]” – Mayumi – (Mt Roskill, younger)
d. “mostly because there's so many places [(**where**/that/Ø/*when) they can hide up in the rooves]” – Ronaldo – (Titirangi, younger)

In examples 39 a and b the antecedent is temporal and both *when* and *where* are acceptable, whereas in examples 39 c and d, the referent is a physical location and only *where* is acceptable.

This is not the only complementary distribution in adverbial relative clauses. Consider the examples 40 a-d:

40. a. “and it got to the point, [(**where**/that/?Ø/when), um, I couldn't do it any more,] because she was just too, too abusive.” – Sharon – (Mt Roskill, older)
b. “having like some instances [(**where**/?that/*Ø/when) I was like, whoa, this is different!]” – Flora – (Mt Roskill, younger)
c. “I grew up with a family [(**where**/?that/*Ø/*when) my parents liked to show us different parts of the country]” – Kirsten – (Titirangi, older)
d. “Yep, cos he is from the city [(**where**/*that/*Ø/*when) he was first mayor].” – Chelsea – (Papatoetoe, younger)

Example 40a allows all four variants, with the possible exception of *Ø*, which is awkward if not entirely ungrammatical. Example 40b allows *where* and *when* and possibly *that* but not *Ø.* Example 40c allows *where* and possibly *that*. Example 40d only allows *where* unless the preposition *of* is added to the end of the clause which would make both *that* and *Ø* acceptable.

The coding for this study does not allow for a quantitative analysis on the variation in these adverbial relatives. However, there is both complementary distribution and optional variation that would make such an analysis worthwhile, just as is the case with the more frequently studied relative clause types explored earlier. Just as coding for animacy shows us variation in the distribution of *who* and *which,* coding adverbial relatives for whether they are temporal or locative would give us quantitative data on the constraints on *where* and *when.* It would also be useful to perform quantitative analysis on other syntactic factors that influence the choice of adverbial relativisers. This would give us more information about the variation in the English relative clause system. As adverbial relatives were the first relative constructions in English to use *wh*- relative markers (Gisborne and Truswell 2017) it would be informative to explore their variation in contemporary Englishes. Looking at

variation in adverbial relative markers would make it possible to investigate whether they retain artefacts of their legacy as the "first" *wh*- relatives.

4.8 Summary of results

The results of this study show that, at least superficially, the relative marker system in Auckland English does not differ wildly from that seen in varieties of English from more homogeneous – or less super-diverse - communities. The overall rates of use of the *that, Ø* and *wh*- relativisers in the study are fairly comparable to those seen in other studies of spoken Englishes. This will be discussed in greater detail in the chapter 5.

There are also some features in this data that warrant further exploration in the following chapter in order to better understand the grammar of relative clauses in Auckland English. The definiteness of the antecedent is the most consistent predictor of relativiser choice in Auckland English. There are some parallels between how definiteness patterns as a constraint in this data and the patterns seen in other studies of relative clause variation. There is also comparability with Sigley's (1997) results (despite coding differences), in the only other New Zealand study of relative clause variation.

There is little evidence of a change in progress for this variable although dividing the age groups into three rather than the original two proscribed by the study design is productive and suggests some possible age grading effects for the choice of relativiser in both subject and non-subject relatives.

There is some inter-community variation in the data, such as a favouring effect for *who* in subject relatives in Titirangi and a favouring effect for Ø in non-subject relatives in Mt Roskill. There are also a lack of constraints on relative marker variation in Papatoetoe for both subject and non-subject relatives. Possible explanations for these findings and how they relate to our original research questions will be discussed in the following chapter.

5. Discussion

This chapter places the results of this study in a wider linguistic context. Section 5.1 looks at the overall distribution of relative markers in this and other studies, in order to see if there is a consistent pattern to the rate of use of the different relativisers across different varieties of English and how Auckland English patterns in relation to other varieties. I then compare the syntactic (section 5.2) and social (5.3) factors that condition relativiser choice, that have been found to be significant in other varieties and compare these with both the Auckland corpus as a whole and then with the individual communities studied. As part of this comparison, I discuss the implications of this study's results, both in terms of our understanding of relative clause variation and of Auckland English more generally. In section 5.4 I discuss the implications of the qualitative analysis on those types of relative clauses generally excluded from variationist studies.

5.1 Overview

In this section, the first basis for comparison between this and other studies of relative clause variation is the overall frequency of each of the relative markers.
Table 5.1 shows a comparison of the total distribution of relative markers in this data with three other previously discussed studies. These studies were selected because they use exclusively spoken data as is the case for the Auckland Voices corpus. Levey's (2014) results listed here are the adult data from his study.

Relative marker	**Auckland Voices Corpus**		**Levey 2014 (Adults) (Ottawa)**		**D'Arcy and Tagliamonte 2010 (Toronto)**		**Tagliamonte et al. 2005 (UK)**	
	%	N	%	N	%	N	%	N
That	60.7	1373	56	409	55.6	1790	64	1230
Ø	20.6	466	24	175	26.7	859	28	528
Who	13.7	309	20	143	17.2	554	7	131
Which	5.0	113	0	1	0.1	4	1	19

Table 5.1 – The distribution of relative markers across the dataset for four different studies

The rate of use of *that* by our Auckland speakers appears to be fairly typical, falling as it does almost precisely between Tagliamonte et al.'s (2005) UK study and the two Canadian studies. Auckland speakers use of *who* also seems to fall at a mid-point, less frequent than for both sets of Canadian speakers but more than the UK speakers.
Aucklanders seem to use less Ø than the speakers in all three other studies and use more *which.* We would expect to see a higher rate of use of Ø in Tagliamonte et al.'s (2005) data as they look at varieties in which Ø subject relatives are acceptable, thus increasing the number of possible opportunities for Ø to be used as a relativiser. However Auckland speakers' use of Ø is also lower than the Canadian speakers', despite Ø being "largely restricted to non-subject function" (Levey 2014: 29) for the Canadian speakers in both of these studies. In Auckland, I found that Mt Roskill speakers used more Ø than the speakers in the other two communities (see table 4.5). In fact, at 26% (n=221), the Mt Roskill speakers' use of Ø relativisers was more comparable with the other studies, indicating that it is the Titirangi and Papatoetoe speakers' rather low usage (both 17% n=126 and 119 respectively) which accounts for the overall difference between Auckland English and the other varieties.

Auckland speakers' relatively high frequency of using *which* as a relativiser also stands out in table 5.1. High frequency of *which* relatives tend to occur in those studies that include written data (Ball 1996, Sigley 1997, Huber 2012, Hinrichs, Szmrecsanyi and Bohmann 2015). The *which* relative clauses in the Auckland corpus include examples which fall somewhere between canonical restrictive relative clauses and non-restrictive relative clauses. As noted in Chapter 2, many previous syntacticians have attempted to unify restrictive relative clauses and non-restrictive relative clauses and Sigley (1997) argues that all relative clauses fall on a cline of restrictiveness. It is possible that by taking Sigley's proposal seriously and by being more inclusive with respect to relative clauses in this study, the number of *which* relative clauses is slightly higher here. Otherwise I have no explanation for this difference.

A commonality between this study and the previous analyses is the complementary distribution of relative markers. Like D'Arcy and Tagliamonte (2010), I found that "only *that* is used productively across the restrictive domain, competing robustly in subject and non-subject positions" (D'Arcy and Tagliamonte 2010: 391). The second and third most

commonly used relative markers (*Ø* and *who*) are only used with any frequency in the same environments they have been found in the other studies, namely *Ø* relativisers occurs in non-subject relatives and *who* occurs in subject relatives with human antecedents. This is unsurprising given the well-documented, highly-circumscribed nature of the English relative pronoun system (Ball 1996, Nevalainen & Raumolin-Brunberg 2002, Sigley 1997, Tagliamonte 2002).

In the case of *Ø,* its distribution is restricted due to its ungrammaticality as a subject relativiser. Even in varieties that do allow *Ø* subject relatives, its use is still highly circumscribed and restricted to particular types of constructions (Tagliamonte et al. 2005: 101). *Wh*- relatives meanwhile, due to their status as relative pronouns rather than complementisers, are restricted due their semantics; they can only appear in relative clauses when their semantic features agree with their antecedent. This highly circumscribed system explains why, with some minor differences, varieties of English show comparable patterns in the distribution of the different relative markers. It may be assumed that the similarities in overall rates of use can be explained by common tendencies across varieties of English for speakers to talk about the same kinds of information (cf. Fox and Thompson 1990).

5.2 Review of syntactic constraints

The most consistently significant factor determining which relativiser is used is the definiteness of the relativised antecedent. It was found that, in non-subject relative clauses, indefinite pronouns favoured *Ø* (61.5%)*,* indefinite NPs disfavoured it (32.2%) and definite NPs slightly disfavoured *Ø* (46.2%). I say “slightly” because within 5% either side of 50, any favouring or disfavouring effect is fairly small.

These results are fairly comparable with other studies of relative marker variation. Levey (2014) and Tottie and Harvie (2000) both show a favouring effect for *Ø* relativisers with pronominal antecedents. While Tagliamonte et al. (2005) found no significant effects for definiteness in non-subject relatives, in subject relative clauses they found that “indefinite NPs are ranked higher for zero than definite NPs” (Tagliamonte et al. 2005: 100) which is the

reverse of what this data shows for non-subject relative clauses. However, Huddleston (1971) and Sigley (1997) both argue that there is a correlation between indefinite antecedents and overt relative markers which may help to explain why our indefinite NPs disfavoured Ø more than the definite NPs did.

In subject relative clauses with human antecedents, definite NPs disfavoured *who*, indefinite NPs had a close to neutral effect and indefinite pronouns favoured the use of *who*. There are fewer comparisons to be made here with other studies. Tagliamonte et al. (2005) do not explore the syntactic conditioning of the variation between *that* and *wh-.* While D'Arcy and Tagliamonte (2010) do look at *wh-* relativisers, they are more interested in the social rather than the syntactic constraints. The adult speakers in Levey (2014) show no significant conditioning effect for the definiteness of the antecedent for subject relatives.

As discussed in section 4.4.1, Sigley (1997) does provide a potential answer for why definiteness might be conditioning the choice between *who* and *that* in subject relatives. Looking at both antecedents and relative markers as having a hierarchy of explicitness, there is a negative correlation between the explicitness of the relativiser and the explicitness of its antecedent. Thus, *who* which is more explicit than *that* will most disfavour definite NPs which are the most explicit antecedent.

The consistency with which definiteness of the antecedent appears as a significant constraint on relativiser choice in this study suggests a characteristic of the grammar of Auckland English is for speakers to use the definiteness of the antecedent – if nothing else – as the main constraint on relativiser selection. To the extent that this finding aligns with Sigley (1997), it might even be reasonable to suggest it is a characteristic of New Zealand English as a whole.

Another significant syntactic factor in this data is the grammatical role of the matrix NP, that is, the role in the main clause of the NP that is being relativised. In the case of non-subject relatives, subject matrix NPs weakly favour Ø while direct object matrix NPs weakly disfavour it. A possible explanation for this lies in Fox and Thompson's (1990) discourse related explanation of the grammar of English relative clauses, outlined in section 2.3.1.

They argue that combinations where the NP-Rel and NP-Matrix have the same role are dispreferred – that is constructions where the NP matrix is the subject and the NP-Rel is the object (S-O) are favoured; object and object (O-O) constructions are not. While I found little evidence of this in this data (O-O constructions were the most prevalent) this does not necessarily mean that there are not some explanations to be found in this theory for the patterns that I did find. What is seen in this study is that S-O combinations favour Ø while O-O combinations disfavour it. Fox and Thompson's explanation for why S-O combinations are favoured over O-O is that subject matrix NPs need to be grounded in relation to the rest of the discourse and object relative clauses serve this function. Object matrix NPs are usually already grounded in relation to the subject in the matrix clause.

41. a. But the first job [I got at Middlemore Hospital] was in my school holidays. – Dot – (Papatoetoe, older)
    b. we have a Tongan hymn [that we s- both sing together]. – Catarina - (Papatoetoe, older)

Looking at these two examples, it is clear that the relative clauses are doing different jobs in terms of the information flow of the discourse. When the NP-matrix is a subject, as in 41a the role of the relative clause is to ground it – place it in the context of the wider discourse, which in this case is done by relating it to the speaker – "I got". However, when the NP-matrix is an object (41b), it has often already been grounded (by the pronoun in subject position) and thus the function of the relative clause is something else, in the case of 41b, to expand on the grounding from the matrix clause. This is shown in that the subject pronoun "we" in the relative clause is coreferential with the subject "we" in the matrix clause. It is possible that the preference seen in this study for Ø in S-O constructions is because this grounding function of the relative clause is best achieved without an overt relativiser intervening between the NP-matrix and the relative clause, while in O-O combinations an overt relativiser is used to point to and specify the additional information.

I do not find, as evidence from other studies (Ball 1996, Sigley 1997) did, that where Ø is an option, it is favoured with human antecedents. Ball explains this trend with the argument that when *who* entered the relativiser system, it was associated with "personal" (human)

antecedents causing *that* to move away from them and become associated with non-personal antecedents. As a result, Ø (and of course *who*) "expanded to take its place" (Ball 1996: 246). In this data, this historical legacy is only evident insofar as *who* is only used with any frequency in human subject relatives. There is no evidence in non-subject relatives that Ø takes the place of *that* with human antecedents.

5.3 Review of social constraints

Gender only appears as a significant social factor in non-subject relatives. In a choice between Ø and *that* as a non-subject relative marker, the men in this study used Ø 49.3% of the time and the women used it 37.9% of the time. This is perhaps surprising in that both Levey (2014) and D'Arcy and Tagliamonte (2010) find some (though admittedly small) correlation between *who* (in human subject relatives) and female speech. D'Arcy and Tagliamonte (2010) in particular link this preference to the perceived formality of *wh*-forms. D'Arcy and Tagliamonte use level of education as a correlate of formality in their study and show that the group with a higher level of education used more *who* relativisers. Although level of education was not coded in this study, community was a significant social factor in determining the rate of use of *who*, and, like D'Arcy and Tagliamonte, the speakers from the community with the highest socioeconomic status (Titirangi) use more *who* relativisers. D'Arcy and Tagliamonte (2010) found that in the group with a higher level of education, women use significantly more *who* relativisers than men. Given this, one might expect Titirangi, the community with the highest rate of *who,* to show some gender effect, however, no community showed significant results for *who* subject relatives.

As previously discussed, non-subject relatives show a significant correlation between Ø and male speech. Neither Levey (2014) nor Tagliamonte et al. (2005) show a significant gender effect for Ø in non-subject relatives. Sigley (1997) found that Pākehā men had a higher rate of use of Ø relativisers than Pākehā women and both Māori men and women. He argues that this is a function of privilege; that overt relativisers are a floor holding device and white men are less likely to be interrupted in conversation and therefore less likely to need to use overt relativisers. This argument is not supported by the data from this study. The older

Titirangi male speakers in our corpus are arguably the most traditionally privileged speakers and therefore might be expected to pattern like the Pākehā men in Sigley's study. While it appears that this group uses the most Ø in the data, this is mainly the effect of a single speaker and thus I can make no wider claims about this result. Without a larger number of older male Titirangi speakers, Sigley's argument does not provide a convincing explanation for our results. However, while a gender effect is seen for use of Ø in the data as a whole, when I examine individual communities this effect is only significant in Mt Roskill. Younger Mt Roskill speakers, who produce 272 out of the 414 non-subject relative clause tokens, are a very diverse group. None of the male speakers in this group identify as Pākehā, and they would not be typically seen as highly privileged speakers entitled to hold the floor.

What then could be the explanation for the preference for Ø by male speakers in Auckland? It is possible that there could be an association of formality with this variable. When the corpus was split into three age groups rather than two, there appeared the v-shaped curve of age graded variables (D'Arcy and Tagliamonte 2010) with speakers in the middle age group favouring Ø relativisers less than both older and younger speakers. This could mean that Ø has an association with non-standardness or low prestige (Cheshire 2005). If this variation is stable (and there is no evidence in the data of change taking place) we might expect to see female speakers using less of this non-standard variant (Labov 1990, 2015). While Ø in non-subject relatives is not particularly correlated with informality in the Canadian and UK studies, Sigley (1997) does observe an increase of Ø in informal texts in Wellington English. Possibly, this formality distinction between Ø and *that* is a peculiarity of New Zealand English.

When the age groups were divided into three, it appeared that non-subject relatives may in fact be losing the gender dimension to their variation. Male and female speakers in the younger (under 25) age group are converging on a much more similar rate of Ø relativiser use than the older groups have. It is possible that for younger speakers of Auckland English it is not as important to index gender with your choice of non-subject relative marker.

Although this corpus was designed to interrogate language change, and the age groups chosen reflect this design, it is possible that two broad age groupings (young versus old)

might be effective in looking at phonological change, but they are not nuanced enough to observe change when examining grammatical variables. Certainly, when the age groups are divided into three the results are much closer to significance. It seems plausible that if I looked at a larger sample with more age groups, including speakers in the 25-40 age bracket that was omitted, some clearer patterns might appear that could tell us what is happening in the relative marker system in Auckland over time. The decision to focus on the two very broad age groups with a clear gap in between was largely made on the basis of available funding (M Meyerhoff, personal communication, 21.02.2019) and it is hoped the corpus will grow in due course.

Community only appears as a significant factor in the choice of relative marker when speaker is removed as a random effect. As previously discussed, there is considerable inter-speaker variation in this dataset. Not only is there considerable variation in the number of relative clauses produced by each speaker but for both subject and non-subject relatives there are a number of speakers who have categorical behaviour, producing 100% of one or other of the variants. Given this, I feel justified in undertaking parallel analyses that do not include speaker as a random effect and in discussing the community differences shown in the data.

For subject relatives with human antecedents, *who* is favoured by speakers in Titirangi, very slightly disfavoured by Mt Roskill speakers and disfavoured more by speakers in Papatoetoe. Given the demographics of these three communities, this is consistent with the proposition that *who* in Auckland English has some of the prestige and standardness associations that *who* has been shown to have in other studies (D'Arcy and Tagliamonte 2010, Hinrichs, Szmrecsanyi and Bohmann 2015, Tagliamonte et al. 2005). This is further supported by the distribution of *who* across the three age groups, that is with a peak in use for the middle age group, the reverse of the v-shaped curve seen for non-standard variants.

For non-subject relatives, Ø is favoured by Mt Roskill speakers and slightly disfavoured by speakers in both Titirangi and Papatoetoe. Mt Roskill speakers' favouring of Ø actually makes their rate of use of this variant more comparable with studies of other varieties of English, while Titirangi and Mt Roskill speakers have an unusually low rate of use of this

variant. Mt Roskill as a community, has undergone a lot of recent migration since the 1990s and now has no ethnic majority. While Papatoetoe is also very diverse, the immigration patterns that led to this were less recent and started in the 1960s-1970s. It is possible that a lower rate of use of Ø relatives has been indexical of a particularly New Zealand pattern of relativisation and Mt Roskill speakers are orienting to a more international pattern of relativisation where Ø is used more frequently. It is beyond the scope of this study to evaluate how languages and varieties other than New Zealand English that have a strong presence in Mt Roskill might be influencing the favouring effect for Ø relatives in this community.

When looking at the constraints on relativiser variation in each of the communities individually, it is clear that in both subject and non-subject relatives it is the community that favours the variant other than *that* – i.e. *who* in subject relatives and Ø in non-subject relatives – that shows the most constraints on its use. Mt Roskill speakers favour Ø in non-subject relatives and of the independent variables I tested, three appear as significant constraints for Mt Roskill speakers, two for Titirangi speakers and none for Papatoetoe speakers. Titirangi speakers favour *who* in subject relatives and they have three significant constraints on the distribution of *who,* while the other two communities have none.

Possibly, what marks someone as a Titirangi speaker is not just favouring *who* in human subject relatives, but also having a system where the distribution of *who* relativisers is circumscribed by a number of factors. The same might also be the case for Mt Roskill speakers and Ø in non-subject relatives. What might this say about Papatoetoe speakers, who favour neither Ø in non-subject relatives nor *who* in subject relatives and show no significant constraints on their use? It is possible that their relativiser variation is constrained by factors that have not been included in this study. Some other studies include the distance of the relative clause from its antecedent but other than this there are no obvious factors that could have been included in the analysis but were not.

Another possible explanation for the lack of constraints on relativiser variation in Papatoetoe lies in Papatoetoe's demographic make-up. As previously mentioned, Papatoetoe has the longest history of diversity of our three communities. This makes

Papatoetoe the community that is most comparable with studies of other superdiverse cities such as London (Cheshire et al. 2011) and Toronto (Nagy 2011) which have a longer history of diversity. As discussed in section 2.4, diverse communities with an abundance of variants in the feature pool (Mufwene 2001) often show linguistic levelling (Cheshire et al. 2011) and select one, unmarked variant and generalise it to all environments, abandoning complex systems of constraints. Papatoetoe may well be showing signs of this kind of levelling. Speakers in Papatoetoe disfavour the alternate variant to *that* in both subject and non-subject relatives. *That* is the obvious choice for selection as a single unmarked variant for a levelled system as it fulfils more criteria for selection from the feature pool (Siegel 1997). Unlike the other relative markers, *wh*- and *Ø,* there are few exceptions on which relative clauses *that* can be used in. This regularity is an important factor in the selection of a variant. If Papatoetoe speakers have developed or are developing a levelled system in which *that* reigns supreme, the lack of constraints on relative marker variation in this community is to be expected.

5.4 Review of qualitative analysis

While not the primary focus of this study, the qualitative exploration of Ø as a subject relativiser and of adverbial relative clauses allowed us to make further observations both about the Auckland English relative marker system and about the standard methodology for variationist studies of relative clauses.

5.4.1 Ø Subject relatives

Tagliamonte et al. (2005) provide us with more insight into the issue of Ø as a subject relative as their study explores dialects in which this is an acceptable and more widely used construction. They found that the most significant factor in determining Ø subject relatives was sentence type. According to Tagliamonte et al. (2005), Ø subject relatives are favoured by, and in fact "virtually restricted to" (Tagliamonte et al. 2005: 101) existentials, clefts, and possessives. Across all three communities studied by Tagliamonte et al., existential constructions most highly favoured Ø relatives. While clefts were not coded in our study, existential constructions were. Of the 11 clear Ø subject relatives in this data, 6 appeared in

existential constructions. Considering that existentials only account for 11.9% (268 tokens) of the overall dataset, this seems quite a high proportion. While we have to be extremely cautious about making any generalisations based on so little data, it is certainly worth noting that the same kind of construction that favours Ø subject relatives in varieties of English in which they are grammatically acceptable seems to be the same construction in which they appear even when they are supposedly ungrammatical, as is the case for New Zealand English.

Possessive constructions were not coded separately in this corpus so I cannot determine what percentage of the total data they make up, however they do account for 3 of the remaining 5 Ø subject relatives. I would suggest that this is an argument in favour of these constructions, rare though they are, being examples of true Ø subject relatives in Auckland English. If they were merely random anomalies or errors in either production or coding there would be no reason for them to behave like Ø subject relatives in other varieties. This in no way suggests that Ø subject relatives are entering the Auckland English vernacular. However it does suggest that they are not entirely ungrammatical in Auckland English and their distribution is constrained by similar factors as in other varieties.

5.4.2 Adverbial relative clauses

As we have seen, adverbial relatives offer optional variation in a very similar way to other relative clauses, that is, with *that, Ø* and *wh-* relative markers all being possible in some constructions but with complementary distributions depending both on the grammatical structure of the relative and matrix clauses and characteristics of the antecedent. For this reason, extending the variationist study of relative clauses to include adverbials might be warranted in the future. This would be particularly valuable in light of the fact that recent research by Gisborne and Truswell (2017) has found that the adverbial relativisers *where* and *when* were the first *wh-* relative pronouns to enter the English dependent relative system. By including adverbial relatives in a variationist analysis I could assess whether synchronic patterns of variation continue to show reflexes of this linguistic history. Features of the variation in contemporary language varieties can be used to interrogate the historical variation and change of a language (Labov 2008).

Questions to be asked in such a study would include: (i) how frequently are *wh*- relativisers used in adverbial relatives compared to *that* and *Ø;* (ii) how does this compare with the frequency of *wh*- relativisers (such as *who* and *which*) in other types of relative clauses; and (iii) what are the constraints (both syntactic and social) on *wh-, that* and *Ø* in adverbial relatives and how does this compare with other relative clauses? As the first constructions in which *wh*- relative markers entered the system, one might expect adverbial relatives to have a higher frequency of *wh*- than other types of relative clauses. If *wh*-relativisers were first used in adverbial constructions they might have infiltrated this domain more thoroughly before moving up the hierarchy to object and subject constructions.

6. Conclusion

This study of the variation in relative markers in Auckland English has illustrated several points. First, it adds support to the contention that the variation in relative clauses in English is highly circumscribed and varieties tend to have comparable patterns in terms of their overall rates of use of the various relativisers (Ball 1996, D'Arcy and Tagliamonte 2010, Levey 2014). Second, restrictiveness is a far from binary concept and warrants serious consideration in any variationist study of relative clauses as even slight ambiguity can affect results. Communities such as Auckland that have undergone dramatic and rapid change can still show distinctive, community marking features – such as how speakers orient to the definiteness of the antecedent through their choice of relative marker. Despite this, smaller communities within a larger one can also show distinctive individual constraints on variation. Finally, there is value in looking at grammatical "anomalies". Even very low frequency tokens can provide information about a variable.

I must now consider how well this study has answered our key research questions. Turning to our first question, namely 'Is there change taking place in the Auckland relative marker system?', the answer based on the data in this study is that there is little evidence of change. I find no significant apparent time/generational difference in relativiser distribution when the speakers are divided into two age groups, but when the speakers are divided into three age groups, there is some evidence of an age-grading effect, and this would suggest that this is a stable variable (Labov 1994). The fact that the age-grading pattern observed in this study is directly comparable with that found by Sigley (1997) further strengthens the inference that this is a stable variable in New Zealand English as a whole, and within Auckland this is not changing.

Our second question, as to whether change is emanating from more diverse communities is rendered irrelevant if in fact there is no change taking place, so I will set this aside. However, the effect of community diversity on variation still warrants some further discussion, so I will return to this shortly.

The third question, 'How comparable are the constraints on relativiser variation in Auckland English and in other varieties?', requires a more complex answer. In this study, I have demonstrated that certain aspects of the variation of relative clauses are highly circumscribed. There are many similarities in relativiser variation between different varieties of English both in terms of the overall rates of use of different relative markers and their complementary distribution. This shows that some features of relative clause variation in English are, if not universal, at least very pervasive across geographically distant varieties of English. For instance, the overall percentage rate of use of *that* seems fairly consistent in this study and in other studies of spoken English (D'Arcy and Tagliamonte 2010, Levey 2014, Tagliamonte et al. 2005). Ø is consistently the second most frequently used relativiser, used predominantly in non-subject relatives, even in varieties where Ø subject relatives are acceptable (Tagliamonte et al. 2005). Animacy is a consistent factor in determining relativiser choice across varieties, and Auckland English is no exception. *Who* is used almost exclusively with human (or at least animate) antecedents and mostly in subject relatives.

There are, however, some features of relative marker distribution that appear to be more idiosyncratic of Auckland English. One such peculiarity is the role of speaker gender in constraining relativiser choice. While there is no gender dimension to the variation between *that* and *who* there is a gender effect for the use of *Ø.* This is almost the reverse of what is found in other studies (D'Arcy and Tagliamonte 2010, Levey 2014). It is possible that this is because Ø relativisers are indexed as informal in Auckland English. The hypothesis that Ø might be associated with informality is supported by the v-shaped pattern of age-grading that can be seen for its distribution when the data is reanalysed over three age groups. Gender and formality may in fact be interrelated here with men using more of the informal variant, as is often seen in cases of stable variation (Labov 2001). This, along with the age-grading pattern observed are arguments for, though not conclusive evidence of the stability of this variable in Auckland English.

Variation in subject relatives also appears to be sensitive to some kind of notion of formality with *who* correlating positively with formality. This can be seen both in its n-shaped age distribution and in the preference for *who* in Titirangi, the community with the highest socio-economic demographic. This is directly comparable with other studies of relative

marker variation, especially Tagliamonte et al. (2005) and D'Arcy and Tagliamonte (2010) who demonstrate formality associations for *wh*- relative markers.

Our final research question regarding whether the relative marker system in Auckland English shows evidence of levelling, necessitates a return to the question of community diversity. Given that there was no evidence of language change in the data, we might expect to see no evidence of levelling either, given that levelling is a type of, or the result of, change. While there is little evidence of levelling in the data overall, when looking at the communities individually, Papatoetoe shows a dispreference for any relative marker other than *that* and there are no significant syntactic or social constraints on relativiser variation. Given that Papatoetoe is the community in our dataset that has been highly diverse for the longest, this is precisely the community in which we would most expect to find levelling. I have no clear explanation for why we see a levelled system in Papatoetoe with no evidence of change, in which younger speakers' relativiser systems are becoming even more levelled than their older counterparts' systems, or why there appears to be no diffusion of this levelling to other communities in Auckland. If the levelling in Papatoetoe started early enough, it may be that there is no evidence of inter-generational change there because the shift to *that* was completed more than a generation ago.

Further research could interrogate the issue of change versus stability in the Auckland relative marker system by extending the number of age groups surveyed. Collecting data from a sufficient number of participants to create groups of speakers in their twenties, thirties, forties, fifties, etcetera and plot relativiser variation across these groups, would provide a more nuanced picture of how the variation patterns according to age. In addition, having a greater number of speakers of all ages between 20 and 90 might provide a better understanding of the possible age-grading effects seen in Auckland overall and the apparent levelling and simplification of rules in Papatoetoe.

Another avenue for future research is, as previously discussed, to examine variation in adverbial relative clauses. The quantitative methods used in this study to explore the variation in NP relative markers could be applied to the variation in adverbial relativisers. This would not only tell us more about the relative marker system in Auckland English but it

would also contribute to a theoretical framework for extending variationist studies of relative clauses in general.

This study sought to explore the variation in relative markers in Auckland English through the lens of the extensive demographic change that has taken place in this city. I found that the relative marker system in Auckland English still shows some comparability with the patterns of variation seen in studies from more homogeneous communities worldwide (D'Arcy and Tagliamonte 2010, Levey 2014) as well as with New Zealand English from over two decades ago (Sigley 1997). However, certain features of the dataset, including the considerable inter-speaker variation and the lack of constraints on variation in some cases, especially in Papatoetoe, suggest that there may be some disruption going on in Auckland's relative marker system. As previously noted, Auckland's status as a superdiverse city is relatively new and it may take some time for the effects of this diversity to register itself in the language to the extent seen in cities (cf. Cheshire et al. 2011) with longer histories of diversity.

References:

Adger, David 2003. *Core syntax: A minimalist approach*. Oxford: Oxford University Press.

Ball, Catherine N. 1996. A diachronic study of relative markers in spoken and written English. *Language Variation and Change*. 8(02): 227-258.

Blommaert, Jan 2013. *Ethnography, Superdiversity and Linguistic Landscapes: Chronicles of Complexity*. Bristol: Multilingual Matters.

Bresnan, Joan W. 1972. *Theory of complementation in English syntax*. Cambridge, MA: MIT dissertation.

Buchstaller, Isabelle 2015. Exploring linguistic malleability across the life span: Age-specific patterns in quotative use. *Language in Society*. 44(4): 457-496. doi:10.1017/S0047404515000391

Carnie, Andrew 2013. *Syntax: A generative introduction*. 3rd edn. Chichester, West Sussex: Wiley-Blackwell.

Chambers, Jack K. 1995. *Sociolinguistic theory: Linguistic Variation and its social significance*. Oxford: Blackwell.

Cheshire, Jenny 1982. *Variation in an English dialect: A sociolinguistic study*. Cambridge: Cambridge University Press.

Cheshire, Jenny 2005. Age and generation-specific use of language. In Ulrich Ammon, Norbert Dittmar, Klaus Mattheier and Peter Trudgill (eds.), *Sociolinguistics: An introductory handbook of the science of language and society*. Berlin: Mouton de Gruyter. 1552–63.

Cheshire, Jenny 2013. Grammaticalisation in social context: The emergence of a new English pronoun. *Journal of Sociolinguistics* 17(5): 608-633.

Cheshire, Jenny, Paul Kerswill, Sue Fox, and Eivind Torgersen 2011. Contact, the feature pool and the speech community: The emergence of Multicultural London English. *Journal of Sociolinguistics* 15(2): 151-196.

Cofer, Thomas M. 1975. Performance constraints on relative pronoun deletion. *Linguistics* 157: 13-32.

D'Arcy, Alexandra, and Sali A. Tagliamonte 2010. Prestige, accommodation, and the legacy of relative who. *Language in Society* 39(3): 383-410.

Demirdache, Hamida 1991. Resumptive chains in restrictive relatives, appositives and

dislocation structures. PhD thesis, Massachusetts Institute of Technology, Massachusetts, USA.

Denison, David 1998. Syntax. In Suzanne Romaine (ed.) *The Cambridge history of the English language. Volume 4, 1776-1997*. Cambridge: Cambridge University Press. 92-329.

Douglas, Jamie 2017. Unifying the that-trace and anti-that-trace effects. *Glossa: a journal of general linguistics* 2(1).

Downes, William 1998. *Language and society*. 2nd edn. Cambridge: Cambridge University Press.

Eckert, Penelope 1989. The whole woman: Sex and gender differences in variation. *Language Variation and Change*. 1: 245-26

Emonds, Joseph 1979. Appositive relatives have no properties. *Linguistic Inquiry* 10(2): 211-243.

ELAN (Version 5.2) Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands 2018. https://tla.mpi.nl/tools/tla-tools/elan/ (4 April 2018)

Erlewine, Michael Yoshitaka. 2016. Anti-locality and optimality in Kaqchikel agent focus. *Natural Language & Linguistic Theory* 34(2): 429–479. DOI: https://doi.org/10.1007/s11049-015-9310-z

Fox, Barbara A. and Sandra A. Thompson 1990. A discourse explanation of the grammar of relative clauses in English conversation. *Language* 66(2): 297-316.

Gisborne, Nikolas, and Robert Truswell 2017. Where do relative specifiers come from? In Eric Mathieu & Robert Truswell (eds.), *Micro-change and Macro-change in Diachronic Syntax*. Oxford: Oxford University Press. DOI:20.500.11820/f86f0f6f-3425-40fb-8862-f178661c0411

Guy, Gregory R. and Robert Bayley, 1995. On the choice of relative pronouns in English. *American Speech* 70: 148-161.

Haegeman, Liliane 1994. *Introduction to government and binding theory*. Oxford: Blackwell

Hay, Jennifer, Katie Drager & Brynmore Thomas 2013. Using nonsense words to investigate vowel merger. *English Language & Linguistics* 17(2): 241-269.

Hinrichs, Lars, Benedikt Szmrecsanyi, & Axel Bohmann 2015. Which-hunting and the Standard English relative clause. *Language* 91(4): 806-836.

Hodges, John C. 1951. *Harbrace College Handbook*. New York: Harcourt Brace.

Holmes, Janet, Allan Bell & Mary Boyce 1991. *Variation and change in New Zealand English: A social dialect investigation*. Project Report to the Social Sciences Committee of the Foundation for Research, Science and Technology. Wellington: Victoria University of Wellington, Department of Linguistics.

Huber, Magnus 2012. Syntactic and variational complexity in British and Ghanaian English. In Bernd Kortmann, and Benedikt Szmrecsanyi (eds.) *Linguistic Complexity: Second Language Acquisition, Indigenization, Contact*. Berlin: Walter de Gruyter. 218-242

Huckin, Thomas N., Elizabeth H. Curtin, and Debra Graham 1986. Prescriptive Linguistics and Plain English: The case of "Whiz-deletions". *Visible Language* 20.2: 174-187.

Huddleston, Rodney 1971. *The sentence in written English.* Cambridge: Cambridge University Press.

Huddleston, Rodney and Geoffrey K. Pullum 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Hyde, Janet S. 2005. The gender similarities hypothesis. *American Psychologist* 60(6): 581—92.

Jackendoff, Ray 1977. *X syntax: A study of phrase structure*. Linguistic Inquiry Monograph 2. Cambridge, Massachusetts: MIT Press.

Johnson, Daniel, E. 2009. Getting off the GoldVarb Standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass* 3(1): 359–383.

Kandybowicz, Jason 2014. Comp-trace effects explained away. In Donald Baumer, David Montero, and Michael Scanlon (eds.), *Proceedings of the 25th West Coast Conference on Formal Linguistics.* Somerville, MA: Cascadilla Proceedings Project. 220-228.

Kautzsch, Alexander 2002. *The historical evolution of earlier African American English: An empirical comparison of early sources.* Berlin: Mouton de Gruyter.

Keenan, Edward L. 1985. Relative clauses. In Timothy Shopen (ed.), *Language typology and syntactic description.* Cambridge: Cambridge University Press. 141-170.

Keenan, Edward L. and Bernard Comrie 1977. Noun phrase accessibility and universal grammar. *Linguistic inquiry* 8(1): 63-99.

Kerswill, Paul & Ann Williams 2000. Creating a New Town koine: Children and language change in Milton Keynes. *Language in Society*. 29: 65-115.

Labov, William 1972. *Sociolinguistic patterns*. Philidelphia: University of Pennsylvania Press

Labov, William 1990. The intersection of sex and social class in the course of linguistic

change. *Language variation and change*. 2(2): 205-254.

Labov, William 1994. *Principles of linguistic change, Volume 1: Internal Factors.* Oxford: Blackwell.

Labov, William 2001. *Principles of linguistic change, Volume 2: Social Factors.* Oxford: Blackwell.

Labov, William 2008. Mysteries of the substrate. In Miriam Meyerhoff & Naomi Nagy (eds.) *Social Lives in Language: Sociolinguistics and multilingual speech communities*. Amsterdam:John Benjamins. 315-326.

Labov, William 2015. The discovery of the unexpected. *Asia-Pacific Language Variation*. 1(1): 7-22.

Lakoff, George. 1968. *Deep and surface Grammar*. Bloomington, Indiana: Indiana University Linguistics Club.

Levey, Stephen 2014. A comparative variationist perspective on relative clauses in child and adult speech. In Rena Torres Cacoullos, Nathalie Dion & André Lapierre (eds.), *Linguistic Variation: Confronting Fact and Theory*. New York: Routledge. 22-37.

McCawley, James D. 1981. The syntax and semantics of English relative clauses. *Lingua* 53: 99-149.

Maclagan, Margaret, and Elizabeth Gordon 1999. Data for New Zealand social dialectology: the Canterbury corpus. *New Zealand English Journal* 13: 50-58.

Meyerhoff, Miriam 2010. *Introducing Sociolinguistics*. London: Routledge

Meyerhoff, Miriam 2017. *Community dependencies: Connections and discontinuities in Auckland City.* Key note lecture at the New Zealand Linguistic Society Conference, Auckland, 23-24 November 2017.

Meyerhoff, Miriam, Erik Schleef and Laurel Mackenzie 2015. *Doing Sociolinguistics*. London: Routledge.

Meyerhoff, Miriam, Catherine Watson, Elaine Y. Ballard and Helen Charters 2015. Breaking Babel – Rethinking Language Change in a super-diverse city. Marsden Grant Proposal.

Milroy, Lesley and Mathew Gordon 2008. *Sociolinguistics: Method and interpretation.* Massachusetts: Blackwell Publishing

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

Mufwene, Salikoko S. 2001. *The ecology of language evolution*. Cambridge: Cambridge University Press.

Nagy, Naomi 2011. A Multilingual corpus to explore variation in language contact situations. *Rassegna Italiana di Linguistica Applicata* 43(1-2): 65-84.

Nevalainen, Terttu and Helena Raumolin-Brunberg 2002. The rise of the relative who in Early Modern English. In Patricia Poussa (ed.), *Relativisation on the North Sea littoral*. 109-121.

Oxford English Dictionary 2018*a.* http://www.oed.com.helicon.vuw.ac.nz/view/Entry/200178?rskey=35Cnxf&result=2#eid (15 June 2018)

Oxford English Dictionary 2018*a.* http://www.oed.com.helicon.vuw.ac.nz/view/Entry/228284?redirectedFrom=which#eid (15 June 2018)

Romaine, Suzanne 1982. *Socio-historical linguistics: its status and methodology*. Cambridge: Cambridge University Press.

Ross, Háj 1995. Defective noun phrases. In Andrea Dainova, Rachel Hemphill, Barbara Luka, Barbara Need, and Sheri Pargman (eds.) *Papers of the 31st Regional Meeting of the Chicago Linguistics Society*. 398–440.

Roy, Joseph, and Stephen Levey 2014. Mixed effects models and unbalanced sociolinguistic data: The need for caution. *Poster presented at NWAV 43*. Illinois.

Sankoff, Gillian 2006. Age: Apparent time and real time. In Brown, Keith (ed.), *Encyclopedia of language and linguistics*, 2nd edn. Amsterdam: Elsevier. 110–16.

Sells, Peter 1985. *Restrictive and non-restrictive modification*. CSLI Report 85-28. Center for the Study of Language and Information, Stanford University.

Siegel, Jeff. 1997. Mixing, leveling and pidgin/creole development. In Arthur K. Spears and Donald Winford (eds.), *The Structure and Status of Pidgins and Creoles*. Amsterdam: Benjamins. 111-149.

Sigley, Robert J. 1997. Choosing Your Relatives: Relative Clauses in New Zealand English. PhD thesis, Victoria University of Wellington, Wellington, New Zealand.

Statistics New Zealand n.d.*a*. http://archive.stats.govt.nz/Census/2013-census/profile-and-summary-reports/quickstats-culture-identity/languages.aspx (28 October 2018)

Statistics New Zealand n.d.*b*. http://archive.stats.govt.nz/Census/2013-census/profile-and-summary-reports/quickstats-about-a-place.aspx?request_value=13170&tabname= (28 October 2018)

Tagliamonte Sali A. 2002. Variation and change in the British relative marker system. In Patricia Poussa (ed.), *Relativisation on the North Sea Littoral*. Munich: Lincom Europa. 147–165.

Tagliamonte, Sali A. 2006. *Analysing sociolinguistic variation*. Cambridge: Cambridge University Press

Tagliamonte, Sali A. Jennifer Smith & Helen Lawrence 2005. No taming the vernacular! Insights from the relatives in northern Britain. *Language Variation and Change* 17(1): 74-112.

Thompson, Sandra A. and Anthony Mulac 1991. The discourse conditions for the use of the complementizer that in conversational English*. Journal of pragmatics* 15(3): 237-251.

Tottie, Gunnel and Dawn Harvie 2000. It's all relative: Relativization strategies in early African American English. In Shana Poplack (ed.), *The English history of African American English.* Oxford: Oxford University Press. 198–230.

Trudgill, Peter 1974. Linguistic change and diffusion: Description and explanation in sociolinguistic dialect geography. *Language in society* 3(2): 215-246.

Peter Trudgill 2004. *New-dialect formation*. Edinburgh: Edinburgh University Press.

Trudgill, Peter, Elizabeth Gordon, Gillian Lewis, and Margaret Maclagan 2000. Determinism in new-dialect formation and the genesis of New Zealand English. *Journal of Linguistics* 36(2): 299-318.

Vertovec, Stephen 2014. *Super-Diversity*. London: Taylor & Francis.

Warren, Paul 2006. Oops, I've done a futt: quality and quantity in a New Zealand vowel contrast. *Te Reo* 49: 125-143.

Wittenburg, Peter, Hennie Brugman, Albert Russel, Alex Klassmann, and Han Sloetjes 2006. ELAN: a Professional Framework for Multimodality Research. In *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation.* 1556-1559.

Appendices

Appendix A – Content of Participant Information Sheet

Project title: Auckland Voices: Language in a superdiverse city
Researchers: Miriam hoffrhoff (Victoria University of Wellington)
and Elaine Ballard, Helen Charters & Catherine Watson (University of Auckland)

Introduction

This information sheet explains why we are making recordings in Auckland and what we will do with them. This sheet is yours to keep so please take your time to read through it and to ask questions about anything you'd like to clarify.

We know that recent migration patterns have had a big effect on how Auckland looks. Our study wants to see whether migration is affecting the way Aucklanders talk. This project is funded by the Royal Society of New Zealand for three years and involves researchers at Victoria University of Wellington and the University of Auckland. We would like to record a conversation between you and one of our team members to include in our project.

What happens

Participation is completely voluntary. If you agree to participate, one or two members of our team will digitally-record a conversation that will last between one and two hours. We can stop for breaks, if you need refreshment, and you are in control of the conversation. If at any point you want to stop, for any reason, you can say "Let's stop now" and we will.

The recording can be completed in any quiet place where you feel comfortable. This may be your home, a community centre or another quiet place you know and like. We will probably talk about what life is like living in Auckland now and what life was like when you were growing up (whether in Auckland or someplace else). There are no 'right' answers in the discussion – we're just interested in finding topics you enjoy talking about.

We will always respect your confidentiality. All recordings will be assigned a reference code and we can make up a name to refer to you by if you would like this (a pseudonym), and we will only refer to the recordings in our research using the code or your made-up name. We will transcribe the recordings and we will keep an original copy as a backup, but only our core team members will have access to the original recordings. When we use the recording for research or teaching, we will remove references to your name and to names of other people or places which might identify you.

You may withdraw your consent at any time up to one year from the date of the final recording and we'll destroy your recording. After one year, some of our analysis will have already been completed and it may be difficult to remove your data. We plan to keep these recordings for possible future research after this project is finished, so in the consent form please let us know if you would prefer for your recording not to be used in any future

studies. Also, please let us know if there are any sections of the recordings that you would like to be deleted or not used for future studies.

How will we use the recording

Your recording will be transcribed (written down word for word) and then we will look in detail at some of the ways in which people in different communities in Auckland are pronouncing words or using English. For us, there are no 'right' or 'wrong' ways of talking. The way people talk can tell us interesting things about their identity. We may use excerpts from your recording in public presentations about this project but if we do, any excerpt will be kept strictly anonymous. The project team, based in Wellington and Auckland, will have access to all the materials.

If we do use your recording in any future research, we may be working with other researchers who will have access to the anonymised versions of the recordings. They will only know your reference code, or made-up name and any relevant biographical information about you, such as your gender, your age and how long you and your family have lived in this part of Auckland. If you would prefer for the core research team to be the only ones with access your recording or transcript, you can let us know this too.

Your recording and transcript will be stored as digital files on an external hard drive, which will be stored securely in a locked cabinet at all times. The files may be changed to a new format as technology changes, but they will always be stored securely. We will also store your consent forms securely.

We are very happy to give you a copy of your recording and/or transcript. We are also happy to keep you up-to-date with our research publications on this project. Please let us know if you would like to get copies of your recording, the transcript or our research and provide us with the contact details you would like us to keep for this purpose. Please note that we will not share these contact details with anyone – they will be solely used for providing you with the requested information.

Contact details

If you have any questions, comments, concerns or thoughts about the project or the recordings, please don't hesitate to get in touch with one of the project investigators:

Professor Miriam Meyerhoff (Victoria University of Wellington)
Ph: 027 860 1497
Email: miriam.meyerhoff@vuw.ac.nz

Dr Catherine Watson (University of Auckland)
Ph: 09 923 5979
Email: c.watson@auckland.ac.nz

If you have any questions or concerns relating to the collection or use of data you can also contact:
Dr Meredith Marra (Head of School, Victoria University of Wellington)
Ph: 04 463 5636
email: meredith.marra@vuw.ac.nz

For any queries regarding ethical concerns, you may contact the ethics convenor at Victoria University of Wellington:
Dr Susan Corbett (Convener, VUW Human Ethics Committee)
Ph: 04 463 5480
email: susan.corbett@vuw.ac.nz

Appendix B – Content of Participant Consent Form

Consent Form
This form will be kept indefinitely

Research title: Auckland Voices: Language in a superdiverse city
Researchers: Miriam Meyerhoff (Victoria University of Wellington)
and Elaine Ballard, Helen Charters & Catherine Watson (University of Auckland)

Please read the following, and tick the boxes that you agree to. You may choose to tick only some of the boxes and not others, or you may tick all of them. Your tick(s) plus your signature at the end of the form indicate that you consent to that component of the study.

- ☐ I agree to **have a conversation with a project researcher for at least an hour.**
- ☐ I agree to have this conversation **digitally recorded**.
- ☐ I grant you permission to keep and **use my recording and transcript**, with the understanding that my confidentiality will be maintained at all times.
- ☐ I grant you permission to use the recorded material for **academic purposes**, e.g. discussions, presentations, teaching and any published or unpublished works. My confidentiality will be maintained at all times.
- ☐ I grant you permission to **share anonymised versions of my recording and transcript with other co-investigators** in future collaborative linguistic studies.

I have read the Participant Information Sheet and I understand the intent and nature of the intended research. I understand that:

- my participation is voluntary,
- my real name or any personal, identifying information will not be used in any part of the project,
- the recording will be kept indefinitely – subject to the conditions I just agreed to,
- I can ask to have any portion of the recording erased that I don't want you to keep,
- I can ask to have the conversation stop at any time and for any reason,
- I can withdraw my conversation from the project up to one (1) year after a recording is made in which case the recording will be destroyed,
- I may, at any time, withdraw from any <u>future project</u> not already underway.

I have had the opportunity to ask questions and have received satisfactory answers about the proposed research. otherwise, I understand that the data will be stored in a secure location indefinitely.

Name of participant	Signature of participant (if applicable)
Date	Signature of researcher

please turn over...

Request for Further Contact

I would like to have the following information or documents sent to me:

- ☐ a digital copy of my recording
- ☐ a transcript of my recording
- ☐ summaries of any research findings that stem from this project
- ☐ electronic copies of any journal papers and/or presentations that stem from this project

I would like the above to be sent to me at the following address(es):

Name (for use in correspondence) _____

Email address _____

Postal address _____

_____

_____

_____

Appendix C - Multiple regression analysis of factors contributing to the choice of Ø in non-subject relatives (speaker as random effect).

Non-subject Relatives Ø vs Ø + *that* **R2 =** 0.253 **input probability =** 0.528 **N=** 1064			
	FW	**%**	**Total N**
Definiteness of the antecedent	p.<0.0000		
Indefinite Pronoun	0.67	61.5	104
Definite	0.46	46.2	556
Indefinite	0.36	32.2	404
Gender	p.<0.05		
Male	0.58	49.3	420
Female	0.42	37.9	644
Length of relative clause	p.<0.05		
Short	0.54	48.4	512
Long	0.46	36.8	552
Role of the antecedent NP in the matrix clause	p.= 0.0584		
Subject	0.58	51.8	276
Other Object	0.48	40.9	274
Existential	0.48	34.1	88
Direct Object	0.46	39	426
Community	p.= 0.0916		
Mt Roskill	0.59	51.9	414
Titirangi	0.46	35.9	345
Papatoetoe	0.45	36.7	305
Animacy of the antecedent NP in the matrix clause	p.= 0.192		
Other animate	0.67	64.7	17
Inanimate	0.42	42.6	861
Human	0.40	39.2	186
Complexity of relative clause	p.= 0.329		
Simple	0.53	43.8	940
Complex	0.47	31.5	124
Age group	p.= 0.759		
Younger	0.51	43.9	651
Older	0.49	40	413
Speaker (random)			

AK-M-M01	0.81	95.7	23
AK-T-I10	0.778	81.8	22
AK-P-M03	0.748	76.5	24
AK-M-M03	0.741	76.5	17
AK-MOB-M11	0.72	100	7
AK-T-V06	0.714	60.9	23
AK-T-B11	0.675	60	15
AK-P-R05	0.673	75	8
AK-P-R12	0.665	54.5	11
AK-M-M10	0.655	90	10
AK-P-R03	0.651	50	12
AK-P-R11	0.64	44	25
AK-P-R02	0.638	70.6	17
AK-P-R06	0.629	70	10
AK-T-V07	0.611	41	39
AK-M-M04	0.598	75	8
AK-T-B09	0.59	60	15
AK-P-R01	0.588	53.8	13
AK-M-R03	0.582	56.2	80
AK-P-F06	0.58	36.4	22
AK-T-V08	0.579	46.2	13
AK-MY-M12	0.559	48	25
AK-M-A06	0.558	54.5	11
AK-M-R01	0.556	53.8	9
AK-P-R04	0.554	40	5
AK-T-I05	0.536	37.5	8
AK-MO-M09	0.533	50	6
AK-P-R08	0.532	37.5	8
AK-T-I04	0.523	27.8	18
AK-MY-M13	0.519	66.7	21
AK-T-I02	0.516	30	10
AK-T-I03	0.512	25	8
AK-MY-M14	0.506	60	15
AK-T-I07	0.501	27.8	18
AK-T-I06	0.492	27.3	22
AK-M-R07	0.485	60	5
AK-M-R02	0.479	47.1	17
AK-P-R13	0.478	38.5	26
AK-M-M07	0.474	50	16
AK-M-R09	0.467	56.2	10
AK-T-I01	0.464	25	24
AK-M-M05	0.443	33.3	6
AK-T-B12	0.423	28.6	7
AK-MY-R11	0.42	35.3	17
AK-T-V04	0.41	33.3	6
AK-P-R10	0.401	33.3	9

AK-P-M04	0.401	16.7	18
AK-T-V02	0.392	18.5	27
AK-T-I08	0.391	30	10
AK-M-M02	0.388	20	5
AK-M-A05	0.385	50	20
AK-T-V03	0.369	25	12
AK-P-R07	0.363	27.3	22
AK-P-M07	0.361	20	10
AK-M-R04	0.358	27.3	11
AK-PB-M01	0.346	14.3	7
AK-T-I09	0.328	11.1	18
AK-M-R10	0.322	20	10
AK-M-M08	0.306	22.2	18
AK-P-M05	0.306	15	20
AK-M-R08	0.287	13.3	15
AK-P-M02	0.264	0	12
AK-M-M06	0.239	25	32
AK-P-R09	0.238	38	26
AK-T-V01	0.223	16.7	30

Appendix D – Multiple regression analysis of factors contributing to the choice of *who* in subject relatives with human antecedents (speaker as random effect).

Subject Relatives *who* vs *who* + *that* **R2 =** 0.43 **input probability =** 0.446 **N=** 668 (human antecedents only)			
	FW	**%**	**N**
Definiteness of the antecedent	p.<0.01		
Indefinite Pronoun	0.65	57.1	42
Indefinite	0.49	46.8	365
Definite	0.36	34.9	261
Role of the antecedent NP in the matrix clause	p.= 0.0561		
Other Object	0.58	48.5	163
Direct Object	0.55	45.6	272
Subject	0.52	38.6	145
Existential	0.35	30.7	88
Community	p.= 0.167		
Titirangi	0.64	55.4	168
Mt Roskill	0.49	43	249
Papatoetoe	0.38	34.3	251
Length of relative clause	p.= 0.212		
Long	0.54	46	398
Short	0.47	38.1	270
Complexity of relative clause	p.= 0.533		
Simple	0.52	46.3	121
Complex	0.48	42	547
Age group	p.= 0.75		
Older	0.52	44.5	310
Younger	0.48	41.3	358
Gender	p.= 0.786		
Male	0.52	42.8	283
Female	0.49	42.9	385
Speaker (random)			
AK-M-A05	0.916	100	10
AK-P-F06	0.914	88.9	18
AK-P-R02	0.878	85.7	7
AK-M-R07	0.852	87.5	8
AK-MY-R11	0.844	87.5	8

AK-T-I04	0.81	92.3	13
AK-P-R11	0.795	66.7	12
AK-T-V07	0.793	84.6	26
AK-M-R04	0.793	80	5
AK-T-V06	0.782	100	5
AK-M-M05	0.781	85.7	7
AK-T-V04	0.774	100	2
AK-P-R07	0.77	69.2	13
AK-P-M07	0.719	55.6	9
AK-PB-M01	0.716	60	5
AK-T-V08	0.699	75	4
AK-M-M01	0.698	65.2	23
AK-P-R04	0.691	60	5
AK-M-M04	0.643	60	5
AK-M-M07	0.64	60	32
AK-P-M03	0.64	43.8	11
AK-T-V02	0.634	71.4	7
AK-T-I07	0.631	75	8
AK-P-R10	0.617	44.4	18
AK-M-R01	0.614	60	5
AK-MY-M12	0.609	52.4	21
AK-T-B09	0.594	66.7	3
AK-T-I08	0.593	66.7	6
AK-T-I05	0.571	57.1	7
AK-M-M08	0.568	50	2
AK-T-V01	0.556	57.1	14
AK-P-R08	0.552	31.2	16
AK-MOB-M11	0.548	50	8
AK-MO-M09	0.53	42.9	7
AK-M-M03	0.527	50	6
AK-M-M06	0.511	47.8	23
AK-P-R06	0.472	27.3	11
AK-MY-M14	0.453	33.3	12
AK-T-I02	0.439	50	8
AK-T-I03	0.429	57.1	7
AK-P-R01	0.424	0	1
AK-P-R12	0.422	0	1
AK-T-B11	0.391	40	5
AK-M-A06	0.384	0	1
AK-T-I10	0.365	37.7	8
AK-T-V03	0.353	0	1
AK-P-R05	0.34	15.4	13
AK-T-I06	0.316	36.4	11
AK-P-R09	0.282	9.1	11
AK-M-M02	0.279	0	3
AK-M-R08	0.251	14.3	7

AK-P-M02	0.251	6.7	15
AK-P-M05	0.245	7.7	13
AK-M-R10	0.235	13.6	22
AK-M-R03	0.231	15	20
AK-M-R02	0.216	0	4
AK-T-I09	0.189	15.4	13
AK-T-B12	0.182	0	4
AK-MY-M13	0.176	8.3	12
AK-T-I01	0.165	9.1	11
AK-P-M04	0.156	4.3	23
AK-T-V05	0.154	0	5
AK-M-M10	0.135	0	8
AK-P-R03	0.125	0	14
AK-P-R13	0.124	0	14
AK-M-R09	0.09	0	11