The acquisition of variation: Arab Migrants' acquisition of (ING) and Coronal Stop Deletion in Wellington

by

Rania Ali Zarour

A thesis submitted to the Victoria University of Wellington

in fulfilment of the requirements for the degree of

Doctor of Philosophy

School of Linguistics and Applied Language Studies

Victoria University of Wellington

(2018)

Copyright statement

The author of this thesis (including any appendices and schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") she has given Victoria University of Wellington certain rights to use such Copyright, including administrative purposes. This page must form part of any such copies made.

Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and Reproductions described in it may take place is available in the University IP Policy (see https://www.victoria.ac.nz/fgr/current-phd/ip).

Dedication

To my wonderful parents, loving Husband, amazing children and coffee table.

Acknowledgements

First of all, I would like to offer my greatest gratitude for my father who has always been there for me, believed in me, and who has always been willing to encourage my academic endeavours.

I would also like to acknowledge the contribution made by various people who have been involved in this project over the last few years. Firstly, I would like to thank my supervisors: Miriam Meyerhoff and Janet Holmes, whose knowledge and advice on all matters has been invaluable. I would also like to acknowledge the help provided by Paul Warren, who gave me access to the New Zealand Spoken English Database, Daniel Ezra Johnson, who has been kind enough to help me with statistics-related matters as well as invaluable knowledge and insights for the use of R and Rbrul, and Lisa Woods who has also helped me with my statistical endeavours.

Secondly, I would also like to thank the administrative staff at the school of Linguistics, and Applied Language studies especially Janet Attrill for all the help and advice she has provided over the years, always having a spare key for me to get to the office.

Finally, I would like to offer my gratitude to all the participants who took part in the study, who, without exception, were generous with their time, hospitality, and willingness to help. Thank you.

Abbreviations and Acronyms

1. CSD: Coronal stop deletion

2. (ING): Coronal and velar alterations

3. NS: Native speakers

4. NNS: Non-native speakers

5. NZE: New Zealand English

6. NZSED: New Zealand Spoken English Database

7. AM: Arab migrants

8. L1: First language

9. L2: Second language

Abstract

This dissertation investigates the patterns of acquisition found among Arab migrants to Wellington for two stable variables: coronal stop deletion (CSD) and (ING). CSD is the alternation between retained and deleted final consonant clusters, i.e. */wɛst/* vs. */wɛs/* and (ING) is the realisation of the final nasal in unstressed word-final syllables i.e. */dɹaivɪŋ/ vs. /dɹaivɪn/.* CSD is a phonological variable that is mainly conditioned by articulatory constraints while (ING) is a morpho-phonemic variable with syntactic conditioning as well.

An emerging trend in variationist sociolinguistics is to study variation in non-native varieties by analysing how far non-native speaker (NNS) patterns of variation replicate constraints on variation found among native speakers (NS) of a target variety.

This study applies variationist methods to investigate the following questions: 1. What are the linguistic and the non-linguistic constraints that condition variation in the production of (ING) and CSD among NS in the New Zealand Spoken English Database (NZSED) in Wellington? 2. What are the linguistic and the non-linguistic constraints that condition variation in the production of (ING) and CSD among Arab migrants in Wellington (AM)? 3. Based on the results for (1) and (2), is there any evidence for "transformation under transfer" (Meyerhoff, 2009a)

Interpretation of the results is done in line with the so-called "three lines of evidence", and considers significant and non-significant constraints, constraint hierarchies and rank ordering of constraints (Tagliamonte & Temple, 2005).

I consider the proposition that AMs, of all ages, are prone to transformation under transfer of NS constraints on the variables CSD and (ING), mainly illustrating strong and weak transfer. It is expected that old and middle-aged AMs will have patterns different from those found among young AMs .

I also consider the possibility that articulatory constraints may be more readily transformed by AMs into ethnolectal marking, whereas grammatical constraints may be more likely to be strongly transferred by AMs. Old and middle-aged AMs seem to be more likely to display strong transfer of NS constraints, but they do not seem to be using variation in the L2 stylistically. By contrast, young AMs stylistically use articulatory constraints to convey important social indexicalities.

The results suggest that oold and middle-aged AMs with developing grammars are like NS children acquiring variation of their L1, in the sense that old and middle-aged AMs are sensitive, in both CSD and (ING), to dialect-specific constraints on variation as they display strong transfer of the highest ranked NS constraint, be it articulatory or grammatical in nature; they also seem to perceive NS frequencies of occurrence of variables.

Old and middle-aged AMs have an advantage over NS-children in their cognitive abilities that enable them to apply global constraints on variation by filtering their previous exposure to English, to replicate grammatical constraints of the L2 variables. Old and middle-aged AMs also seem to replicate the articulatory constraints that are perceptually salient, or that can host L1 transfer. They sometimes innovate articulatory constraints that are meaningful to them probably because of the influence of their L1.

Young AM, who have arrived in New Zealand at an age of six years or younger, would be expected to illustrate strong transfer for stable variables like CSD and (ING). The results, nevertheless, illustrate that although young AMs share the same significant constraints found among NS of NZSED, they have different rank orderings, internal hierarchies and frequency of variants. Young AM, unexpectedly, diverge from NS norms and exhibit weak transfer of NS articulatory constraints on CSD, while they show strong transfer of NS grammatical constraints for the variable (ING). I suggest that young AMs seem to be using articulatory constraints in the L2 stylistically, to convey important social indexicalities.

In addition, young AMs seem to hold an intermediate status between NS of NZSED and first-generation AM. Like old and middle-aged AMs, they replicate global-grammatical constraints on (ING) with an internal hierarchy that follows the Labovian nominal-verbal continuum, rather than the local, internal hierarchy. This suggests that (ING), as a morphophonemic variable with syntactic interfaces, has less room for the stylistic use of variation patterns as a reflection of identity marking. Articulatory constraints may be more subject to L1 transfer and these may become a marker of ethnicity among a Second-generation of migrants.

Table of contents

List of Tables

List of figures

Chapter 1: Introduction

1.1 Variationist Sociolinguistics

Multi-ethnic communities are becoming more the norm rather than the exception in today's world, and immigration is a phenomenon that is highly associated with this. New Zealand steadily welcomes an increasing number of immigrants, for example, in the 2016-2017 financial year alone, 52,455 residence visas were issued for people from over fifty countries of origin (Ministry of Business, 2017).

New immigrants, who come from non-English speaking countries, have the choice of maintaining their languages as well as acquiring the variety of the host country. The acquisition of a host country's linguistic variety may be the result of two different but related processes; language acquisition and language shift. Language shift occurs when migrants gradually stop using their first language (L1), accommodate to the second language norms resulting in a change in language proficiency or use in the L2 (Jaspaert & Kroon, 1993). This shift generally occurs in favour of the host language and may be completed within three generations of migrants (Holmes et al. 1993; Coles, 1994).

A very important aspect of a second language acquisition is that competence in a second language (L2) is not merely the acquisition of obligatory native-speaker-like (NS) linguistic competence, but also the acquisition of sociolinguistic competence, which is the acquisition of NS patterns of variation (Drummond, 2012).

Variationist sociolinguistics is a subfield of sociolinguistics that attends to sociolinguistic competence; the way language varies and changes, diachronically and synchronically, in communities of speakers as a result of the interaction among social factors,

linguistic factors and cognitive processes. Variation patterns available across the various levels of language are the result of orderly or "structured heterogeneity" (Labov, 2001); this structured variation is a part of the human language capacity.

Variationist sociolinguistics, which can be traced to the early 1960s, moved away from the Chomskian (1952) paradigm of formal grammar and traditional dialectology and focused attention on the concept of linguistic varieties. Such varieties might refer to dialects, ethnolects or more divergent varieties such as creoles.

The American sociolinguist, Labov (1963) established the field of variationist sociolinguistics in his famous work on diphthongs on Martha's Vineyard, an island in the United States. He devised the notion of the linguistic variable to capture variation, quantitatively. A linguistic variable is a set of related linguistic forms (originally focusing on phonology, subsequently expanded to morphology, syntax etc.), all of which mean the same thing and which correlate with some social meaning in a speech community.

An example of a sociolinguistic variable is coronal stop deletion in word-final clusters in English where the final stop is either deleted or is retained, i.e. [wɛst] vs. [wɛs]. The two options are collectively called the variants of the variable coronal stop deletion (CSD). Variants are not necessarily binary choices, they could be continuous as well, like vowels. Identifying all possible contexts where the variants can alternate is called establishing the envelope of variation.

Variationist sociolinguistics targets the most unmonitored form of language, the vernacular, by implementing sociolinguistic interviews to collect examples of speech data across several linguistic tasks that are believed to display varying degrees of linguistic monitoring including different styles such as reading a passage, reading a word list, reading minimal pairs and conversations (see Labov 2001).

The goal of the interview is usually to get participants as emotionally involved as possible to the end of engaging in a narrative where they become so immersed in the content of what they are saying that they almost forget that they are being observed. The purpose of this is to elicit more spontaneous speech, and therefore the more consistent norms of the vernacular (Labov, 1984).

This data collection toolbox called for new statistical analysis concepts and programs. The method's mathematical implementation was first introduced by Cedergren and Sankoff (1974), who provided a description of the variable rule program which has repeatedly been improved ever since. The newest version of the variable rule program, Goldvarb X is an updated version of its predecessors Goldvarb 2.0 and Goldvarb 2001. Another statistical analysis program that is implemented by many sociolinguists is Rbrul which was designed to deal with many of the shortcomings presented by the Goldvarb family (Johnson, 2009).

What the variable rule programme does is to quantitatively model variation as a probabilistic output of several factors that are believed to collectively affect the probability of a variant choice (response). In other words, a variable rule analysis computes a logistic statistical model, where the effects of many factors that collectively affect a response are measured, based on observed data distribution[1], such that each determining factor is assigned a numerical factor weight that describes how much it influences the probabilities of choice of a response.

---

[1] Observed data distribution in this study is the observed distribution of the tokens, representing the variables of this study, in the corpora.

The output from the variable rule programme is presented under what Tagliamonte (2006) calls "the three lines of evidence" including (i) statistical significance, (ii) rank ordering and (iii) constraint hierarchies. Having constraints on variation presuppose that variant choice and its distribution in linguistic contexts is conditioned by a group of independent predictors called "constraints".

Each of these constraints has several lower-level factors called "levels". For example, if we consider one of the constraints on CSD in English, the following segment, then all possible following-sounds will be the levels of that constraint. Each of the levels has a unique effect in predicting one variant over others, and this effect is presented by a numerical value called "factor weights". The order of the internal levels according to their factor weights results in a continuum called the "constraint hierarchy".

The constraints are ordered according to the relative strength of influence each constraint has on the conditioning of the variable. The range between the highest and the lowest factor weights for levels within a constraint hierarchy is indicative of the overall strength a constraint has in predicting variation (Hazen, 2011).

Since its inception in the 1960s, quantitative variationist sociolinguistics has enhanced our understanding of how social, linguistic and more recently, cognitive factors relate to language variation and change with a huge focus on variation patterns among native speakers of a linguistic variety including NS children. Eckert (2012:87) identified three waves of variation studies based on the focus of these studies.

"The first wave of variation studies established broad correlations between linguistic variables and the macro-sociological categories of socioeconomic class, gender, ethnicity, and age. The second wave employed ethnographic methods to explore the local categories and configurations that inhabit, or constitute, these broader categories. … the third wave, arguing that (a) variation constitutes a robust social semiotic system, potentially expressing

the full range of social concerns in a community; (b) the meanings of variables are underspecified, gaining more specific meanings in the context of styles, and (c) variation does not simply reflect, but also constructs, social meaning and hence is a force in social change."

Variationist sociolinguists have recently directed their attention to the acquisition of variation among non-native speakers (NNS) in various contact settings like study abroad settings, selective migration, forced migration, business relocation and the like. These studies show that variation in the speech of NNS is also systematic and not random.

Some studies have examined variation in a target language as a reflection of second language (L2) processing. They considered synchronic variation in the learners' speech as a reflection of second language learners' underlying developing grammars. Selinker (1972) laid the foundation for this concept as an individualized system of a language learner that is rooted in both the native language and the target language but suggested that it cannot be explained entirely in comparison to them. Referring to this approximate system, he coined the term "learner interlanguage'".

Recent work in variationist sociolinguistics consider variation in NNS speech in comparison to NS patterns of variation; meaning that NNS variation patterns are analysed to see how far they replicate the patterns of variation found in the vernacular of NS. This field of inquiry is, nevertheless, relatively underdeveloped; a blind spot, for most contact-induced language studies, let alone is the role of the interplay of variation, change and contact-settings in multilingual communities (Léglise & Chamoreau, 2013).

The ethnic styles of young people with migrant backgrounds have recently become of interest in variationist research with a focus on new features of everyday talk in the speech of young people of migrant origins in various social settings (Kern, 2011). This emerging interest implies that there are shared patterns of variation existent among NNS in contact

settings. I hypothesize, bearing that in mind, and in line with Schleef et al's (2011) suggestion that the inherent complexity of the acquisition of variation would inevitably affect the patterns of acquisition of variation among NNS of a second language (L2), that NNS in contact settings may display similar tendencies in the acquisition of variation.

The current study is a typical variationist sociolinguistic study of the acquisition of variation by NNS of English in Wellington, the capital of New Zealand, as a contact setting. The main scope of this study is to investigate the acquisition of variation by Arab migrants (AM) to Wellington; first-generation of AMs and their 1.5 generation of AMs[2].

I aim to outline patterns of variation across AMs to illustrate how far they replicate NS of New Zealand English's forms of variables (variants), the raw frequencies of forms (variant frequencies), and the underlying constraints on the distribution of forms (variant distribution) which collectively represent the underlying grammar AMs have and whether it is similar and/or different from NS' grammar.

To embark on this, I have decided to work on two extensively-studied and stable variables in the English language: coronal stop deletion in final consonant clusters (CSD) and the realization of the final nasal in unstressed syllables, (ING).

The reason behind my choice of these variables is that they are stable. This means that I have a very good understanding of the aspects of NS-patterns of variation for these two variables including forms (variants), raw frequencies of forms and the constraints conditioning the distribution of these forms across many NS of English varieties.

The fact that these variables are stable makes my task of understanding how NNS acquire the aspects of variation more feasible since there is an underlying assumption that the

---

[2] 1.5 generation is the term used in variationist sociolinguistics to refer to younger NNS who were not born in a host country but who have relocated to it very young.

constraints conditioning the variation for stable variables among NS have not changed for a long time. There is also an elaborate literature on the variables across a decent number of NS varieties of English and contact-induced varieties of English including ethnolects and English-based creoles.

The reason why I have chosen Wellington as the contact setting for my study is that this contact setting is particularly different from other contact settings, in the sense that the AMs in this study have willingly decided to move to Wellington where there are no visible signs of social isolation, discrimination or the like.

These migrants left their countries in search of a better future for themselves and their offspring. It is only logical, therefore, to assume that this fact will influence their attitudes towards New Zealand, New Zealand English, belonging and social networking, all of which have been found to be important social factors that affect the acquisition of variation (see Edwards (2011); Schleef et al. (2011)).

Consequently, AMs to Wellington might be particularly predisposed to replicate NS patterns because they do not have any attitudinal, social or physical and/or regional boundaries that would lead to divergence from NS norms.

The other reason that makes this group worthy of studying is the potential effects of the Arabic language on the acquisition of variation. Firstly, through possible transfer effects of the strong diglossic norms of Arabic that may render Arabs predisposed to acquire style-shifting constraints on variation.

Specifically, language transfer, overt attitude, social evaluations and social constraints are more readily accepted to influence the acquisition of variation in migrant communities. Evidence suggests that not only surface structures, but also "high level" structures are subject to attitudes and social evaluations (Levon & Buchstaller, 2015).

Assuming that the social meaning of a variable is at least in part conditioned by speakers' previous knowledge and previously held views of each other i.e., the social histories of the speakers, then potentially, sociolinguistic precepts that are important for NNS might affect how they perceive the social meaning of variation as well as the non-linguistic constraints governing such variation in the L2.

Arabs may thus be predisposed to see style in New Zealand English (NZE) and to look for standard and prestigious variables as a way of reproducing Arabic diglossia, where a high and a low variety coexist as registers and are subject to conscious code-switching (Al-Wer, 2009; Holes, 1993). This "nativist dualism" is a crucial attitude for Arabic-speakers (Suleiman, 2011) and an integral aspect of the Arab post-colonial self (Miller & Caubet, 2009).

The second reason to study Arab migrants is that the phonemic inventory of Arabic may introduce a wide range of variants in the realization of one variable but not the other. For example, Levantine Arabic has many coronal stops that may affect the acquisition of CSD whereas the number of nasal phonemes in Arabic is rather limited.

The AM sample group that I have worked with are from the same geographical region called the Levant; the term refers to a collection of countries in the Eastern Mediterranean including Lebanon, Syria, Jordan and Palestine. This would make L1 transfer, when relevant, a group effect on the acquisition of variation.

The group also share similar social indices of linguistic choices, an issue that will be relevant when I discuss how AMs tend to reinterpret the social indices of some variables and might even end up creating new social constraints based on how they interpret the meaning of variants in the L2.

1.2 Thesis structure

The structure of the thesis is as follows: The remainder of this introduction, chapter one, provides a profile of Arab migrants in Wellington, including some background information on the established Lebanese community in the city and the reason they were excluded from this project. Chapter two, the literature review, introduces the complexity of the acquisition of variation by NNS in contact settings which compels them to handle variation using different techniques. We will see that different contact settings are associated with specific trends of handling L2 variation among NNS; as NNS living in multilingual metropolises share patterns of acquisition of variation that are different from patterns found among NNS living in less ethnically diverse contact settings. I then introduce the variables of the study: CSD and (ING) and the chapter concludes with the research questions.

Chapter three introduces the methodology applied to study the acquisition of variation by NNS in Wellington. The recent trend in Variationist sociolinguistics to study the acquisition of variation among NNS is to see how far they replicate NS constraints. I had, therefore, to create a benchmark of New Zealand English patterns of variation on CSD and (ING) by analysing NS' norms in the New Zealand Spoken English database (NZSED).

I have also compiled a non-native speaker corpus and then analysed it to see how far AMs replicate the NS norms of variation on CSD and (ING). I then discuss sampling methodology, and the sociolinguistic interview and how the interviews were transcribed and coded for the variables. I also introduce the coding rationale for NS and NNS for the variables CSD and (ING). I then present methods used for analysing the data including the use of parametric and non-parametric tests.

The main body of the thesis is introduced in chapter four which presents the results for CSD and (ING); first for NS of NZSED and then for AMs. Chapter five, the discussion, highlights the most relevant findings from the study and illustrates patterns of acquisition

found across NS of NZSED and NNS living in Wellington. Finally, chapter six, the conclusion, examines the extent to which the research questions have been answered, with a focus on the contribution of this study to the existing body of knowledge.

1.3 The social profile of Arabs in New Zealand

According to the Encyclopaedia of New Zealand; Te Ara (2014) the existence of Lebanese Arabs in New Zealand goes back as early as the early settlers in the country. In the late 1800s and early 1900s, thousands of Lebanese migrated to the Americas, a few went to Australia and still fewer to New Zealand.

Upon arrival in New Zealand, the Lebanese settled in clusters in Dunedin and Auckland and as early as 1890 blended into the community and became integrated. After the early arrivals in the 1890s, the number of Lebanese migrants dropped. But relatives of earlier settlers continued arriving in fewer numbers. By the 1940s the Dunedin Lebanese were largely assimilated, and many have since moved to other parts of the country. In the 1990s some 1,000 Otago and Southland residents were descended from the original migrants (Ministry for Culture and Heritage of the New Zealand Government, 2014).

Smaller numbers of Lebanese people arrived in New Zealand in the time span 1975-1990 as they fled the civil war back home. They settled mainly in Auckland, Wellington and Dunedin.

In the late 1900s and early 2000s people started to arrive from other parts of the Arab world, for example, from Kuwait, the United Arab Emirates, Saudi Arabia, Syria, Jordan and Bahrain. This is reflected by residency applications in the years 1997-2012 (Immigration New Zealand 2017).

In 2013, the Arab ethnic group comprised less than one percent of the New Zealand population most of whom cluster in Auckland (see Figure 1). The Lebanese were excluded

from my study mainly because they identify themselves separately from Arabs. The data from the New Zealand Census (New Zealand Statistics, 2013) confirms this observation; the Lebanese populations in New Zealand, do not report their race as Arab (they choose Lebanese).

The reason is likely to be that the Lebanese have been in New Zealand as far back as the 1850s and are for this reason a well-established community. Therefore, the focus of the study narrowed to include only Levantine Arabs excluding the Lebanese.

The Wellington region hosts a few number of Arab migrants (AM), most of whom cluster in Newtown, Kilbirnie and Miramar. My fieldwork investigation and my connections with formal brokers of the Arab community in Wellington helped me realize that there are approximately two hundred families who arrived in Wellington from the Levant and I have made acquaintance with a large part of the Arab community and have excluded families with intermarriages from outside the Levant.

Figure 1.1: Arab ethnic group by region, 2013 NZ Census
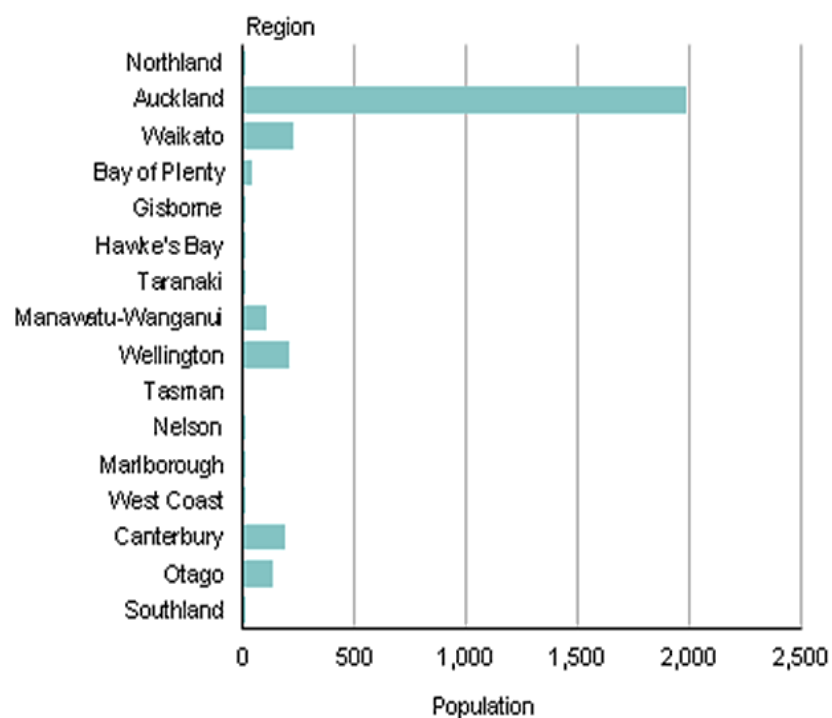
Note: Some counts may be too small to show on graph.
Source: Statistics New Zealand

Chapter 2: Literature review

As suggested earlier, the multi-ethnic atmosphere of Wellington, as in other major cities in New Zealand, and the complexity of the process of the acquisition of variation, and the need to display NS-like constraints on variation by NNS, as a sign of community-membership, may compel NNS to devise techniques to handle the new sociolinguistic situation they are exposed to. NNS experiences may involve multilingualism, language and dialect levelling, linguistic innovation and the "transformation under transfer" (Meyerhoff , 2009a) of NS patterns of variation.

But before I discuss the last-mentioned aspect of acquisition of variation as a typical linguistic behaviour of NNS in contact settings, I will try to illustrate how and why the acquisition of variation is a complex process for NNS.

2.1 The Complexity of the acquisition of variation by NNS

First, variation is a part of our linguistic competence not just our performance (Meyerhoff & Schleef, 2012).This means that a NNS's task is not only to replicate NS surface structures and their frequencies, but they must achieve a certain level of linguistic competence to be able to start targeting NS variation patterns. In other words, the acquisition of variation in the L2 by NNS has the acquisition of sufficient L2 linguistic competence as a prerequisite, and this is a very complex process for adult NNS, to begin with.

More specifically, we can define linguistic competence as the acquisition of obligatory native speaker (NS) forms while sociolinguistic competence is the acquisition of NS norms for language use, these include NS patterns of variation (Drummond, 2012), though both competencies are interconnected (Hymes, 1972). NNS must acquire sufficient linguistic competence that will then allow them to target and work on the details of sociolinguistic competence.

In a similar vein, Sharma and Sankaran (2011) propose that the acquisition of linguistic constraints is indicative of gaining new linguistic systems by NNS, whereas acquiring non-linguistic constraints indicate gaining sociolinguistic competence.

There is a growing awareness and concomitant interest in sociolinguistics that NS-like linguistic proficiency requires awareness of the patterns of variation that characterize what NNS identify as the L2 target variety. Moreover, researchers are also increasingly aware of the work L2 speakers may have to do to evaluate and emulate the socially indexed values of variables, as they are used in the speech community of the target language (Eckert, 2008a; Schleef et al., 2011; Meyerhoff & Schleef, 2012; Schleef, 2013b).

This growing focus in variationist sociolinguistics indicates that NNS need to achieve "social awareness" of linguistic variables after they accurately identify the linguistic variables of a speech community (Labov, 2001). Not only that, but presumably, they also should develop their "sociolinguistic monitors" (Labov et al., 2006) which is a concept that proposes that social judgement, perception and attitudes are operated through a cognitive mechanism that makes speakers sensitive to the social indices of variation in a linguistic variety. These sociolinguistic monitors are argued to be shared by members of a speech community and therefore entitle them to community agency.

If, following Eckert (2008a), we assume that the expression of systematic patterns of socially stratified variation is a means by which speakers can express social agency, then it is easy to see that for L2 migrants, the acquisition of variation might also be wrapped up in a larger task of finding new ways to express new identities and ways of belonging.

The first challenge in relation to the acquisition of social indexicality of variables is to work out the link between the social indices of variants and social constructs like age, gender and social status and this link is not straightforward (Ochs, 1992; Silverstein, 2003; Eckert, 2008a, 2008b). Some of the most obvious cues of this link might be the physical cues of the

vocal tract, but even in this case, there is no one to one correspondence between social meaning and variables (Foulkes & Docherty, 2006).

What adds to the complexity of this task is that some social constructs are not always easily recognized because they may not have corresponding social traits that are visible in a community like age or gender. Even social status, which is sometimes taken for granted in sociolinguistic studies as a cumulative trait of wealth and education, is not necessarily defined as such among NNS communities.

Another challenge for NNS acquiring the indexicality of variation in an L2 is encapsulated in Eckert's concept of the "indexical field". Understanding variables as being situated within a broad indexical field means that the social meanings ascribed to variants are multiple, though presumably not completely unrelated (Eckert, 2008a). The implications of this to NNS is that they must try and tease out among a multiplicity of meanings which ones are relevant to them or that they can reasonably identify with or that others would be open to having them identify with.

For example, the velar variant of the variable (ING) is associated with the concept of education which has many positive attributes such as professionalism, formality and elite social status. Nevertheless, educated speakers can also be viewed as condescending, pretentious and self-conceited depending not only on context but also on the attitudes of the speakers towards each other and the identities they would like to assume.

To illustrate this process, I use the work of Podesva (2007) where he discussed how a gay lawyer who usually assumes a "gay diva" persona among his friends switched to a more subdued version of himself when he was on a radio show. Probably to pass as a professional, trustworthy person rather than a flamboyant stereotype. This was manifested in the lawyer's use of tokens of released /t/.

The social meaning of a variable is also, at least in part, affected by the interlocutors' previous knowledge and previously held views and attitudes about each other and the social histories of the speakers.

Accordingly, NNS' understanding of social indexes and social categories in the L2 might be affected by their knowledge of social indexicality in their first language. This would potentially affect how NNS perceive the social meaning of variation as well as the non-linguistic constraints, like age and gender, governing such variation in an L2.

For example, Schleef et al. (2011) found that migrant Polish teenagers in Edinburgh and London seem to perceive gender to be an important factor in the conditioning of the stable variable (ING). The authors explain the emergence of this constraint as an innovation by suggesting that the Polish teenagers might be perceiving gender as a pre-existing and accessible social indexicality. Probably caused by what the Polish children perceived as hints of gender in their already-established social norms and accordingly social meaning.

Sociolinguistic indexicality is also relative to contexts as well as the speakers' need to establish a certain position to others in an interaction. For example, Okamoto (2014) suggests that what might qualify as polite in one context is not necessarily so in another.

Not only is the social indexicality of variants multi-fold and context-bound, but it is also ever-expanding and not stable. For example, Silverstein (2003) discusses the concept of "other order indexicalities" which means that the indexical field of a given variable is always open for new additions of senses. This means that NNS have a moving target to approach; not only do they have to acquire the social associations of a given variant, but also all derived indices. In short, social indices are "fluid" and are challenging to acquire by NNS (Meyerhoff & Schleef, 2014).

NS may share a degree of conventionalization of meaning across communities of speakers of a language resulting in similar indexicality being associated with the same

variable, such as the social meaning of the velar nasal variant of (ING) (Hazen, 2008) and released /t/ being consistently linked to education or high social status (Podesva, 2010). NNS, on the other hand, may reinterpret these meanings according to their already-established L1 social constructs and meaning preferences.

It is worth mentioning that two other concepts are interconnected with indexicality; these are perceptual salience and linguistic markedness. The salience of variants has been found to be a conditioning factor in the acquisition of variation by NNS (Bayley & Langman, 2004; Major, 2004; Brown, 2011) as well as markedness which is also interconnected with indexicality; the less marked the variable, the sooner it seems to be acquired by NNS (Regan, 1996; Mougeon et al. , 2004; Uritescu et al. , 2004; Dewaele, 2005; Howard et al. , 2006; Drummond, 2011; Hoffman & Walker, 2010) . This interconnection and interdependencies across these three concepts make the task of the NNS very complex.

Another aspect of the complexity of acquiring variation in a second language is heightened by the existence of more than one potential target group and their variety of the target language. A NNS does not only have to identify target variants (Labov, 2001) and their social indexicality (Eckert, 2008a; Meyerhoff & Schleef, 2012; Schleef et al., 2011; Schleef, 2013b), but a NNS must also identify a target group and their language variety as a part of the development of their sociolinguistic competence (Bayley, 2005; Eckert, 2008b). Unfortunately, this link is not necessarily in one-to-one correspondence with the speakers of the dominant variety in a speech community.

Identifying a target group is not merely a matter of establishing an understanding of the vernacular of the majority in a contact setting, because the prevalence of a given linguistic variety does not necessarily render it a target variety for NNS. Some social factors and social attributes of linguistic varieties may render it more appealing to NNS, even if it is a variety spoken by a minority.

For example, Zentella (1997) discussed how experiencing social segregation could motivate NNS to target a minority linguistic variety rather than the dominant variety. Peurtoricans in New York City identified more with African Americans with whom they may have shared a similar experience of marginalization and racism, consequently, they ended up acquiring AAVE norms on variation rather than mainstream American English.

More examples can be drawn from work on NS teenager-communities, where what social group a teenager relates to or aspire to belong to, determines the linguistic variety they end up adapting to. For example, (Eckert, 1989, 1996) showed this for the Jocks and Burnouts in a Detroit high school and for pre-adolescent heterosexual girls in California. Likewise, Mendoza-Denton (1996, 2008, 2014) shows this for teenage Latina girls. Moore and Podesva (2009) make a similar point for Townies and Eden Village girls in the UK and (Benor, 2004) describes similar processes shaping the speech of Orthodox Jewish boys.

This issue is particularly of importance when trying to understand the acquisition of variation by migrant communities. Because migrants may also have similar experiences of marginalization and younger migrants (such as Generation 1.5 or 2 speakers) may feel an additional need to differentiate themselves from their parents' generation of migrants.

This complexity of the acquisition of variation may also be discussed in terms of the type of variables in the L2. Some variables are more complex than others in terms of the number of linguistic interfaces that a variable has, i.e., morpho-phonemic, morpho-syntactic etc. or the number of standard/formal and non-standard/informal variants that a variable has.

This may pose a problem for NNS especially when their mother tongue does not have linguistic distinctions of markedness across variants (Howard et al., 2006; Drummond, 2011; Schleef, 2013a). The literature on the acquisition of variation by NNS implies that the more complex the variable the later it is acquired by NNS (Nagy et al. , 2003; Blondeau & Nagy, 2008; Clark & Schleef, 2010; Brown, 2011; Meyerhoff & Schleef, 2014).

Adult NNS acquiring variation are confronted with a task like that of children acquiring variation and the social significance of that variation in their mother tongue. However, adult NNS lack many of the resources available to NS children, such as the lengthy exposure and feedback from the community. In addition, adult NNS must deal with biological constraints that may limit the plasticity of their linguistic systems. These are summarized in various hypotheses about the critical period (Jiang et al. , 2009; Abrahamsson, 2012).

NS children are attentive to variation patterns that they hear in both the speech around them and Child Directed Speech (CDS), the latter includes the strategic use of variation and hence constitutes an important resource.

Infants' perceptual categories are shaped by the degree of contrast provided by the input of their caregivers (Cristià, 2011, 2013; Cristià & Seidl, 2013), and gendered variants may be learnt from selective input (Babel et al. 2014). Children, as early as infanthood, orient themselves to the local community's norms of speech (Floccia et al. , 2012). When a child's variety is different from the mainstream variety, children, typically, acquire the norms of their peer group including the local patterns (Kerswill & Williams, 2000; Labov, 2014).

Interestingly, evidence suggests that less community-involved children are more likely to conform to patterns of variation of their closed social circles, whereas more-community-involved children produce community-like variation (Kerswill & Williams, 2000). This point illustrates the importance of the role of social networks as a constraint on the acquisition of variation. This point's implications to the acquisition of variation by NNS is that networking may indeed affect how much they approach NS-like constraints on variation.

When children are exposed to variation other than that of their homes, not only do they develop an understanding of the speech norms of their larger communities, but also an

understanding of the social indexicality of that variation. It seems likely that this understanding is achieved before they even reach the very creative period of adolescence (Sankoff & Blondeau, 2007).

Children are indeed sensitive to social evaluations as well as dialect-specific constraints alongside other grammatical ones (Smith et al. , 2009). Adult NNS obviously lack the exposure to all these child-directed resources that help NS children acquire variation.

2.2 The acquisition of variation in migration contexts

I believe it is beneficial to present an overview of the contact setting that I study because I aspire to identify points of similarity and difference across migrant communities worldwide and show how the different circumstances of each contact setting affect the process of acquisition of variation, by focusing on social factors and NNS-specific social factors. I hope that I will eventually be able to draw some generalizations regarding the role of the contact setting in the process of acquisition of variation.

The contact setting that I study is one between two mutually-unintelligible languages where the NS speak New Zealand English and the migrants speak varieties of Levantine Arabic. These migrants to New Zealand have become migrants by choice; they did not suffer forced migration of any type, they were mainly looking for a better future for themselves and their families.

This point is important when I discuss migrant's attitudes towards the target language, the host country and the local speakers because this affects the migrants' social networks and the acquisition of variation, consequently. I believe, it is reasonable to consider that Arab migrants to Wellington share a lot in common with migrants in low-linguistically-diverse contact settings and should, therefore, be placed within the framework of this contact setting.

Another point to mention is the migrants' English language proficiency. All previous exposure to English for first-generation migrants in this sample, prior to arrival to New Zealand,[3] was either through formal classroom-instruction and/or passive contact through multimedia communications and technology.

Recent work in variationist sociolinguistics considers variation in NNS speech in comparison with NS patterns of variation. This means that NNS variation patterns are analyzed to see how far they replicate the patterns of variation found in the vernacular of NS. Although very little is known about NNS rank ordering of constraints when acquiring variation in an L2 (Meyerhoff & Schleef, 2012), and despite the fact that real-time, developmental studies of NNS hierarchies are lacking (Geeslin & Long, 2014), some researchers have suggested that when language users adapt patterns or languages at large, they always impose some kind of change in NS' patterns (Kachru, 1982, 1992).

The change is either in form and/or meaning and is sometimes substantial, in the case of patterns of variation, involving changing the ranking of constraints that condition variation (Bayley, 2005; Meyerhoff & Schleef, 2012). More specifically, Bayley (2005) hints that the reordering of constraints and internal ordering of levels within constraint hierarchies is a typical tendency of NNS with developing grammars.

Schleef et al. (2011) and Meyerhoff and Schleef (2012) suggest that NNS may be "inclined to replicate" some NS constraints on variation, reinterpret some constraints and reject others. NNS may also create constraints that are meaningful to them (Schleef et al., 2011).

In order to study the suggested phenomenon of NNS changing NS constraints, I will be borrowing the concept of "transformation under transfer" (Meyerhoff, 2009a) to refer to

---

[3] excluding younger Arab second generation migrants who were born in New Zealand

all changes NNS make to NS constraints collectively. The term "transformation under transfer" was originally applied to describe the amount of constraint transfer from a source language (English) to an English-based creole (Meyerhoff, 2009a).

Meyerhoff suggested that "[t]his typology considers similarities both of kind and of quality. In doing so, it allows us to ask linguistically and statistically meaningful questions about whether the same surface-level form found in two languages is in fact "the same." It adopts well-established terms from the field of language contact and attempts to relativize them to the kind of phenomena that variationist study."

And she (2009a) proposed a continuum of three types of replication patterns, namely:
1. Weak transfer or replication occurs when the same constraints are significant constraints on a variable in the model and in the replica varieties.
2. Strong transfer occurs when the same constraints are significant in both model and replica, and the ordering of these constraints is the same in both model and replica.
3. Calquing occurs when the same constraints are significant in both model and replica, and the rank ordering of these constraints is the same in both model and replica, and the factors within groups have the same ranking in model and replica (same internal hierarchies).

I hypothesize that the patterns 1, 2, 3 above and the concept of "transformation under transfer" can be stretched to apply to NNS patterns when acquiring L2 variation.

2.2.1 The acquisition of variation in multilingual metropolises

Previous research on variation in contact settings, particularly in multilingual metropolises, suggests that young mobile adults, who live in urban centers across the world, tend to demonstrate broadly similar patterns of variation regardless of their first language. Sometimes they produce structures (lexical items, syntactic structures, phonic variation and prosody) that do not have counterparts in the peer NS target language, i.e., linguistic

innovations (Mougeon et al. , 2005; Wiese, 2009 ; Torgersen et al. 2011; Torgersen &
Szakay, 2012)

For example, some innovations were detected in the patterns of variation amongst
speakers of minority Ontario French in Canada. Such innovations were possibly the result of
first language transfer from English in this contact setting (Mougeon et al., 2005). The
pragmatic marker "you get me" is an innovation typical of multicultural London-English
(Torgersen et al., 2011). Torgersen and Szakay (2012) show that rhythm in multicultural
English is a shared attribute of all ethnicities in Metropolis London.

Another interesting form of innovation exceeds first language transfer and lexical
innovations to describe the emergence of so-called multiethnolects. Wiese (2009) discusses
the emergence of new linguistic constructs in urban, multi-ethnic Europe as she discusses
Kiezdeutsch as a typical adolescent-related, contact-induced variety of language that is not a
mere reduction (levelling) of major ethnic minority ethnolects, but a rather productive system
in its own right. And Wiese (2009) makes interesting suggestions about the emergence of
new particles and complex morpho-syntactic constructs typical of this new variety and claims
that this tendency holds true for multiethnolects across languages and countries.

Furthermore, Britain (2010) suggests that modernity, social mobility and linguistic-
accommodation practices motivate supralocal language levelling in contact settings. Hornsby
(2007) argues that levelling and geographical diffusion are typical of urban centers, like
urban Paris and London, through the spread of the characteristics of a culturally and
economically more attractive variety in a contact setting.
Interestingly, even speakers of mutually-intelligible dialects seem to perform linguistic-form
levelling, particularly for the forms that have stigmatized social indices. For example,
Salvadorian-Spanish migrants to Toronto, Ontario, seem to acquire less marked variables
when several Spanish varieties coexist (Hoffman, 2004).

Despite the previously mentioned attempts to understand the acquisition of variation by NNS, the investigation of the way NNS target NS variation is a developing field of inquiry (Meyerhoff & Schleef, 2012).

Previous research on NNS acquisition of variation has often been conducted in cities like London, Berlin and Toronto where many speakers of many different languages are part of the sociolinguistic ecology. I believe that it is interesting to see if migrants in less linguistically diverse contact settings exhibit patterns of variation similar to those seen in these previous studies, i.e. innovation and levelling. It is possible though that less-linguistically-diverse contact settings will have patterns of variation specific to them.

Before I go into detailed discussion of the literature on this topic, I would like to outline what I mean by low-linguistically-diverse contact settings. I am referring to contact settings where NNS groups are either few, regionally scattered or face forced or self-imposed segregation that is evident in their accommodation and settlement trends, daily routines and jobs. For example, do NNS live in ethnic neighbourhoods, work mainly with NNS within the same ethnic group and have no networks with other ethnic groups and NS in the host country? If this is true, are they less compelled to resort to linguistic levelling or innovation as found in urban centers like London?

This is not to imply that migrants living in less linguistically diverse contact settings are confronted with a task less complex than that facing migrants in multilingual metropolises because, as noted above, the acquisition of variation in a linguistic system is inherently a complex process (Schleef et al., 2011 ; Meyerhoff & Schleef, 2012).

2.2.2 The acquisition of variation in low-linguistically-diverse contact settings

I propose that the "transformation under transfer" of NS constraints on variation by NNS is an inevitable output of the complexity of the process of the acquisition of variation outlined previously, but this pattern is also associated with NNS-specific social constraints as well as the type of the constraints on variation; articulatory, grammatical or stylistic. NNS-specific constraints cover a wide range of issues from identity marking to exposure to the target language and attitudes towards speakers of the L2 and the host country. One recurring social constraint in the literature of acquisition of variation by NNS is ethnicity (Boberg, 2004; Brown, 2011), it is, nevertheless, only one aspect of identity that may affect the process of acquisition of variation.

For example, the role of ethnicity in identity marking, when combined with social isolation, visible minority status and settlement arrangements, may encourage segregation and consequently affect linguistic choices (Boberg 2014). More specifically, Jewish and Italian migrants brought up outside of Montréal demonstrated low, if any, ethnic marking in their variety of English, unlike those who were raised in ethnic neighbourhoods in Montréal.

Hoffman and Walker (2010) measured the correlation between the acquisition of a stable variable and ethnic orientation[4] among NNS. Interestingly, ethnic orientation was only significant as a constraint conditioning the acquisition of variation when correlated with networks that the speakers maintained

The contrast between Hoffman and Walker's (2010) findings regarding Chinese NNS in Toronto and how they tend to divert from the mainstream language variety, and

---

[4] Ethnic orientation (EO) is an operationalization of a qualitative investigation of ethnicity as a conditioning factor that affects variation patterns in multilingual Toronto that was proposed by Hoffman and Walker (2010). For further details see Hoffman and Walker (2010: 46-480).

observations drawn about Asians adopting mainstream American English in the USA (Boberg, 2004; Lo & Reyes, 2004) provide further evidence that neither ethnicity nor social isolation can be discussed independently of one another (and other social factors).

Sharma and Sankaran (2011) considered social isolation in relation to the NNS-specific networks as reflected by the level of Asianness of their networks. The Asianness of the networks was a significant factor predicting transfer of L1 features for first-generation Punjabi-speaking Indians. In their study, these speakers maintained Indian-only networks in the U.K, and that fact directly affected their acquisition of variation as well as causing strong L1 transfer effects. These results showcase the importance of NNS networks as a constraint conditioning variation.

Nagy et al. (2003) studied the role of NNS networks which was differentiated in terms of its orientation (NNS- networks, mixed, NS networks) and how this orientation may encourage or discourage L2 usage and consequently affect acquisition of variation in an L2. If NNS are NS-oriented, they are expected to converge with NS constraints, if NNS are NNS-oriented, they are more likely to diverge from NS-constraints, probably displaying L1 transfer and/or ethnic marking (Schleef et al., 2011; Drummond, 2012; Meyerhoff & Schleef, 2012; Schleef, 2013b; Davydova et al, 2017; Schleef ,2017 a).

Finally, data from taught language learners is more mixed on the effect of social networks (Regan, 1996; Uritescu et al., 2004 ; Brown, 2011; Geeslin & Long, 2014; Sayahi, 2014) suggesting that the effect of networks is more pronounced in more naturalistic, immersion contexts of the acquisition of variation.

To summarize, NNS-specific constraints include social isolation, network involvement and settlement history that have effects on the acquisition of variation. In fact, these factors overlap and result in the differential acquisition of variation (Major, 2004)

There is also a positive correlation between acculturation, simply defined as integration with a host community, and the acquisition of variation in an L2 (Dewaele, 2005). However, Jiang et al. (2009) investigated the role of acculturation on the acquisition of a native-like pronunciation by NNS and their results indicated that acculturation levels only proved significant when the NNS wanted to belong to the host country, this was reflected in their overt attitudes towards the host country, the target language and perceived discrimination (Gambino, 2001; Lybeck, 2002).

Another example is the reported positive correlation between NS-NNS network involvement and acquisitions of variation beyond the categorical use of formal variables which is typical of classroom environments (Howard et al., 2006).

Exposure to the L2, not surprisingly, seems to be relevant in the acquisition of variation. It is, ultimately also, intertwined with social isolation and/or network involvement. For example, there seems to be the threshold for the emergence of stylistic constraints on vernacular variation as implied by the literature (see Drummond 2011; Sharma and Sankaran 2011; Schleef 2017a). NNS who have spent more than three years in a host country are more likely to acquire stylistic variants. Wolfram (1984) suggested that NNS are concerned with grammatical constraints at first but once these are acquired, NNS start to look for ways to sound more local.

Another facet of exposure is the proficiency and linguistic competence of the NNS which usually has a positive correlation with NNS acquisition of variation. For example, developing linguistic competence resulted in a change in factor weights and hierarchies for NNS across different time intervals (Young, 1991; Major, 2004; Howard et al., 2006; Hoffman & Walker, 2010; Schleef, 2013a, 2017a; Davydova et al., 2017). There is, however, room for individual differences in the acquisition of NS-like constraints (Mougeon & Nadasdi, 1998; Geeslin & Long, 2014).

Another way of understanding the role of increased proficiency on the acquisition of variation among NNS is the relationship between increased proficiency and approaching NS-like frequency of occurrence of a certain variable. For example, in studies of CSD, Edwards (2016:144) states that:

"[p]revious studies that have focused on proficiency and contact have found that the lower the proficiency, the higher the deletion (Bayley, 1996). The findings for both VE [Vietnamese English] and HKE [Hong Kong English] confirm this finding. Specifically, as the participants become more proficient in English, they tend to reduce or simplify clusters less, and this then reduces the rate of CSD. It is possible that more proficient speakers of CE [China English] have more exposure to English and therefore, more contact with features of English and that speakers of CE acquire sociolinguistic features of English – including CSD – as they become more proficient. It is also possible that they have changed their norm-orientation. "

A point to mention here is that only a few researchers provide a criterion for measuring proficiency. Some used TOEFL scores (Young, 1991 ;Bayley & Langman, 2004), others resorted to self-reported proficiency which is often tied to age at arrival in the host country (Blondeau & Nagy, 2008; Schleef et al., 2011 ; Schleef, 2013a, 2017a). Sometimes proficiency is indirectly measured through the level of use of the target language and speakers' linguistic history (Nagy et al., 2003 ; Dewaele, 2005; Howard et al., 2006).This means that caution must be exercised when generalising about the effects of proficiency on the acquisition of variation in general.

Finally, attitudes have been found to condition NNS acquisition of variation. Mainly, overt attitudes towards the host country, the sense of belonging to the host country and intention to stay there have been found to be significant constraints on the acquisition of variation by NNS (Schleef et al., 2011; Drummond, 2012; Schleef, 2013a, 2017a).

The literature, to date, is unclear about whether certain types of variables at different levels of linguistic structures or certain kinds of constraints (cognitive, stylistic) are more or less likely to be associated with NNS replication patterns 1,2,3 mentioned earlier (Adamson & Regan, 1991; Regan, 1996 ; Mougeon et al., 2004; Dewaele, 2005; Howard et al., 2006 ; Hoffman & Walker, 2010 ; Schleef et al., 2011 ; Meyerhoff & Schleef, 2014). It seems plausible, however, that constraint type affects the acquisition of NS constraints on variation by NNS, although the exact definition of constraint type is fluid and there seems to be no consensus when discussing the role of constraint type in the acquisition of variation in variationist sociolinguistics literature.

For example, Labov (1989) distinguished four types of constraints: articulatory, grammatical, stylistic and social constraints. Articulatory constraints are both phonological and phonetic in nature and Labov argues that these constraints apply universally[5], whereas grammatical, stylistic and social constraints may be dialect-specific. For example, the stylistic/prestige indexicalities associated with the alveolar nasal as a variant of (ING) are much fewer in some dialects of American English compared with British English dialects.

The implications of Labov (1989) typology of constraint types to NNS is that it might help me identify NNS tendencies and patterns by differentiating dialect-specific (social and stylistic constraints) from universal constraintsa and we can then try to see the patterns NNS have.

To the best of my knowledge, only one study Edwards (2011) tried to integrate Labov's typology to test whether NNS follow patterns like NS-children acquiring variation. I believe that NNS will tend to exactly replicate and partially replicate phonetic/phonological

---

[5] Although articulatory constraints are phonetic by definition and although phonological constraints can be traced to articulatory consideration, these constraints are not necessarily universal.

constraints (articulatory constraints in Labov's terms) that have some universal attributes to them.

Reviewing the literature, evidence implies that NNS seem to be able to replicate articulatory constraints on variation regardless of their proficiency level, the length of stay in a host country and network involvement.

For example, Cambodian and Vietnamese immigrants in Philadelphia and Washington D.C in the United States replicated NS phonological constraints on (ING), regardless of their level of proficiency (Adamson and Regan 1991). Thus, illustrating strong transfer of NS constraints. Brazilian Portuguese speakers managed to replicate constraints in the production of near native-like variants of nasal codas regardless of their proficiency in English and the fact that nasal codas are non-existent in their L1 (Kluge et al. , 2007) illustrating strong transfer of NS constraints. Hoffman and Walker (2010) studied CSD among three generation NNSs of Italian and Chinese descent to Toronto and all three generations of the different language groups could reproduce NS phonological constraints on the variables illustrating weak to strong transfer across generations.

Schleef (2017a) conducted an aparent-time investigation of the constraint hierarchies by Polish migrant teenagers to the U.K. This study examines the acquisition of T-glottalling among teenage migrants in London. Results show that constraint hierarchies based on native input begin to be approached after two years in England as migrants replicated phonological constraints early on in their development. A point to mention is that in phase two, respondents could apply phonological constraints word medially, a position that is more complex, as their proficiency advanced (ibid). In phase three, the respondents produced native like phonological constraints.

Drummond (2011) looked at /t/ release and /t/ glottalling in word-medial and word-final positions by Polish migrants to Manchester, U.K. His respondents could replicate

phonological constraints on less stigmatized variables show-casing strong transfer, i.e., the /t/release and the glottalized /t/ word finally. Drummond (2012) on the same sample, investigated the constraints on (ING) and found that Polish migrants replicate phonological constraints as well illustrating strong transfer.

In studies of the acquisition of French phonemic variables, NNS seem to acquire phonological constraints regardless of their proficiency level, exposure to French and the markedness of the variable and whether a variable is formal or informal. For example, highly advanced learners of French acquired phonological constraints on the production of "ne" deletion, a mildly marked informal variable, even before arriving in France (Regan, 1996 ; Dewaele, 2005) and illustrated strong transfer. Uritescu et al. (2004) investigated the acquisition of schwa deletion by low proficiency immersion students learning French in Canada who replicated phonological constraints too and illustrated strong transfer.

Even when NNS acquire non-phonological variables, phonological constraints seem to be among the first constraints to acquire, which may indicate strong transfer from the L2. For example, NNS acquired phonological constraints in the acquisition of the alternation between overt and null complementizers in Anglophone Montréal French (Blondeau & Nagy, 2008). Bayley and Langman (2004) investigated the acquisition of verbal morphology of English and Hungarian by Chinese students in California and learners with both L1s get the English variation in a similar pattern. The respondents displayed similar trends in the acquisition of phonological and grammatical constraints on the verb marking acquiring phonological constraints and grammatical constraints in both languages, although Hungarian has a much more complex verb morphology system than Mandarin.

As far as grammatical constraints are concerned, NNS can replicate them partially (weak to strong transfer) or exactly (calquing) depending on NNS-specific features. For example, Howard et al. (2006) investigated the acquisition of /l/ deletion by advanced

learners of French and found that they could replicate grammatical constraints, although the variable is informal and such variables usually call for a threshold of linguistic exposure (Drummond, 2011; Schleef, 2017a). Drummond (2012) reported replication of grammatical constraints by NNS and Young (1991) also detected replication of grammatical constraints among advanced learners of English.

An interesting concept that is worth mentioning in tandem with AM's ability to replicate NS grammatical constraints, aside from proficiency, is that of passive contact and what the application of supra-local constraints. Sayahi (2014:120) defines passive language contact as a situation "where speakers are exposed to bilingual or bi-dialectal material without themselves producing similar utterances".

The implications of passive contact to NNS is that NNS are sometimes exposed to L2 constraints without realising that. Schleef et al. (2011) illustrated that NNS display a broader range of grammatical constraints on (ING) than those found in the target variety. For example, migrant Polish adolescents to the U.K acquired a verbal-nominal constraint in the production of (ING) with distributions that do not have counterparts in the varieties of the peer NS groups but that are existent in other varieties of English. Schleef (2013b) investigated the same tendency of acquiring nominal-verbal continuum, but he referred to it as a re-allocation of grammatical category.

Persistence, defined in terms of a tendency to repeat the realization of a prior variant, is also a constraint that seems to be acquired by NNS probably because it has some cognitive associations. For example, Polish migrants to two cities in the UK could acquire priming and phonological constraints on (ING) regardless of their proficiency, exposure to English and the city they moved to (Schleef et al., 2011).

2.2.3 Reallocation, rejection and creation of constraints among NNS acquiring variation in an L2 (weak transfer and created constraints)

The acquisition of non-linguistic constraints on variation is the locus of most reallocation and rejection of NS constraints as well as the innovation of NNS-only constraints. This is presumably because the connection between the indexicality of variation and social categories is inherently indirect (Ochs, 1992; Silverstein, 2003; Eckert, 2008a).

The interpretation of the meaning of social indices relies on the context, expressed identities of speakers (Sclafani, 2009), and their assumed social identity (Eckert, 2008b; Campbell-Kibler, 2012; Mendoza-Denton, 2014).

Meyerhoff and Schleef (2012) suggest that when non-linguistic constraints emerge, they are usually associated with group categories that have the most social value for NNS. For example, Meyerhoff and Schleef (2012) reported the emergence of gender and friendship networks as non-linguistic constraints conditioning (ING) among Polish migrants to London and Edinburgh. Mougeon et al. (2004) and Uritescu et al. (2004) reported the emergence of a social network constraint among NNS in Canada and Sharma & Sankaran (2011) reported the emergence of gender as a significant constraint for female heritage Hindi Punjabi speakers residing in the U.K.

Although Meyerhoff and Schleef (2012) suggested this framework as an explanation for the reversal of some constraint hierarchies, the reported innovations may be reallocations of social constraints within the framework of Silverstein's (2003) discussion of second-order-indexicality.

Several studies of the acquisition of variation in a range of different contexts have found reallocation of constraints as markers of style (Regan, 1996; Mougeon et al., 2004; Dewaele, 2005; Schleef, 2013b). Abrahamsson (2001) and Major (2004) propose that formal style favours more native-like variation. There also seems to be some evidence that NNS L1

stylistic variation can be an influence on the forms they acquire in the target variety. For example, GafarSamar (2000) investigated potential substrate effects from Persian in the acquisition of resumptive relative pronouns in English. Earlier research had assumed the forms found in NNS English were transferred from colloquial Persian (Gas, 1979), but empirical evidence suggested that transfer is more likely from an elevated, formal variety of Persian (GafarSamar, 2000). Moreover, Meyerhoff and Schleef (2014) suggest that Polish migrants to the U.K may use L1 forms as an additional "stylistic resource" to mark formal style in the L2.

Accounts proposed by researchers for the reallocation of style and gender constraints vary. Some researchers have suggested that there are limitations to the accommodation of target group variation (Adamson & Regan, 1991; Howard et al., 2006; Meyerhoff & Schleef, 2012), while others speculate that NNS are trying to get the appropriate form and end up overdoing one variant (Bayley & Langman, 2004). Regan (1996) also cautions that we may assume that NNS have missed the mark, but in some cases, we may be using out-of-date data or inappropriate NS data as the assumed target. Finally, rejection of NS constraints is by far the hardest transfer pattern to prove since the absence of evidence is not necessarily evidence of absence.

Non-native speakers living in multilingual communities need to prove their new affiliations by adapting to their new environments. If linguistic community membership is demonstrated by mastery/acquisition of NS-like constraint hierarchies, then NNS may feel some social pressure to acquire NS constraints on variation (Tagliamonte 2002; Hoffman and Walker 2010) .

Nevertheless, surface frequencies of target variables do not necessarily indicate underlying constraints at work (Guy, 1991; Tagliamonte, 2002). This holds true not only for

migrant communities transforming to new countries, but also to mobile people of various L1 living in metropolis multilingual communities.

In conclusion, I suggest that replication of linguistic and non-linguistic constraints, reallocation of constraints and rejection and innovation of constraints, sum up the overall tendencies of NNS in the acquisition of L2 variation, collectively referred to as a tendency for "transformation under transfer" of NS constraints, but they also seem to illustrate other patterns of acquisition of variation:

1. NNS tend to acquire linguistic constraints before the acquisition of non-linguistic constraints.

2. Within the acquisition of linguistic constraints, weak and strong transfer of the NS' occurs for phonological constraints and sometimes with priming, both of which can be construed as universal, cognitive constraints.

3. Situations characterised by passive contact, supra-local constraints and higher proficiency in the L2 may also result in weak or strong transfer of syntactic constraints.

4. Most cases of reallocation of constraints, rejection and innovation occurs for non-linguistic constraints. This reallocation of constraints sometimes surfaces as a reversal of typical norms of style and gender. For example, we see more non standard variants in the speech of NNS even in most formal scenarioes. Innovations might be 2nd order indexicalities (Silverstein, 2003) or a reinterpretation of the indexical field within important themes for language learners (Meyerhoff & Schleef, 2012).

2.3 The variables of this study

I have studied and analysed the constraints governing the variation of two well-studied variables of the English language; coronal stop deletion (CSD) and (ING) among native speakers of the New Zealand Spoken English Database (NZSED) and Arab migrants in Wellington with a focus on Non-native-speaker specific (NNS-specific) social variables.

There are many reasons why I decided to study these two variables. Firstly, the two variables provide a good control and a point for comparison since they are stable in varieties of native speaker English in North America, Great Britain, Australia and New Zealand and are present in spontaneous speech (Labov, 1966; Trudgill, 1974; Bell & Holmes, 1992; Tagliamonte, 2004; Campbell-Kibler, 2007; Hazen, 2008; Hoffman & Walker, 2010; Schleef et al., 2011; Drummond, 2012).

This means that we have a very good understanding of the constraints conditioning variation for CSD and (ING) among NS of varieties of English. An important point to mention here is that stable variables are not necessarily as stable for NNS as they are for NS of English (Drummond, 2012; Meyerhoff & Schleef, 2012), so they are worth further investigation because they pose an interesting challenge to language learners.

Some previous studies have examined these variables among NNS. These studies provided insights into how NNS acquire pattern of variation (Wolfram, 1984, 1985; Adamson & Regan, 1991; Bayley, 1996; Edwards, 2001; Drummond, 2010; Edwards, 2011; Schleef et al., 2011; Meyerhoff & Schleef, 2012). There is, nevertheless, need for further research in this area.

Both (ING) and CSD have variants that have social indices associated with intelligence, education, high social status and professionalism (the velar nasal and an overt release of /t/ respectively) (Bell & Holmes, 1992; Podesva & Campbell-Kibler, 2002; Romaine, 2003; Benor, 2004; Campbell-Kibler, 2007, Podesva, 2007; Moore & Podesva,

2009; Kendall, 2010; Campbell-Kibler, 2012) . This means that there might be room for creating new indices and/or reinterpreting the NS indices among the Arab migrants to New Zealand, for instance, social status, correctness and education which are important values in Arab communities.

Moreover, there is relatively little research studying stable variation in NZE, compared with attention paid to variables undergoing change (Maclagan & Hay, 2010). To the best of my knowledge, there are only a few studies of (ING) and CSD in NZE (Bell & Holmes, 1992; Holmes and Bell 1994; Bayard & Krishnayya, 2001; Schreier, 2003, 2005).

2.3.1 Coronal stop deletion in English (CSD)

A coronal stop that appears in a word-final consonant cluster is subject to variation in English. This variation is usually discussed in terms of retaining the final consonant or deleting it in words like *west*. CSD is one of the most studied variables in the English language; it has been the object of investigation not only in standard varieties but also in children's developmental varieties and ethnolects. There is also a growing body of research on the acquisition of variation on CSD in contact-induced varieties of English.

Starting with studies of CSD variation in NS varieties of English, Pan-American varieties have shown grammatical category, the following phonological context and the preceding phonological context to be consistent constraints conditioning variation in CSD. Grammatical category is usually reported to have a strong effect with monomorphemes deleting at a higher rate than irregular past tense forms with regular past tense forms favouring deletion the least (Guy, 1980, Neu, 1980; Guy, 1991; Hazen, 2011).

The following phonological environment constraint shows a consistent tendency towards deletion in pre-consonantal contexts, whereas the rates of deletion in pre-pausal and pre-vocalic context are dialect-specific but always less than pre-consonantal contexts. For example, in Philadelphia, U.S.A, pre-pausal contexts are the least favouring of deletion, whereas in New York pre-vocalic contexts are (Guy, 1980, Neu, 1980; Guy, 1991).

As for the preceding phonological context constraint on CSD, non-sonorant sounds seem to favour deletion at a higher rate than preceding sonorant contexts. This constraint seems to have a weaker effect compared with the grammatical category and the following phonological environment in these NS varieties. There is also a slight gender effect with women favouring deletion at a lower rate compared to men (Neu, 1980).

Canadian English seem to share the same constraints on CSD variation with American English varieties (Hoffman & Walker, 2010), showing a sensitivity to grammatical category, following segment and preceding segment constraints.

New Zealand English shares the same constraints found in other varieties of NS English with differences in the level of strength each constraint displays. For example, the grammatical category constraint has a weaker effect, one that is smaller but very close to that of the following segment. Within the grammatical category constraint, regular past tense forms and semi-weak forms delete at a similar rate, both disfavouring deletion. Moreover, (Holmes & Bell, 1994) reported a strong gender-effect with men favouring deletion at a higher rate than women.

NS children and teenagers seem to share the same constraints on CSD variation as their adult counterparts (Labov, 1989; Guy & Boyd, 1990; Roberts, 1994, 1997; Smith et al., 2009). Younger NS not only display mastery of articulatory constraints on variation, but also display a mastery of dialect-specific constraints (Labov, 1989; Roberts, 1997; Smith et al., 2009), and grammatical category appears to be subject to emerge in developmental stages, that is, younger children delete semi-weak past tense at a similar rate to monomorphemes, presumably because their grammatical development of semi-weak verbs as a distinct grammatical class is developing. Relatedly, Guy and Boyd (1990) find that CSD is subject to age-grading.

Although CSD is a stable variable that is conditioned by a stable set of phonological and grammatical constraints, the significance of the grammatical category might be dialect-specific (Temple, 2003; Tagliamonte & Temple, 2005; Guy, Hay, & Walker, 2008; Hazen, 2011).

The major difference between NS norms of variation on CSD and NNS patterns is the rate of deletion. For example, Pan-American Englishes have a deletion rate of around 33% (Guy, 1991), British English 24% (Tagliamonte & Temple, 2005), and New Zealand and Australian English 31% (Holmes & Bell, 1994).

By contrast, contact-induced varieties of English usually have higher rates of deletion. For example, in English based-creoles rates can be 75% (Patrick, 1991, 1999), in ethnolects like AAVE, Tejano and Chicano Englishes rates are around 50% (Patrick, 1991, 1999, Edwards 2016) and in NNS Englishes like Vietnamese English deletion rates can be 71%, Hong Kong English 63% and Chinese English 45% (Edwards, 2016).

Implications from the literature on CSD suggest that not just the frequency of deletion but also the ranking of the constraints and the internal hierarchies of constraints may differ for contact-induced varieties compared with NS varieties of English. For example, AAVE, Chicano and Tejano Englishes illustrate that the preceding segment constraint on CSD may be a stronger, or at least as significant, a constraint, as the following segment (Cohen & Labov, 1967; Fasold, 1972; Santa Ana , 1991; Bayley, 1996). In contrast, this constraint generally has a lower effect size in mainstream NS varieties of English.

Phonological constraints may be the locus of first language transfer which may or may not become a marker of ethnicity or identity at a later stage. For example, the older version of Maori English (19th century Maori English), had a constraint hierarchy for CSD that is different from contemporary Maori English (Schreier, 2003).

Evidence suggests that NNS may share the same constraints on the variable as their NS peers but the factor weights for such constraints may be different also due to L1 transfer (Bayley, 1996; Edwards, 2011). For example, Edwards (2011) shows that the preceding segment is a stronger constraint on CSD than grammatical category for NS of Mandarin learning English in the U.S.A. Her work also illustrates that the internal hierarchy of the

constraints may differ for NNS; she suggests that this illustrates how the preceding segment has room for first language transfer. Her Mandarin native speakers deleted more coronal stops syllable-finally when they were preceded by an alveolar nasal, which is different from English NS. She suggests that this is influenced by the phonotactics of Mandarin and what constitutes some acceptable singleton codas in Mandarin. Hoffman and Walker (2010) also found differential factor weights and orderings that may be due to L1 transfer in the first generations of migrants and ethnic marking in subsequent generations of speakers from the same ethnic heritage groups.

To illustrate this point, one may regard the tendency of NNS to delete or keep the final onset as a result of re-syllabification (see Labov 1997) as conditioned by the first language. NNS may favour or disfavour deletion of the final /t/, /d/ in clusters depending on whether they can form plausible onsets in their L1s. This possibility of re-syllabification is also interesting for my sample because Arabic, unlike English, allows initial /tk/ and /tl/ clusters, so Arab migrants might be less likely to produce CSD deletion with following /k/ and /l/ because they can form plausible clusters in Arabic. If this is the case, it would be suggestive of L1 transfer.

The implication of this argument to my sample of NNS is that I can examine their speech for traces of ethnolectal marking by looking at how dialect-specific aspects of variation are realised by the 1.5 generation of Arab migrants.

Another way to consider CSD as a potential host of ethnic marking is to investigate dialect-specific stylistic constraints on variation among NS of English and see how NNS target these. The idea is inspired by Labov (1989) work on children, where he considered the dialect-specific attributes of CSD, i.e. the role of the following segment in CSD, mainly pre-pauses, and (ING) i.e., the acquisition of stylistic constraints, to find out whether NS children follow a universal trend in the acquisition of variation (by acquiring articulatory constraints,

grammatical constraints and finally social ones), or whether dialect-specific constraints on variation emerge alongside linguistic ones. The results indicated that NS children acquire both articulatory constraints and stylistic constraints (dialect-specific) while grammatical constraints are tied to underlying developing grammars (Labov, 1989; Guy & Boyd, 1990 ; Roberts, 1994, 1997; Schleef, 2013a; Smith et al., 2009). Few researchers have proposed that adult NNS follow an acquisition pattern like that of NS children (Edwards, 2011).

Other studies of the acquisition of past tense marking in English indirectly tackle the acquisition of variation of CSD by NNS. The results from these studies indicate that NNS seem to acquire phonological constraints on CSD variation, replicating the patterns found in the speech of NS (Adamson et al. , 1996).

NNS' acquisition of grammatical constraints is, nevertheless, tied to their length of stay in a host country as well as their social network involvement patterns. For example,Wolfram (1984, 1985) reports that NNS delete monomorphemes at a lower rate compared to NS. He suggests that NNS are more concerned about grammatical accuracy during early acquisition, he also suggested a three-year threshold after which NNS start targeting NS social constraints. Accordingly, one can find out if NNS follow a universal trend in the acquisition of variation and whether adult NNS acquire dialect-specific and stylistic constraints in their L2 variation.

2.3.1.1 Global and dialect-specific constraints on CSD in varieties of English

As we have seen, there are recurring patterns for the constraints on CSD but there is also the possibility of dialect-specific and developmental variability in the ranking of these constraints. Some studies, influenced by the recent move to focus on lexical frequency, have argued that this is the principal, cross-varietal constraint on CSD. For instance, (Guy et al., 2008:53) argue that it is lexical frequency, rather than morphological status, that affects CSD and that lexical frequency in fact "predict[s] much of the morphological effect that has received so much attention in prior studies". Guy et al. (2008) also found that lexical frequency was a highly significant predictor of CSD in their study of early New Zealand English (speakers recorded in the 1940s).

Other researchers, however, have found that frequency effects cannot explain the CSD in their data (Hazen, 2011; Walker, 2012). "It is likely that frequency effects are mitigated by other variables and that differences exist across varieties of English in terms of how frequency effects impact CSD" (Edwards, 2016:167). Moreover, earlier studies of CSD among NS of English have long recognized that deletion occurs more in high-frequency words, and therefore they are often omitted such as *just*, *and, went* for the speakers of American English, and New Zealand English (Neu, 1980; Holmes & Bell, 1994).

2.3.2 Coronal and velar nasal alternation in (ING)

The variable (ING) is one of the most studied variables in sociolinguistics and variation research and it is one of the most stable variables synchronically and diachronically, with consistent social evaluations. This does not mean, however, that the variant is not complex, especially from a second language acquisition of variation point of view, where research suggest that (ING) is not as stable for NNS as it is for NS (Drummond, 2012).

Even younger NS take longer to acquire this complex variable compared with other variables (Roberts, 1994). Hazen (2006:583) notes that "(ING) is one of the consummate sociolinguistics variables. Its complete story involves both diachronic and synchronic variation, both internal and external factors".

The main two variants that are usually studied for the (ING) variable are velar (coronal) and alveolar variants of the final nasal segment and the alternation between these two variants is a stable pattern in most dialects of English (Wagner, 2008). Nevertheless, (ING) variants are not simply velar as opposed to alveolar nasals in unstressed syllables, other possible variants exist in dialects of English with varying distributional patterns across speech communities (Wagner, 2008).

For example, Canadian English has a third potential variant, the tense alveolar (Woods, 1978), that is also available in some dialects of British English (Houston, 1985). Other British English dialects introduce a fourth variant [ɪŋg] (velar nasal plus stop) (Wells, 1982). This describes the variant [ɪŋg] where words like *during* and *watching* are realised with the same pronunciation as *Linger* and *singer,* i.e., with a velar stop after the nasal. Australian and New Zealand varieties have an [ɪŋk] variant (Shopen, 1978; Bell & Holmes, 1992; Gordon, 1998) as do some varieties of (mainly southern) British English (Wells, 1982; Meyerhoff & Schleef, 2014).

Studies of (ING) variation in NS varieties of English find consistent linguistic constraints that condition variation – these are grammatical, phonological, social and stylistic. The grammatical constraints on the variable (ING) are best explained by the Labovian nominal-verbal continuum; where nominal categories favour the velar nasal variant and more verbal categories favour the alveolar nasal variant (progressive – gerundial participle – adjectives – gerunds and nouns) (Labov, 2001). This constraint seems to be consistent across varieties of NS Englishes: American English (Labov, 2001, Tamminga 2014), British English (Tagliamonte, 2004) and New Zealand English (Bell & Holmes, 1992).

The following phonological context and the preceding phonological context are also linguistic constraints that condition variation on (ING) (Houston, 1985). The sound segment preceding the (ING) variant dissimilates from it for ease of articulation in a process known as progressive homorganic dissimilation. Conversely, the realisation of the final nasal seems to converge with the initial segment in the following words. This process, known as regressive homorganic assimilation, seems to comply with universal principles whereby adjacent sounds assimilate according to place for ease of articulation.

Social constraints and stylistic ones have a constant effect on conditioning (ING) in varieties of NS English. Socioeconomic status is the most recurring constraint that conditions variation, with more alveolar pronunciations as one goes down the social ladder. Not surprisingly, the social evaluations associated with this variant, the alveolar nasal, are usually negative (Labov et al., 2006; Campbell-Kibler, 2007).

Social evaluations of a variant and its social salience are interconnected. For example, the (ING) variable is socially salient in North America, but the variable lacks that social prominence in the UK (Levon and Fox, 2014, Schleef et al. 2016).

Style-shifting is also a conditioning constraint with casual and less formal stances favouring alveolar variants. Age-grading is also a constraint on (ING) (Labov, 2001; Wagner, 2008), but this tendency needs to be discussed in tandem with the social and stylistic associations of the variants. In North Carolina in the U.S, for instance, there are similar rates of the alveolar variant across different age cohorts, consistent with the idea that it is not the variant that matters, but rather its social associations (Forrest, 2015). Where there is age-grading for (ING) the non-standard variable is mostly used by adolescents and least used by middle-aged members of a speech community, this adolescent peak of non-standard usage decreases as speakers leave their adolescence (Wagner, 2008).

New Zealand English displays the same linguistic and social constraints on (ING) that are found in other varieties of NS Englishes (Bell & Holmes, 1992). More specifically, the grammatical category constraint follows the Labovian continuum with nominal categories favouring the velar nasal and verbal categories favouring the alveolar nasal. The following phonological segment constraint has a strong effect in this dialect of English; when the following segment is either a pause or a velar, the velar variant is preferred. If the following segment is a sibilant fricative or a vowel, there is a preference for the alveolar nasal, with other consonants having a lower effect in predicting the (ING) variant (Bell & Holmes, 1992).

The preceding segment is also a strong constraint on (ING) in NZE, and there is also a strong gender effect, such that women use fewer alveolar variants compared to men who had double the rate of the alveolar nasal /n/ compared to women (Bell & Holmes, 1992). Socioeconomic status also patterns with other varieties of English, while NZE only diverges from other dialects of English in that speaker ethnicity does not seem to affect (ING) variation (Bell & Holmes, 1992).

Labov (1989) suggests that children acquire dialect-specific constraints on variation alongside articulatory constraints while they take longer to acquire grammatical constraints. In the case of (ING), children as young as 6 years of age notice and conform to stylistic constraints in their speech communities (Labov, 1989). Similar results were found among younger children (3-4 years old) who had, for the most part, mastered the process of variation of (ING) (Roberts, 1994,1997). As children grow up and interact more in their speech communities, they become more acquainted with stylistic constraints on variation, and older children apply style-shifting constraints at a higher rate compared with younger children (Roberts, 1997; Kerswill & Williams, 2000)

A point to mention though, is that this apparent acquisition of style-shifting may be affected by the fact that children do seem to acquire the alveolar nasal before the velar nasal (Grunel, 1987), and therefore, younger children's apparent inability to completely match community norms for style-shifting, may not necessarily be a reflection of the (non-)acquisition of social evaluations, but may rather be related to the later acquisition of the velar nasal, the one that is usually associated with formal, more careful speech. Nevertheless, constraints on (ING) are acquired at a slower rate than CSD by NS children of English in Philadelphia U.S.A (Labov, 1989).

High frequencies of the non-standard variant for the variable CSD are usually associated with contact-induced varieties, and ethnolects usually have higher rates of CSD than NS varieties. However, higher rates of the non-standard variant of (ING) are not necessarily always associated with ethnolects. Presumably, this is because the non-standard, alveolar variant of (ING) is not favoured by SLA processes of simplification.

It is true that high rates of alveolar variants are found in ethnolects of the U.S.A., such as in AAVE (Kendall, 2010), but some NS varieties of English that are not assumed to have a period of language contact in their history are also almost exclusively dominated by the

alveolar variant. These include Southern American dialects (Wolfram & Christian, 1979; Laturnus et al., 2016), Northern British English and Scottish English. Labov (2001) claims that the latter uses the alveolar variant even in the most formal situations.

NNS seem to acquire the same constraints on (ING) as their NS peers, but sometimes they tend to reorder and reinterpret the constraints. Vietnamese and Cambodian L2 learners of English in Philadelphia and Washington D.C. tend to reorder NS constraints on (ING) even for the constraints with the strongest effects in NS varieties.

For example, NNS seem to reorder grammatical constraints and they do not fully replicate the details of the NS constraints based on the part of speech, they only replicate the extreme ends of the nominal-verbal continuum. Adamson and Regan (1991) suggest that this is either due to migrants replicating what they most hear or the learnability of forms; where "frozen forms" that are members of a small set are easier to acquire than those generated by various rules (say progressive, participle and gerunds). They also found that migrants reordered the effects of gender and style. For example, male migrants preferred alveolar nasals even in more monitored speech, perhaps because NNS male respondents want to sound more like male NS. However, NNS converge with NS in the effect of the following segment.

Schleef et al. (2011) compared the (ING) variable in the speech of local and Polish-born adolescents in Edinburgh and London and found that the Polish-born groups in both cities replicated the nominal-verbal constraint on (ING) more than the locally born teens did. The authors suggested that the speech of the Polish-born adolescents is influenced by supra-local constraints and not just by the local constraints modelled by their locally born peers.

Even within the most constant constraint on variation in (ING); grammatical conditioning, some dialects seem to display a nominal-verbal continuum slightly deviating from the Labovian classic hierarchy (Forrest, 2015). In particular, compound nouns ending in –*thing* seem to have a dialect-specific effect for the (ING) variable. For example, in Southern

dialects of the U.S, the alveolar nasal is almost exclusively present in *everything* and *anything* (Houston, 1985). This pattern is explained by suggesting that the ING-syllable in *everything* and *anything* receives secondary stress whereas *nothing* and *something* are dissyllabic (ibid).

Gordon (1998) suggested that all *-thing* compounds favour the velar nasal because these are not realisations of (ING) but a suffix *-thing* that has overlapping variants with (ING) in NZE.

The strength of the following phonological context also appears to be dialect-specific. For example, in Southern dialects of the United States, the following phonological constraint has a strong effect (Houston, 1985), whereas, in other dialects of the U.S., there seems to be "no strong phonological conditioning before the following velars or apicals" (Labov, 2001:87)

Other differences between NS and NNS in the significance, strength, and ordering of constraints are explained by the idea of imperfect learning which is argued to be reflected as the tendency to reorder constraints, especially social ones (Meyerhoff & Schleef, 2014). Drummond (2012) investigated the acquisition of (ING) constraints by Polish migrants to Manchester, U.K. and his results support those of Meyerhoff and Schleef (2014).

Both studies showed broadly similar patterns for grammatical category and preceding phonological context to those found in NS studies. Nevertheless, not all the statistically significant constraints exhibited the same internal patterns as in the NS data; there is a tendency to reverse effects in social class, gender and style by NNS.

An example is how linguistic constraints vary regionally. In London, for example, the grammatical category is not a constraint on NS' production of (ING), whereas it is so in Edinburgh. This may result in a reallocation of constraints which can, in turn, involve "transformation under transfer" (Meyerhoff, 2009a). Schleef defines this as a reallocation of the relative importance of variable input constraints in the output variation. This reallocation

seems to be more likely with some kinds of variation like categorical constraints which are easier to acquire than gradient ones (Schleef, 2017a).

The emergence of non-linguistic constraints in NNS realisation of (ING) may be associated with factors that are socially meaningful within the social networks of the NNS (Schleef et al., 2011; Meyerhoff & Schleef, 2012), this might in part explain the innovation of some constraints that are non-existent in NS peers. If the NNS are living and working within only NNS networks, then what is of social significance will be what is relevant to what the group perceives as important.

For example, Benor (2004) shows that Jewish yeshiva boys tended to produce more velar nasal variants of (ING) than their female peers. Benor explains this reversal of the typical gender pattern for (ING), in light of the association between formal speech and education, demonstrations of which are more appreciated in boys in the Orthodox Jewish community in question. So here the value of the velar nasal is that of education and not femininity.

2.4 Summary and Objectives of the study

In this chapter, I have reviewed all the related literature of the acquisition of variation by NNS in several contact settings, highlighting some repeated patterns that are typical of these settings. For example, we have seen that young mobile people living in urban centres like Paris and London seem to share patterns of variation. We have also seen evidence that NNS living in less ethnically dense contact settings have other patterns of variation.

I propose that AMs to Wellington may share some patterns of acquisition of variation with the second type of NNS outlined above. I have also tried to outline the dialect-specific aspects of variation for the two variables of the study as a source of understanding whether NNS are sensitive to dialect-specific aspects of variation or whether they are driven by universal aspects of some of the constraints on variation as well as cognitive aspects.

I then introduced the two variables of the study, CSD and (ING), highlighting why they are a very good choice to assess the hypotheses raised in this chapter about the acquisition of variation by NNS in contact settings.

Having reviewed the social context for this study and the two linguistic variables that will be the focus of investigation, we can recap the primary objectives of the study. This study is designed to investigate the following questions:

1. What are the linguistic and the non-linguistic constraints that condition variation in the production of (ING) and CSD among NS of NZ English in Wellington?

2. What are the linguistic and the non-linguistic constraints that condition NNS variation in the production of (ING) and CSD by Arab migrants in Wellington?

3. Based on the results for (1) and (2), is there any evidence for weak transfer, strong transfer or calquing in the acquisition of variation by NNS (Meyerhoff 2009a)? Is there evidence for a tendency towards "transformation under transfer" (Meyerhoff 2009a)?

The next chapter introduces the methodology that I applied to study the acquisition of variation by Arab migrants to Wellington. I first introduce the benchmark of NS constraints on CSD and (ING) as suggested by the New Zealand Spoken English Database (NZSED). I then introduce the NNS database and illustrate how far they replicate NS patterns of variation on the variables CSD and (ING). I also illustrate data collection and data analysis tools and conclude the chapter by discussing statistical concepts applied in the analysis.

Chapter 3: Methodology

The acquisition of variation by AMs in Wellington is the major focus of my research project. The approach most recently applied by variationists to study the acquisition of variation among NNS assesses how far NNS replicate patterns of variation found among NS of the target variety. In other words, we first need to create a benchmark of the constraints on variation across the target group of NS and then compare patterns of variation found among a NNS group to judge the degree of acquisition of variation among such groups.

To establish a benchmark for Wellington English, I utilised the New Zealand Spoken English database (NZSED)[6]. Then, I extracted the speakers from the Wellington region that were recorded in early 2000s and assumed that these speakers are indicators of the patterns of NS variation in CSD and (ING): (i) the overall frequencies of variants, (ii) the significant and the non-significant constraints on variation, (iii) the rank ordering of the constraints and the associated strength of effects, (iv) the internal hierarchies of the constraints and (v) dialect-specific aspects of variation.

The initial step towards understanding patterns of acquisition of variation among AMs is to embrace similar assumptions about variation across NS and NNS when analysing the NNS corpus. The first assumption is that variation is an integral part of our linguistic competence and its capacity is the brain and not in utterances; such knowledge is highly structured and rule-governed even for NNS acquiring variation in their interlanguage (Schleef et al., 2011).

Secondly, the utilization of variationist methodologies presupposes that the knowledge of linguistic and non-linguistic constraints on variation is highly abstract and subconscious; native speakers and non-native speakers of a language find it equally difficult

---

[6] The structure of NZSED and methods used for making the recordings are detailed in Warren (2002).

to voice the reasons behind a variant's choice in any given context. Both NS and NNS are attentive to the frequency of sociolinguistic variables, their associated social indices and dialect-specific constraints on variation (Labov, 1989; Edwards, 2011).

The data collection and data analysis methods for variationist sociolinguistics target the vernacular as its source of data by collecting naturalistic data via sociolinguistic interviews. The naturalistic data is then transcribed either auditorily or acoustically and the data can then be coded for potential constraints that condition variation for a variable using a software like ELAN (Nagy & Meyerhoff, 2015).

Words including the target variable are then extracted and the resulting data is converted into a spreadsheet that is compatible with the statistical tools that will be used to analyse the data. After data cleaning, regression analysis can be used to find the constraints that condition variation for a variable as well as details of constraint hierarchies and strength of effects of constraints are calculated to create an understanding of variation for a variable in a linguistic variety.

3.1 Data collection

Since I needed to create a benchmark for the constraints on variation among NS of NZSED, I had to use two data sets in the form of two corpora. One corpus represents the NS speech, and the other represents NNS speech. I used the New Zealand Speaking English database (NZSED) (Warren, 2002) as the NS corpus and I have recorded and compiled a NNS corpus of Arab migrants living in Wellington.

3.1.1 New Zealand Speaking English database (NZSED): Native Speaker corpus

The New Zealand Speaking English Database (NZSED) is an equivalent of the Australian National Database of Spoken Language, i.e., ANDOSL (Warren, 2002). This corpus comprises linguistic data from major New Zealand cities. Wellington is represented by a sample of 72 self-identified Pakeha and Maori speakers of both sexes subdivided into three age groups (18-30, 31-45, 45-60) which were categorized as young, middle-aged and old, respectively. The speakers were audio recorded performing a variety of activities that may be regarded as corresponding to varying levels of formality.

For each New Zealand city, including Wellington, the corpus comprises three main components. The first part targeted the vernacular of New Zealand English, being the least monitored speech style of all. This part was two-fold encompassing a theme-based conversation and a map task. As for the theme-based conversation, the researchers utilized a theme prompt; the drink-and-driving ads that were regularly streamed via television and radio at the time of the corpus collection (the early 2000s). A researcher introduced the respondents to the theme prompt and facilitated the conversation when needed.

This naturally resulted in recordings that are less natural than unstructured, free conversational speech but the tone of the conversations is reasonably consistent. One issue, that I will return to shortly, is that the speakers in the Wellington-based corpus rarely deviated from the topic, with an average conversation length of ten minutes.

The second part of the conversation corpus is a map task, which included turn-taking in the description of a path to a location (Brown & Yule, 1983). Two respondents have different maps with different signposting and landmarks and the task involves navigating through the different maps. The third part included a sentence reading task and minimal pair reading task.

Another point to mention about the Wellington respondents is that they are not entirely a randomized selection of the larger population; the recordings revealed prior acquaintances and friendships between the respondents. I created a spreadsheet with information about the respondents. This included demographic information as well as information about the recordings such as their length, quality and whether the respondents were at ease talking to each other.

Figure 3.1: NZSED demographic information
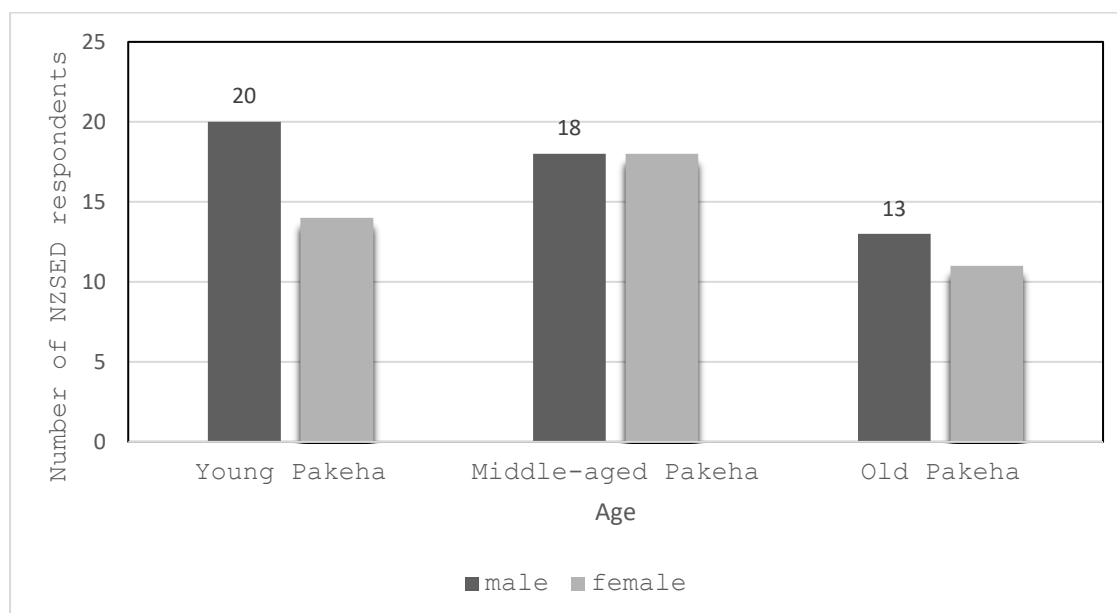
1.1.1.1 Utilizing the NZSED

The audio files were converted into the .wav file format, and when necessary, the file quality was cleaned using Audacity software (2012). I then transcribed all the recordings for the drink-and-driving conversations for the Wellington sample in ELAN. I excluded any samples with Maori speakers because Maori New Zealand English represents a distinct variety of NZE and the Arab migrants would have been exposed to it very little in Wellington.

I used the original NZSED text files, when available, to double check my transcription, as a preliminary phase, and then had a native speaker double check my transcription. I provided a transcription protocol to ensure consistency and comparability of transcripts (see appendix).

I also transcribed the sentence-reading task for the Wellington speakers (200 sentences), eliminating the sentences that had no (ING) and/or CSD tokens. The remaining sentences provided valuable context for variation in the conversational data. This context was also helpful for checking whether AM's patterns of style-shifting were similar or different from NS patterns of style-shifting (on a casual-careful continuum).

I chose to work with this NZE corpus because it is convenient for my personal circumstances as well as research interests. The NZSED was collected relatively recently (2002), this places it within a timeframe that coincides with the arrival of the last significant wave of Arab migrants to Wellington[7] as well as being representative of local NZE. The NZSED corpus is also controlled for demographic information as well as covering a range of tasks of varying formality levels (style-shifting).

---

[7] Excluding the latest wave of Syrian refugees to New Zealand (forced migration) who have continued to arrive since 2014.

3.1.2 Non-native speakers: AM corpus

According to the NZ Census (2013) and New Zealand Statistics (2013), the Arab ethnic group comprise less than one percent of New Zealand's population, and most of them reside in Auckland, with the rest scattered over Wellington and Dunedin.

The second phase of my research was to create a NNS corpus by conducting interviews with Arab migrants (AM) to Wellington. Since I reside in Wellington, it was convenient for me to network and build rapport with the Arab community in the greater Wellington region.

I utilized various methods to begin recording in the Arab community in Wellington starting with an intensive investigation of the community. I also consulted the New Zealand Census to understand the demographic aspects of the community and their counts and clustering patterns in Wellington. I then started to introduce myself to the community. I tried using formal brokers defined as people with institutional authority and power, such as religious clergymen, and informal brokers like older people in the AM community in Wellington.

I started to attend the weekly sermon at a mosque in Kilbirnie, a suburb in Wellington where Arabs cluster, and approached the sheikh of the mosque for help. The sheikh was helpful and introduced me to some women who did community service, teaching New Zealand-born Arabs and Muslims the Arabic language since the Quran is written in Arabic. I introduced myself as one of the community and then offered my help because I have a linguistic background and can help the women choose the best ways to teach Arabic and literacy as well.

Although this arrangement seemed useful at first, it was not. I was always told to wait for the next meeting or the next class and ended up with no respondents whatsoever. The reason for this might be the fact that most of these women-volunteers were relatively older

and not comfortable with being interviewed in English. Nevertheless, I managed to create a network of well-connected people who introduced me to AMs outside the mosque circle.

Formal brokers were not the best way to reach out to the AM community; although people would usually initially agree to participate in my study, they seemed to withdraw when the broker was not there. Accordingly, I tried to enter the community through informal brokers, like older respected people in the community, given that Arab culture respects seniority and experience and these elderly groups represented both.

People would initially agree to participate in my study but would withdraw later. Eight months in, I eventually established that the best way to gain access to participants for this community is to become one of them. To do that, I became friends with a shop owner in Newtown, a suburb in Wellington that has a high density of Arabs.

I would go to his shop every other day for a chat, I then offered to help him in the shop and he accepted. After a while, I became familiar to most of the AMs in Newtown and I started to introduce myself informally through friend-of-a-friend networks. I also started to go to a café in Miramar, another suburb where Arabs cluster in Wellington. People living there started to be familiar with me, and they started to initiate conversations with me.

Eventually, I started going to a coffee shop in Wellington CBD, located on one of the most popular streets and night life in Wellington, Cuba Street. The café is the only one that serves shisha in Wellington, so I used to go there, and smoke shisha and people would approach me and say, "you must be Arab".

My research started to take an ethnographic nature and slowly but steadily, I became a member of the Arab community in Wellington. This access provided me with qualitative data that I would have never, otherwise, acquired and which informed my research decisions in terms of theory, methodology, analysis and interpretation.

I gained a deep understanding of the social structure of the community and by association an understanding of the unique social factors for this group of AMs that sets them apart from other Arab migrant groups in New Zealand, particularly, in terms of network patterns and experiences in English. **Appendix D** illustrates the key social factors that will be discussed shortly in terms of their effect on the acquisition of variation among this group of migrants in Wellington.

This set of social factors for this group of AMs, together with statistical techniques informed my decision to split this AM group into two separates second language speaker communities of practice and to discuss the results accordingly. I then started to consider if this set of social factors would engender identities that might surface as linguistic choices, in this case, patterns of variation.

Since my interest in the patterns of variation found among AMs is clearly from a practice theory approach to sociolinguistics, it was only logical to try and find a social construct, a label, that I can use to address the AM group, that is compatible with this view of variation as a social practice.

I have, therefore, decided to employ the concept of the community of practice as outlined by Meyerhoff (2002) where a community of practice is an aggregate of people who share a mutual engagement in some common ways of practices and these practices involve the construction of a shared orientation to the world around them and an identity of their community of practice amongst others.

This operationalization of the concept of the community of practice does not only comply with a social practice approach to variation but is also compatible with current approaches to identity in sociolinguistics where "identities emerge in practice, through the combined effects of structure and agency" (Bucholtz 1999 :209).

Then, I started to break down the concept of the community of practice to its components as outlined in Meyerhoff (2002): (i) mutual engagement by members (whether harmonious or conflictual); (ii) the sharing of some jointly negotiated enterprise (of a relatively specific nature); and (iii) the existence of a shared repertoire (linguistic and other). I then made sure that the AM community in Wellington satisfy these components and they belong to the same Arab migrant community of practice. I will discuss shortly that despite belonging to the same AM community of practice, this group of AMs belongs to two different second language speaker communities of practice.

More specifically, the AM sample from the Levant form a sub-sample of Arabs in New Zealand, and they shared many practices and were equally engaged in these practices. My sample of AMs belong to an Arab community of practice, in the sense that everyone I interviewed is a member of families that roughly arrived in New Zealand at the same time (the mid-1990s), and now live near each other in Newtown and Kilbirnie. Crucially, they participate in shared practices such as coming together for social gatherings, observing religious and ceremonial activities and events, and actively maintaining friendship ties.

The fact that my sample comes from the Levant brought them closer together as well as sharing the same plight of living in diaspora. The men of this community are engaged in a closed circuit of work and contacts; most of them own their businesses or work as taxi drivers. This fact has brought their families even closer and the wives and children of these men have become very good friends.

The women in this community have also formed affiliations of their own. These mainly involve getting together every other day or so to go together to the same English classes, and they even take their children to the same kindergartens and schools.

Because of the mutual engagements of their parents, the children in this community became active members of the community through association, drawn in by their parents'

networks and the shared practices of Arabic-language school and their mothers' social gatherings.

My sample consists mainly of Muslims who share many religious practices that entail getting together at least once a week such as for Friday prayer. Ramadan and the Eid holidays are also examples of engagements that everyone in this specific community will be equally involved in.

A point to mention here, however, is that one person may be involved in more than one community of practice. So, a person might be involved in various communities of practice, depending on his/her varying interests and affiliations, he/she is then capable of assuming various identities by shifting between the communities of practice. Some of these identities are inherent whereas others are embarked into. For example, familial, ethnic and religious communities of practice are ones that one is born into.

I do not believe that the older members of the community have multiple memberships in many communities of practice, but younger AMs certainly do. All younger Arabs that I have spoken to confirmed having some sort of a strong connection with their origins but at the same time, they all identified themselves as Kiwis i.e., New Zealanders, whereas their parents mainly categorized themselves as Arab New Zealanders or even just Arabs. We will see in the results chapter that although my sample do belong to the AM community of practice, they have memberships in separate second language communities that result in different variation patterns.

The question that arose here was whether the memberships in different second language communities and the shared membership with the AM community of practice entail the negotiation of different identities for the AMs and whether these identities are expressed through variation patterns.

In other words, do members of the two second language communities of practice use variation as an act of identity? This proposal seems perfectly plausible since variation is indeed employed as a resource to signal identities among various communities of practice. For example, Eckert (1989, 1996) showed that separate communities of practice; Jocks and Burnouts in a Detroit high school and pre-adolescent heterosexual girls in California expressed their affiliations and identities through the use of variation. Likewise, Mendoza-Denton (1996, 2008, 2014) shows this for teenage Latina girls. Moore and Podesva (2009) make a similar point for Townies and Eden Village girls in the UK and Benor (2004) describes similar processes shaping the speech of Orthodox Jewish boys.

Different identity practices and the associated affiliations can be achieved by employing different linguistic resources such as variation. To illustrate this identity work, I adopted the concepts of positive and negative identity practices (Bucholtz 1999) to explain the variation found among the separate AM-second-language-speaker communities of practice in the discussion and the conclusion chapters of this thesis.

Specifically, negative identity practices define a rejected identity; what a person does not wish to be affiliated with: Jocks and not Burnouts and visa-versa, Eden girl and not a Townie and visa-versa, Kiwis and not Arabs, or Arab-New Zealander and not Arabs. Positive identity practices, on the other hand, define what a person wishes to pass as, a welcomed and desired affiliation: a Latina, a Kiwi, an orthodox Jewish boy etc.

I also propose that members of communities of practice may utilize a mixture of positive and negative identity practices to select/emphasise preferred affiliations and, at the same time, reject other affiliations that may be imposed on them sometimes, especially for ethnic minorities. Young AMs, we will see, express their identities both as kiwis and not migrants, whereas older and middle-aged AMs express their identities as Arab-New

Zealanders but not quite Kiwi. **Appendix E** illustrates some quotes from the AMs'

conversations that reveals the kind of qualitative data that I have based my conclusions on.

Familiarity with the community of practice, gained in this kind of natural manner, is an advantage that would most likely result in more relaxed data, an understanding of the cultural constraints on making contacts as well as establishing an understanding of sociolinguistic differences within the core Arab community in Wellington. I accepted the social role assigned to me by the target community of practice and the concomitant identity work as a member of the community who has expertise in language learning and linguistics and who has a responsibility towards the community in Wellington.

Interviewing people became more like chatting with friends over a cup of tea, which was excellent because I needed to get to the vernacular of this community. Older and middle-aged men were the hardest to approach, most likely because they are busy with their jobs or because they did not feel comfortable speaking to a thirty-year-old woman with their wives hovering around. Conversely, younger respondents were very interested in my research and they helped me by introducing me to their friends.

I ended up with a sample representing Arabic speakers from the Levant[8] because they are more numerous compared to other Arabs in the Wellington region and their substrate varieties of Arabic are more similar than e.g., Egyptian Arabic, Yemeni Arabic, Saudi-Arabian Arabic because these other varieties of Arabic have different phonetic and phonological profiles with several phonemes that do not occur in the Levant. It seemed prudent to try and include participants with reasonably homogeneous L1 substrate effects since my study is focused on phonological variation.

---

[8] The Levant is a collective noun that refers to citizens of Jordan, Syria, Lebanon and Palestine.

3.1.2.1 Sampling techniques in my corpus

Once I established a network within the Arab community in Wellington I started to consider a sampling methodology. The most commonly used approach for sampling in sociolinguistic research is a mixture of a random sampling approach and a social network approach (Milroy, 1980; Milroy & Gordon, 2003; Tagliamonte, 2006) via judgement sampling. The social network approach mitigates the observer's paradox while judgement sampling with (3-5) people per cell is the acceptable minimum number of respondents for each social factor that the researcher intends to test (Tagliamonte, 2006, 2012). I, therefore, set a goal of a total number of respondents to comply with that norm.

I have recruited the AM sample in line with Milroy and Gordon's (2003) judgement sampling technique seeking respondents to fill a quota for previously identified social constraints which I expected might have a significant effect on their acquisition of local patterns of variation, principally gender and age. My sample included AMs to Wellington who have low mobility within New Zealand and who arrived in New Zealand no later than 1999.

To provide a sample of a AM community comparable with other studies, I conducted a survey of research in the subfield of acquisition of variation by migrant communities to ensure the comparability of my results with the findings of other variationist sociolinguistic research, and I ended up interviewing 21 AM respondents.

Figure 3.2: AM's demographic information
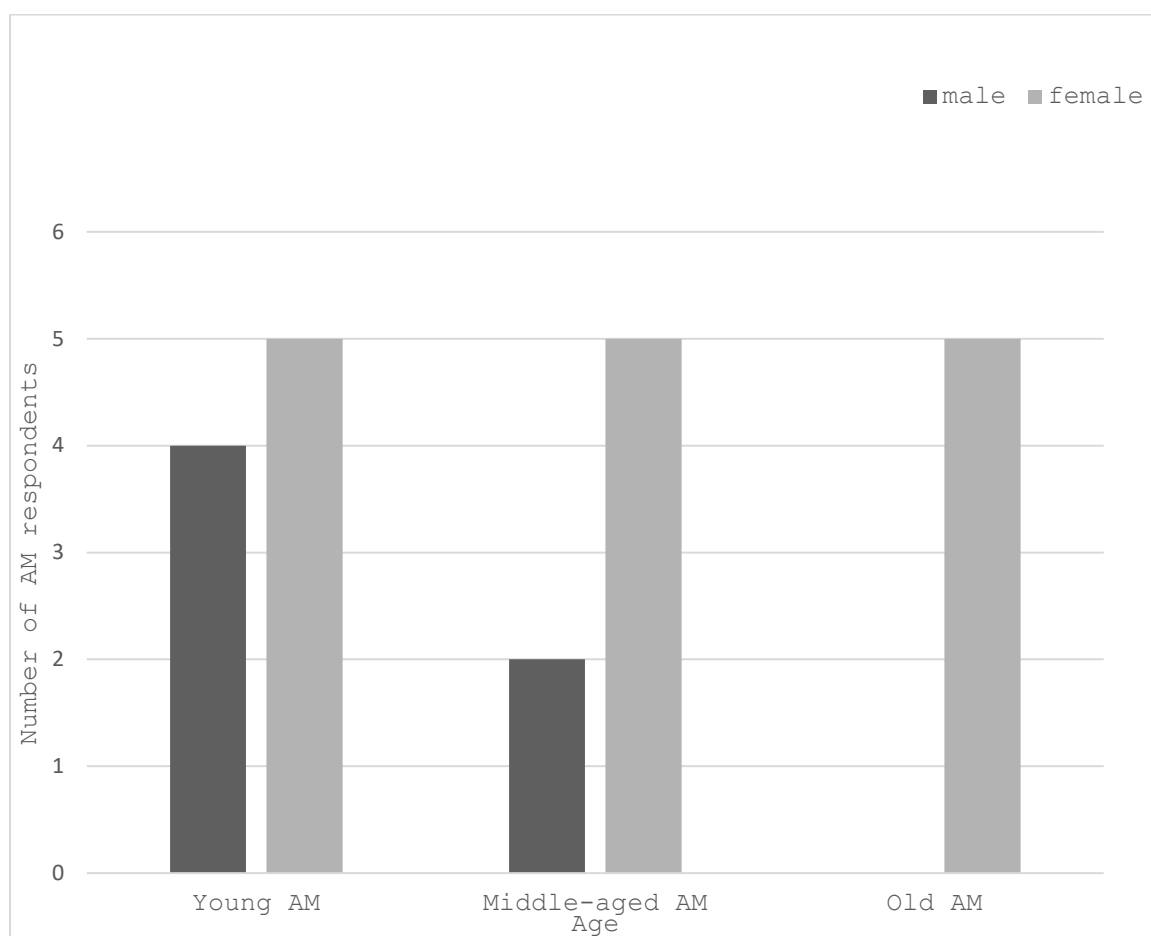
3.2.2 The sociolinguistic interview

I used the sociolinguistic interview as a means of collecting my AM data because it is the most fruitful and efficient tool to elicit natural speech data (Labov 1984; Meyerhoff et al. , 2015). I have only recruited people with whom I formed strong ties and whose homes I had visited at least twice, before conducting the interview. The interviews lasted for an average of an hour and, in line with standard sociolinguistic practice, the setting of the interviews was whatever is most comfortable for participants, usually their own homes.

The goal of the sociolinguistic interview is to elicit a range of contextual styles along a continuum of formality (style-shifting) involving different realizations of the focus

variables (Labov, 1984). Labov (1972) distinguishes five contextual styles based on the level of attention speakers pay to their speech. These are reading minimal pairs, word lists, passages and casual speech; the latter is the most spontaneous, systematic non-careful speech. I initially attempted to collect data with different styles and designed a paragraph reading task for AMs, but most of them declined to do any reading aloud tasks.

I utilized the concept of the interview modules, where the term "interview" is the exact opposite of the final goal, which is to evoke the vernacular and not a formal string of questions and answers (Labov, 1984; Tagliamonte, 2006). Conversational modules are pre-prepared, separate potential themes or conversation prompts for discussion that hold together logically.

The researcher provides a list of themes moving from general to more specific issues within and across modules. The more specific topics are intended to emotionally involve the participants perhaps in a narrative that will result in the use of the vernacular. Tagliamonte (2006) notes that some of the modules have proven successful internationally since people are usually drawn to topics about the supernatural, extraordinary events and the like. Community-specific and culture-specific details can be investigated when researching the target community at large before entering the community.

One of the advantages of using the conversational modules is that the researcher can ensure a steady flow regardless of the respondent's reaction to the topic. Another advantage is that drawing on a set of modules will also ensure reasonable consistency across respondents and this will help elicit comparable data.

AMs' conversations revolve around family-related topics and supernatural and superstitious beliefs, which would also induce the use of a specific semantic field. Not only that, but I have also noticed that older AMs would only discuss topics for which they have the "proper" terminology, which would again result in repetition of keywords.

Recordings were made using a Zoom audio recorder at 44/kHz sample rate that is suitable for instrumental analysis (Zoom, 2017). The Zoom range recorders have several advantages for this kind of research. These recorders are solid state, which means that they directly record to a flash memory. The Zoom range also has high-quality internal microphones that can be set to a variety of angles which helped me get quality recordings from participants regardless of their seating positions and furniture arrangements, as well as avoiding hooking microphones on respondents that resulted in more relaxed sessions. This also helped me avoid "noise floor" created by the recorder itself such as when booting or changing settings, using a remote control.

I devised a unique identification coding system for the interview recordings including the date, time and length of the interview, name of the respondent, the equipment used and indications of permissions used/ signed consent, oral consent, recorded consent (see appendix). I kept copies of the original files with the original participants' names, but I used pseudonyms, that I assigned to each person, from then on.

3.3 Transcription and coding

It is becoming the standard in sociolinguistic variationist studies to have time-aligned audio files of transcriptions; this means using to text-audio linking software like ELAN[9] which can aid this (Nagy & Meyerhoff, 2015)

ELAN is a professional tool that allows researchers to create complex annotations on video and audio resources (Wittenburg et al., 2006). These annotations are speech stretches that are observed in the associated media file, they could be words, clauses, sentences and they can be created on multiple layers, called tiers. Tiers can be hierarchically interconnected

---

[9] (http://tla.mpi.nl/tools/tla-tools/elan, Wittenburg et al. 2006)

as well. A user may add an unlimited number of annotations to audio and/or video streams where an annotation can either be time-aligned to the media or it can refer to other existing annotations.

Having time-aligned annotations means that a researcher has a direct connection to a media file that contains stretches of natural speech that he/she is analysing. This allows the researcher to take account of all contextual data as well as providing more credibility to the analysis since the original data can be easily retrieved and accessed.

The use of ELAN "allows for transcription, extracting, coding, preparation of data for statistical analysis, calculation of some basic frequency statistics, and creation of a concordance all within one program" (Nagy & Meyerhoff, 2015:4). I devised a transcription protocol, namely, a set of conventions to be adhered to when transcribing data auditorily to ensure consistency and comparability across variation studies. My transcription is orthographic for both the NS corpus and the AM corpus, and it followed the guidelines in the Linguistic Data Consortium (LDC), (2017); Tagliamonte, (2006) as well as the guidelines from the University of Pennsylvania SLX Corpus of Classic Sociolinguistic Interviews.

The first task that I conducted in ELAN was transcribing the NS and the AM speech auditorily. I have transcribed the drink-and-drive conversations and the sentence reading components of the Wellington sub-corpus of the NZSED. But I ended up only using the conversation part because I did not manage to compile a sentence reading component for the AMs. I have also transcribed 21-hour long AMs conversations.

The next step was to identify the tokens of the variables of interest in the transcript. The analysis of phonetic and phonological variation is mostly binary in nature. The stable variables under investigation have variants that are well-established across the English-speaking communities, and these reliably reduce to choices between presence/absence of the

final /t, d/ segment in word-final clusters, and the choice between a velar and alveolar nasal in unstressed, word-final (ING).

Being a native speaker of Arabic, which has a wide range of stops gave me an advantage in identifying instances of CSD as I was able to differentiate glottal, deleted, devoiced, retained and reinforced final stops. On the other hand, my sense of what constitutes meaningful variants for (ING) was bolstered, probably because Arabic does not licence final /ŋ/ in coda positions. I, therefore, had difficulty hearing whether speakers were producing an alveolar or a velar nasal in that position. I, therefore, had to double check my coding of (ING) tokens for my NS and AM samples. Hence, I turned to assistance from instrumental analysis[10]. But In the absence of a reliable instrumental check on my own coding of the final nasals, I turned to a trained NS for assistance to double check my coding of the (ING) segments.

3.3.1 Coding coronal stop deletion (CSD)

The literature on the variable coronal stop deletion (CSD) is exhaustive and the recurring constraints that condition CSD in NS English are phonological and grammatical in nature. More specifically these are: preceding segment, following segment and the morphological status of a word. Labov (1989) identifies three additional constraints on CSD, namely (i) whether the syllable to which the [t, d] belongs is stressed or unstressed, (ii)

---

[10] Instrumental acoustic analysis deploys instrumental techniques to study the physical characteristics of speech sounds and it is typically used to study the most meaningful vowel formants (F1/F2). Praat (Boersma & Weenink, 2013) is an acoustic analysis program that creates spectrograms, which are visual representations of sound energy. Each spectrogram is tri-dimensional in the sense that the X axis represents the time of an utterance, the Y axis represents the frequency of the sound wave and the level of shading represents amplitude (loudness).
Formant patterns are natural resonances that can roughly be associated with the size of two articulatory tracks, the oral cavity and the pharyngeal cavity, that are separated by a constriction in the vocal tract. The first formant (F1) represents the pharyngeal cavity and the second formant (F2) represents the oral cavity. Formants are very well developed for vowels, but they are not as clear cut for consonants even for nasals. In the case of nasals, the formant patterns may be coloured by the formant patterns of nearby vowels.

whether the cluster that the [t, d] belongs to consists of two or more consonants, and (iii) the voicing of the segments flanking the [t, d]. but these constraints are less robust. Many studies that do report on these factors have found them not to contribute significantly to the likelihood of [t, d]-deletion. Several studies do not even report on these factors.

The coding of linguistic and social factors conditioning CSD for the NS was based on previous literature available for the variable in native varieties of adult and child English including New Zealand English (Guy, 1980, Neu, 1980; Labov, 1989; Guy & Boyd, 1990 ; Guy, 1991; Holmes & Bell, 1994; Roberts, 1997; Schreier, 2003; Tagliamonte & Temple, 2005; Smith et al., 2009; Hazen, 2011). The length of coda, e.g. CC or CCC was not found to be significant in previous work on NZE (Holmes & Bell, 1994) and was not coded for in the NS corpus. Holmes and Bell (1994) also found syllable stress patterns to be not significant for NZE. Therefore, this was also not coded in the current study. The coding schema implemented for CSD is as follows:

1. The Preceding segment

I have adopted the common coding schema for the preceding segment constraint on CSD with some adaptations, distinguishing sibilant fricatives (/s/, /z/, /ʃ/, /ʒ/, /t ʃ/, /dʒ/), non-sibilant fricatives (/f/, /v/, /θ/, and /ð/), nasals (/m/, /n/, /ŋ/), stops (/p/, /t/, /k/, /b/, /d/, /g/) and liquids (/r/ and /l/) that sometimes pattern distinctively from one another within the preceding segment constraint, therefore, they were coded separately (see Smith et al. 2009). A point to mention is that preceding /r/ was not coded for, among NS of NZSED because it is non-rhotic, so words like *appeared* were excluded not being typical contexts for CSD in NZE. But I included preceding /r/ for the AMs' coding because Arabic allows /rt/ and /rd/ syllable structures.

2. The Following segment

I have adopted the common coding schema for the following segment constraint on CSD with some adaptations, distinguishing sibilant fricatives (/s/, /z/, /ʃ/, /ʒ/, /t ʃ/, /dʒ/), non-sibilant fricatives (/f/, /v/, /θ/, /ð/), nasals (/m/, /n/, /ŋ/), stops (/p/, /t/, /k/, /b/, /d/, / g /), glides (/j/, /w/), vowels and pauses. /l/ and /r/ were coded separately as two categories and not collectively as" liquids" because earlier studies suggested they behave differently as following segments of final consonant stops (Roberts, 1997; Labov, 1997; Tagliamonte & Temple, 2005; Edwards, 2011).

 I have added a separate category of glides to test whether AMs treat them as a separate set of phonemes than vowels, as NS of English do. This step is important as I will make inferences about AMs' acquisition of linguistic and sociolinguistic competence later. Neutralization contexts are excluded because linguists cannot reliably distinguish the realization of the consonant clusters before other coronal stops.

3. The Grammatical category

Reviewing the literature on the variation in CSD for different varieties of NS English as well as AM speaker varieties, I have noticed an inconsistency in the coding of the grammatical function or morphological status across these studies. Specifically, I found that morphological status and word function are usually coded for together.

Ideally, form-based analyses consider the form of words regardless of the function they have in speech. This kind of coding is usually reported in terms of the number of morphemes a word consists of. For example, in a sentence like "she is the most composed person I know", the word *composed* will be coded as bi-morphemic. Whereas function-based analysis codes for the function a word has in speech. In the same sentence above, the word *composed* would be coded as an adjective, but the same word will be coded as a past tense

verb in the sentence "he composed many melodies". Nevertheless, the ambiguous coding of grammatical category for CSD persists in the literature.

For example, the common coding schema distinguishes "semi-weak verbs", "regular past tense verbs" and monomorphemes. This coding system has a verbal bias since the only non-monomorphemic forms coded for are verbal. It also excludes bimorphemic nouns and adjectives (like *-ment* suffixes and superlative *-ist* forms for adjectives).

To the best of my knowledge, only one study of CSD for NS attempted to code strictly according to the morphological complexity of the words containing examples of the variable, e.g. coding for contracted negation morphemes and part of a second morpheme; like *-ment, -ist*) (Holmes and Bell, 1994).

As for NNS studies, many of these exclude all token types other than verbs, which are usually regarded in terms of the acquisition of a second language's morphology. Namely; past tense morphology in English (Wolfram, 1984; Adamson & Regan, 1991).

I decided to go one further step and conduct two separate analyses for CSD based on two separate coding schemas. The first one is purely form-based, meaning that words are coded for the number of morphemes they consist of, regardless of their function, as either monomorphemic or bimorphemic/multiphonemic. In the latter category, I have also included bimorphemic nouns and adjectives.

The other coding schema is function-based; tokens were categorized based on their function in speech. For example, if a morphologically past tense form of a verb was used as an adjective, it was coded as an adjective and not a past tense verb.

My decision to use two coding schemas for grammatical category constraints on CSD in two separate models was to test whether there is a genuine role for the grammatical category constraint in conditioning CSD. Especially that Temple (2003), Tagliamonte and Temple (2005) and Hazen (2011) argue that the grammatical category constraint on CSD in

English may be an artefact of unevenly distributed data across phonological factor groups or even a dialect-specific aspect of CSD. Another reason is to find out whether the underlying grammar is function-based, form-based or whether these distinctions are not perceived at all among NS of English. The two different coding schemas are contrasted in A and B:

A. Form-based coding of grammatical category.

Following Brown's (1973) system for coding categories that emerge in children's first language development, I coded the following as bimorphemic: regular past tense, negated forms, superlatives and bimorphemic nouns. All other lexemes are coded as monomorphemic.

B. Function-based coding.

I used an elaborate functional categorization for words compatible with parts of speech (noun, discourse marker like *and*), adjective, adverb, regular past tense, semi-weak past tense, present and suppletive preterite, i.e., went). Tokens of CSD were coded for their function in speech regardless of their forms. In situations of doubt about the function of tokens, I consulted dictionaries, grammar books and NS of NZE. The difference between coding tokens as adverbs and/or prepositions for some words like *around* and *past* depended on the context (Huddleston & Pullum, 2002). Examples of grammatical categories:

- Nouns*: bimorphemic nouns*: *judge**ment**, advertise**ment**.*
- Nouns: proper nouns*: Auckland, New Zealand.*
- Nouns: Agent*: final**ist**.*
- Adjective: ordinal adjectives: *first, next, second.*
- Superlative forms: *best, worst.*
- Bimorphemic adjectives: *bereaved, observed.*
- Affixed adjectives: *unpleasant* (also bimorphemic).

- Adverbs: *fast, most.*

- Prepositions analysable as adverbial functions: *past, beforehand, around, round, almost.*

- Regular past tense*: aimed, observed, looked, finished, shocked.*

- Semi-weak past tense*: told, kept, left, split, felt, lost.*

- Suppletive preterite*: went.*

- Present tense*: understand, suggest, end, accept.*

4. Style-shifting

Style-shifting as a constraint on CSD does not have a strong effect compared with other variables, like (ING). Nevertheless, it presents itself as a constraint on CSD in NS varieties in English in tandem with the following segment, particularly in pre-pausal settings. For example, Holmes and Bell (1994) and Bell (1977) reported zero deletions in pre-pausal settings in formal styles. Coding style for NS is also important when I compare AM patterns of variation for CSD.

5. Gender: male, female.

6. Age. Age is coded as a discrete variable of young, middle-aged and old.

I also added three factor-groups pertaining to different theories that explain the tendency of the persistence of one variant across speech since some evidence about the potential role of priming have been found in studies of NS English (Abrahamsson, 2001, 2012; Tamminga, 2014). The three factor-groups pertaining to persistence are:

7. The realization of the variable in a prior token: this corresponds to the simple definition of priming effects (cf. Abrahamsson, 2001).

8. The number of morphemes a prior token consists of: this corresponds to (Tamminga, 2014) findings regarding the existence of a simple priming effect across subsequent

tokens within a specific linguistic level; 1. Number of morphemes of the prior token (monomorphemes, bimorphemes). 2. Number of morphemes of current token (monomorphemes, bimorphemes).

9. I added a constraint related to persistence which is the distance of a current variable realization from a previous one. I assumed that the closer the variables the more likely the repetition of a given variant. The distance was measured in terms of the subsequent variants being within same breath unit or not.

Coding for the AMs included the same constraints found among NS of NZSED. I also coded several NNS-specific constraints based on results from CSD studies in contact-induced varieties, learners' interlanguage, ethnolects and English-based creoles (Wolfram, 1984, 1985; Santa Ana , 1991; Bayley, 1994, 1996; Patrick, 1999; Edwards, 2001; Drummond, 2010; Hoffman & Walker, 2010; Edwards, 2011; Schleef, 2013b).

I coded for the social factors that proved to be the most significant in the literature of acquisition of variation among NNS. These include gender, age and age at arrival in the host country that were coded as continuous constraints. The length of stay in the host country was operationalized to test if there is a time threshold after which NNS are ready to acquire variation in an L2; more than three years and less than three years (Schleef et al., 2011; Sharma & Sankaran, 2011; Boberg, 2014) . I have also operationalized network involvement for my sample, as mainly Arabs and mainly-non-Arab; having two options for network affiliations as observed in the literature (Meyerhoff & Niedzielski, 2003; Drummond, 2010; Schleef et al., 2011; Sharma & Sankaran, 2011; Drummond 2012).

Attitude, among NNS, is often measured in terms of the level of belonging to the host country and can be operationalized as the intention to stay in the host country (yes, not sure)

and self-reported affiliation, in my case whether AMs to Wellington identify themselves as New Zealanders, Arab-New Zealanders or Arabs.

Finally, I coded for self-reported English Proficiency (good/excellent/native-speaker) and the way English was acquired/learned (learned formally at school before migration, acquired integratively after migration, acquired as a child in NZ).

3.3.2 Coding (ING)

The literature on the variable (ING) is as exhaustive as that on CSD and the recurring constraints that condition (ING) variation in English are phonological and grammatical in nature. More specifically these are: preceding segment, following segment and the morphological status of a word and/or its grammatical function.

The coding of linguistic and social factors conditioning (ING) for the NS of NZSED was based on previous literature available for the variable in native varieties of English including New Zealand English, for both adults and children (Labov, 1966; Cohen & Labov, 1967; Wolfram, 1969; Fasold, 1972; Trudgill, 1974; Houston, 1985; Labov, 1989; Bell & Holmes, 1992; Roberts, 1997; Gordon, 1998; Romaine, 2003; Tagliamonte, 2004; Labov et al., 2006; Abramowicz, 2007; Campbell-Kibler, 2007; Hazen, 2008; Campbell-Kibler, 2012; Tamminga, 2014).

1. The Preceding segment

I have adopted the common coding schema for the preceding segment constraint with some adaptations, distinguishing preceding apical, preceding velar, preceding /ing/ cluster as in *singing*, preceding /l/, preceding /r/, and preceding other consonants.

2. The Following segment

I have adopted the common coding schema for the following segment constraint with some adaptations, distinguishing following apical, following velar, following pause, following vowel and following other consonants.

3. Grammatical category

I have coded for grammatical categories based on Labov's (2001) work on the continuum ranging from most nominal to most verbal categories where nominal categories prefer the velar nasal, and verbal categories prefer the alveolar nasal, as follows:

/n/ ⟵――――――――――――――――――――――――――――⟶ /ŋ/

Progressive   participle   adjective   gerund   noun   pronoun
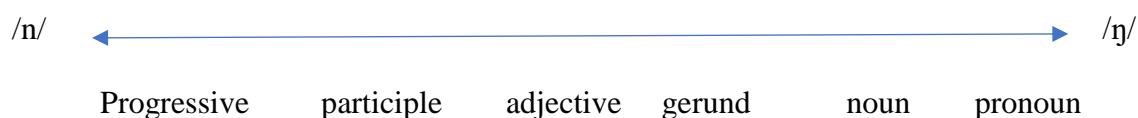
The basic criteria I employed to differentiate the following groups: (i) nouns and gerunds, (ii) gerunds and participles and (iii) participles from adjectives was based on specific characteristics and typical structures of each category. For example, a gerund is a form of a verb used as a noun, and it can be used in the same way as a noun; it could be the subject of a sentence as in (*Driving* too fast is dangerous). A gerund can also follow a preposition like (Can you sneeze without *opening* your mouth?). Gerunds also follow cognitive verbs as in (I like *cooking,* And They hate *milking* cows). Gerunds also exist in compound nouns like (We are going to the *swimming pool*).

Whereas a participle is a form of verb used as an adjective or as a verb in conjunction with an auxiliary verb like (he is *walking*), where it serves as a progressive verb. As for nouns, I have considered words with [ing] as a part of the root like *ceiling* or words that are historically affixed but have become so common that traces of such affixation are not realised any longer like *parking, cooking.* Examples of the grammatical categories from the NZSED are:

- Progressives: were *coming,* is *driving,* was *wearing,* are *saying.*
- Participles: he sat down *watching* T.V, he was jogging *listening* to music.
- Gerunds: *driving* is dangerous, I started *seeing*, *speeding* has many consequences.
- Adjective/participle: *numbing* experience, *exciting* news, *interesting, amazing*.
- Noun/compound nouns: *drink-driving, morning, ceiling, feeling,* a *crossing.*
- Pronoun: *anything, something, everything.*
- Discourse marker: *or anything, and everything.*

4. Gender: male and female.

5. Style. Style was coded for in terms of speech produced along different types of monitoring of language by the speakers: careful as opposed to casual speech.

I added constraints pertaining to several aspects of "persistence" since "[t]he idea that the realization of one ING can then affect the realization of a subsequent ING, favouring similarity" Drummond (2012:107). These were operationalized slightly different from the previous operationalization presented for CSD, because (ING) has morphological, phonological and syntactic attributes.

6. Distance from the previous token (same breath unit, different breath units).

7. Realization of the previous token (velar, alveolar).

8. Prior grammatical category (nominal, verbal).

The coding of linguistic factors conditioning (ING) for AMs was based on results from (ING) studies in contact-induced varieties and learners' interlanguage (Adamson & Regan, 1991; Drummond, 2010; Schleef et al., 2011; Drummond, 2012; Meyerhoff & Schleef, 2012).

I have included the NNS-specific constraints that proved to be the most significant in the literature on the acquisition of variation by NNS and which have been previously outlined for CSD.

An important point to mention here is that I have added separate categories of sounds for the AM group under the preceding segment constraint for CSD. A preceding /r/ context is included for the AMs because Arabic allows /rt/ and /rd/ syllable structures. For the following segment constraint on CSD, I have also added a category of glides for CSD coding for AMs to test whether NS treat them as a separate set of phonemes than vowels, as NS of English do.

This step is important because it allows me to understand the level of acquisition of the L2 phonemic inventory which would enable me (together with qualitative data as well as

information about the AM's special social factors) to make inferences about the AMs

belonging to two different second language communities of practice. It will also inform my

understanding of how far this group approaches the constraint hierarchies of the L2.

3.4 Number of tokens per speaker, Type-token ratio and excluded tokens

Sankoff (1978) suggests that gross constraints on variables can be located with a few tokens whereas fine differences need 30 tokens, at least, to be located. Guy (1980) identifies 30 tokens per person as the reasonable target for tokens with no less than 10 tokens per person as the acceptable norm for sociolinguistic analysis. But Meyerhoff et al. (2015) note that what matters is the number of tokens per cell rather than per person (also see Milroy and Gordon (2003) judgement sampling techniques). Accordingly, I have attempted to get 50 tokens per person for each variable and 30 tokens per cell for each identified social constraint for this study.

The type-token ratio (TTR) is a means of measuring lexical diversity in each corpus; it is usually calculated by the number of types of words divided by the number of tokens. Some linguists exclude the fourth or fifth repetition of a word in the analysis because they believe repetition might skew the data (Tagliamonte, 2006; Tagliamonte & Baayen, 2012). I, nevertheless, have adopted a more flexible approach to TTR for both the NZSED and the AMs' corpora for several reasons. Firstly, the NZSED utilizes theme-prompts: drink and driving ads and map tasks. These kinds of thematic prompts seem to have induced some cognitive priming and consequently the repetition of keywords.

Secondly, using conversational modules (Labov, 1984), I noticed that AM's conversations revolve around family-related topics and supernatural and superstitious beliefs, which would also induce the use of a specific semantic field. Not only that, but I have also noticed that older AMs would only discuss topics for which they have the "proper" terminology, which would again result in repetition of keywords.

Table 3.1 and table 3.2 illustrate token number for AMs and NS of NZSED for the variables CSD and (ING).

Table 3.1: Token grid for CSD and (ING) among AMs				
N of AMs	age	gender	CSD n of tokens	(ING) n of tokens
9	young	5 f, 4 m	455	2616
7	Middle-aged and old	5 f, 2 m	819	1604
5		5 f		
21		15 females, 6 males	1274	4220

Table 3.2: Token grid for CSD and (ING) among NS of NZSED				
Total N of NZSED speakers	age	gender	token N (ING)	token N (CSD)
15	young	5 f, 10 m	413	171
11	Middle-aged	5 f, 6 m	110	475(a split showed in the results; old and middle-aged versus young)
15	old	9 f, 6 m	238	
41		19 females, 22 males	61	646

3.5 Preparing the data for analysis - data cleaning

Once the data for was coded in ELAN, the resulting files were saved in CSV format, which is compatible with the statistical tools I used for this study, namely R and Rbrul statistical tools (cf. Tagliamonte & Baayen, 2012).

The data was carefully cleaned in an Excel worksheet, eliminating all typos, irregular coding, and empty cells. The statistical analysis software utilised is case and space sensitive, which means that if one word was once in upper case and then in lower case, it will be considered as two separate lexemes. The same thing goes for spacing if a word has an indentation and another doesn't, this word will be counted as two separate words,

confounding the analysis. Consequently, I cleaned up the data, mending spelling errors, capitalizations and spacing inconsistencies.

3.6 Statistical tools: R and Rbrul

R is a general-purpose and free software environment for statistical computing, statistical analysis and data visualization (R Development Core Team, 2008). It compiles and runs on a wide variety of operating system platforms. Rbrul is an application running on the statistics software R. This application allows the user to carry out descriptive and inferential statistical analysis usually via the implementation of generalized linear modelling and regressive statistical analyses (Johnson, 2009). Rbrul is a descendant of an old tradition for analysing variable rules, and it can be used to assess the influence of multiple factors on linguistic variables quantitatively.

"A variable rule program evaluates the effects of multiple factors on a binary linguistic 'choice', e.g. the presence or absence of an element, or any phenomenon that can be treated as an alternation between two variants. The factors can be internal (linguistic), such as phonological or syntactic environment, or external (social), for example, speaker gender or social class. The program identifies which factors significantly affect the response variable of interest, in what direction, and to what degree." (Johnson, 2009:359).

The widely available alternative, Goldvarb (Sankoff et al., 2005), has some shortcomings that may isolate its users from the wider community of quantitative linguists (Johnson, 2009). Rbrul attempts to solve the concerns associated with other versions of the variable rule program (for a comprehensive overview, see Johnson (2009)).

The main difference between Rbrul and Goldvarb is Rbrul's use of mixed effects modelling, whereby a distinction is made between fixed effect factor groups like linguistic factor groups and random effect factor groups like speakers and words. The inclusion of

random effects is intended to help quantitative sociolinguist eliminate Type I statistical errors (which may overestimate the role of social variables).

By considering the individual speaker as a random effect, Rbrul considers that the same individual speaker may prefer one variant over the other and this tendency does not necessarily reflect the tendency of the community at large. Such a model "can still capture external effects, but only when they are strong enough to rise above the inter-speaker variation" (Johnson, 2009:365). For the purposes of the present study, Rbrul offered the most meaningful approach to the data.

The Rbrul program provides "the three lines of evidence"(Tagliamonte, 2006) for assessing the structure of variation. It displays the statistical significance of constraints, the rank ordering of constraints and constraint hierarchies.

A constraint is statistically significant if its p-value is below the 0.05 cutoff value which is the norm in sociolinguistics and variation studies. This step is important to the extent that the decision about the underlying grammar of a linguistic variable is largely dependent on which set of constraints are statistically significant. When, for example, investigating a linguistic variable that occurs in two different communities and the analysis shows that the statistically significant constraints are not the same, it becomes obvious that, though the variation can be found in both communities, the underlying system differs markedly (cf. Tagliamonte 2006: 237). This point is important when I compare NS and AM patterns of variation and the associated underlying grammars.

Understanding the constraint ranking provides valuable insight into the relationship between a variable choice and linguistic context. Each constraint is made up of several levels, each of which has a certain effect on a variable reported in factor weight units. If a factor weight, which is a weighted (centred) estimate of the contribution of each factor when all factors are considered together, is over 0.50, it generally favours the application of the rule,

here variant choice, while it disfavours it when it is below the 0.50 level. However, it is sometimes more useful to understand the relative factor weights and their relative position within the constraint ranking.

Constraint rankings are usually reported in a unit called the range. The range is calculated by subtracting the lowest factor weight of a level from the highest factor weight within a constraint. The calculated range is then compared to the ranges of the remaining significant factor groups. The highest figure or widest range equals the strongest constraint. This information enables us to position factor groups according to their relative strengths. The different levels of a constraint with their associated factor weights collectively comprise a constraint hierarchy from the level with the highest factor weight to the level with the lowest factor weight. Constraint hierarchies may differ across different varieties of language and are also used to compare underlying grammars.

Such representation of data is the norm in sociolinguistic variationist studies. This data presentation also enables building solid interpretations and arguments (Tagliamonte, 2002, 2006, 2011) as well as making analysis and results comparable with other studies in the field.

3.6.1 Regression analysis and mixed-models

The type of regression analysis usually deployed in variationist sociolinguistics is binomial logistic regression which best fits linguistic data, typically introduced as a binary choice for a variable. This process estimates the relationships between a dependent variable i.e., the response (here, the response is the application value chosen for a variable: deleted final consonant for CSD and use of the alveolar nasal for (ING)) and one or more independent variables, i.e., predictors (internal and external constraints on variation). Then, it introduces the output of this relationship as a probability of achieving the application value of a variable.

This is done by means of stepwise logistic regression (step-up/step-down), using a maximum likelihood algorithm. In statistical terms, maximum likelihood estimation (MLE) is a method of estimating the parameters (constraints/independent variables) of a statistical model's observations (observed token distribution), by finding the parameter's values that maximize the likelihood of achieving the observations given the parameters (Galili, 2017).

This approach has been very fruitful in variation studies; however, Tagliamonte and Baayen (2012) suggest another way to approach the relationship between the sociolinguistic response and the linguistic and social predictors by applying non-parametric methods. They suggest that conditional inference trees and random forests, together with mixed-effects models, are effective statistical techniques to add to the sociolinguist's "toolkit".

One form of non-parametric tests is that known as recursive partitioning, this process evaluates how close a combination of independent variables, i.e., predictors (internal and external constraints on variation) reflects the distribution of the dependant variable in a corpus.

The recursive partitioning starts by randomly selecting predictors from a suggested pool, this is done in many iterations that exhausts all possible combinations of predictors. Finally, the combination of predictors most voted on by hundreds of recursive partitioning iterations is introduced as the best fit to the data distribution.

To understand the true nature of my data for the variables, I decided to consider two non-parametric tests which are particularly suited to relatively small data sets with possibly collinear factors; conditional inference trees (CIT) and random forests. Tagliamonte and Baayen (2012:22) explain the merits of random forests for sociolinguists:

> "Random forests provide information about the importance of predictors, whether factorial [binary] or continuous, and do so also for unbalanced designs with high multicollinearity, cases for which the family of linear models is less appropriate".

The random forests method starts analysis by partitioning the data and creating many decision trees. This is done by generating random samples of the original data through a process known as bootstrapping, i.e., replacement, and using several predictors selected at random from all the predictors to determine node-splitting; this eventually creates a hierarchy of predictors (Galili, 2017). Multiple subsets of trees are built, and the support for the role of each predictor in each decision is noted.

A useful function to apply based on random forests is the variable importance plot which illustrates how each predictor contributes to distributing the data. The plot shows each variable on the Y-axis, and their importance on the X-axis in descending order as suggested by the position of the associated dot on the X-axis and a predictor with a value greater than (0) would be significant in terms of fit to data distribution.

The variable importance graph shows the consistency with which factors were voted on by forests of CITs generated from randomly selected subsets of the data. Although this measure does not assess statistical significance in terms of goodness of fit or direction of effect (Galili, 2017), they do use significance in terms of distribution.

This is helpful since we do not know much about the constraints that govern variation on (ING) and CSD by AMs in contact settings. In other words, random forests:

> "offers the analyst at the very least a preliminary view on the nature of the dataset and the impact of the predictors. … A conditional inference tree provides estimates of the likelihood of the value of the response variable [in this case] (was/were) based on a series of binary questions about the values of predictor variables." (Tagliamonte and Baayen 2012: 22).

Conditional inference trees add an additional layer of information, reveal interactions among predictors and allow the visualization of interactions which better informs our understanding of variation among AMs, especially if there are factor interactions only for AMs. We can differentiate cases of interactions as opposed to confusing results that may be due to natural data asymmetries or sampling errors, both of which are the norm in sociolinguistic data sets.

Although I initially intended to use non-parametric partitioning as an exploratory stage for data analysis, I realised that I could use them together with logistic regression to triangulate my data and gain a better understanding of the implications of my results regarding patterns of variation found among AMs. Especially since I have relatively small data sets; low token numbers for the NS in the NZSED and low speaker numbers for the AMs.

To summarise, I have used conditional inference trees as an alternative technique to uncover the patterns of variation found among NS and AM groups and I have used random

forests and variable importance plots to confirm the information obtained by the conditional inference trees. Both non-parametric tests were used to validate qualitative observations or double check my understanding and interpretation of the results obtained in logistic regression.

Therefore, the outline presented in the results chapter introducing the results for non-parametric tests followed by the results for the parametric regression (for each of the variables among the three groups of speakers) does not reflect the order in which these tests were conducted. I did not follow a linear order where I applied non-parametric tests and then followed them with logistic regression. The process was more of going back and forth between these techniques to best understand my data, but I felt it was clearer for the reader if I chose a consistent layout and stayed with that.

I decided to conduct mixed-effects modelling in Rbrul since Johnson (2009) suggested that treating speakers as a random effect in logistic regression, i.e., mixed effects logistic regression would account for intra-speaker variability consequently avoiding Type I errors, which may report social variables that are not genuinely significant. However, coding for the speaker separately, which is the norm in contemporary variationist sociolinguistic research, necessarily entails interaction with social factors (Tagliamonte, 2006, 2011, 2012). Lexical items may also entail interaction with word-related variables, e.g. morphological class and part of speech. I, therefore, decided to implement analyses with words as a random effect as well.

I also used mixed-effects modelling in Rbrul, although R is equally capable of doing so, because of the representation of the results involving the "three lines of evidence", that are common in studies of linguistic variation (factor group hierarchies, factor group weights and p-values) and hence enable more direct comparison with previous studies (for an

overview of the benefits of mixed modelling see Johnson, (2009); Tagliamonte & Baayen, (2012)).

3.6.2 Identifying the best models

My goal was to identify the best model for the two variables, coronal stop deletion (CSD) and (ING), in the NZ corpus of New Zealand Spoken English Database (NZSED) and the corpus of AMs of New Zealand English. The best model should:

(1) avoid common problems of data handling like low number of tokens per cell, knockouts, collinearity and interactions, (2) support specific theoretical claims made about CSD and (ING), (3) explain a reasonable amount of the variation of the variables as suggested by the R^2 value, and (4) meet the statistical requirements for the best fit for data.

The statistical sense of "best fit" has already been introduced, but here I will discuss, in detail, the criteria used for choosing the best model for the data sets. These form the basis of the mixed effects models that I will present in the next chapter.

Log-likelihood criterion

The log likelihood focuses on the distribution of the data in each sample. Simply put, this test determines the comparison between the likelihood of the distribution of the data being achieved by a complex model, against the likelihood of the distribution of the data under a less complex model (the null model). The closer to zero the log likelihood value is, in general, the better.

This statistical concept is only useful when there are competing models, and they are fitted to the same data frame and are nested; meaning that one model is more complex than the other (i.e. has more degrees of freedom). If after conducting Chi-square test for two competing models the resulting p-value was less than 0.05, then there is evidence against the null model (R Development Core team 2017).

Akaike information criterion (AIC)

The value of AIC is based on the concept of statistical deviance. Null deviance pertains to how well the response (dependent variable) is predicted by a model with only the intercept (the null model, no added factor groups). Residual deviance pertains to how well the response is predicted by a proposed model. The closer to zero this value, the better the model fit (R Development Core team, 2017) .

This statistical concept is only viable when there are competing models, but the competing models **do not need** to be fitted to the same data frame or to be nested. This last point is important especially since I compare NS models with AMs to judge how far the AMs replicate the patterns found in NS speech.

Another related concept is that of Dxy, where the value Dxy ranks the correlation between predicted probabilities of a model and observed responses. This function can also be applied in R providing an even stronger interpretation of fitness of model by providing the concordance probability of the model (C). The closer the values of Dxy and C are to 1, the better (R Development Core team, 2017).

Fisher scores

This value refers to the number of iterations needed to fit a model, as well as achievement of maximum likelihood at each level of model-building (achieving convergence) (R Development Core team, 2017). In Rbrul, this value is implied, whereas, in R, it is listed as a part of model summary. Sankoff (1988) describes a smooth regression as one where step-up and step-down match, where factor weights fluctuate minimally across iterations, and convergence is reached for each run in under 20 iterations (the maximum limit of the Varbrul program at the time).

3.6.3 Correlation and interaction among constraints

Collinearity, or excessive correlation among explanatory variables, can complicate and prevent the identification of an optimal set of explanatory variables for a statistical model. Collinearity occurs most likely when one constraint on variation is coded for under several constraints. To avoid this happening, I carefully considered the underlying assumptions for the coding scheme I am using and then eliminated any instances of double coding. For example, I coded for the grammatical category for CSD once focusing on form (i.e. coding morphological status).I also coded grammatical category as parts of speech (i.e. a more function-based analysis). But I never combined both constraints in one model in one regression analysis. Instead, by running them in separate models, I could determine which one generated a model that was a better fit to the data, e.g. by comparing log likelihood measures.

The second step was to run detailed cross-tabulations for each potential constraint with the response (the variant choice, i.e. the application value for the variable being analysed), one at a time. This allowed me to identify cases of interactions and to identify places where there were low token numbers.

Interaction happens when the effect of one constraint is conditioned by one or several other constraints. Guy (1980) attributes interactions to one of three reasons: (i) natural data asymmetries, which is the norm in sociolinguistic data; (ii) sampling issues; (iii) real interactions. The worst-case scenario was when I found what is known in the variationist literature as knockouts (categorical applications or non-applications of a variant). This occurred in my datasets because of natural data asymmetries for both the NZSED and the AM corpus, but also because of sampling challenges.

Some of the interactions identified in this way were resolved by collapsing some levels within constraint hierarchies when possible, that is, when collapsing made sense

theoretically, and when levels within constraint hierarchies behaved alike, or when factors behaved alike and were few. For example, a typical case of interaction because of natural data asymmetries surfaced while coding for CSD, I kept preceding /l/ and preceding /r/ for AMs as two levels of the preceding segment constraint because Arabic speakers may be inclined to exhibit first language transfer, pronouncing a post-vocalic /r/ where NS of NZE are non-rhotic.

A typical example of interaction because of sampling issues emerged, for example, with my inability to persuade older AMs males to participate in the study. Having identified instances where the data was skewed heavily like this, I became more cautious about what kinds of generalisations I can make about age and gender in the AM sample.

After finishing this step, I cross-tabulated the social constraints and linguistic constraints; since social constraints are rarely orthogonal, it was especially important to carry out cross tabulations across all social constraints.

Coding for speakers will inevitably cause interactions to occur across social constraints (Tagliamonte, 2006, 2012). I, therefore, cross-tabulated the social constraints for individual speakers. This step revealed interaction between age and non-native speaker-specific constraints; these are networks, self-reported proficiency, and intention to stay in New Zealand and how English was acquired/learned.

For example, younger AMs had mainly non-Arab networks, referred to themselves as native speakers of NZE, all intend to stay in New Zealand or move to Australia, and all of them reported that they picked up English naturally. Older AMs, on the other hand, have mainly Arab networks, listed their proficiency as either very good or excellent (but not native-like). All of them intend to stay in New Zealand, and all of them learned English as adults via integration and English classes provided for migrants to New Zealand (these are obligatory, or they lose their social subsidies).

Middle-aged AMs had mixed networks and reported their proficiency as excellent or very good, but never native-like. They all plan to stay in New Zealand, and only a few received a formal introduction to the English language.

The second step was to consider the variance inflation factor (VIF) which is a value that represents the level of interaction among predictors if any. This value is generated automatically in Rbrul only when interaction occurs amongst explanatory factors. Interpreting the value is straightforward: if the VIF ≤ 1 no interaction exists, if 1< VIF<5 there is a moderate interaction, if VIF > 5-10 then there is a high level of interaction.

I initially encountered a problem of a VIF greater than 2.5, while conducting regression analysis for the variable CSD because the grammatical category constraint and preceding segment constraint were interacting. As a result, I recoded the grammatical category constraint, clustering detailed categories into broader ones, when these subcategories had a similar pattern of behaviour and where combining them was principled on linguistic grounds, e.g. recoding superlative forms with adjectives.

But when the interaction persisted, I went back and recoded the preceding segment constraint according to the broader manner of articulation groups (obstruent consonants versus sonorants, with vowels coded as a separate category of sonorant vowels. The VIF value disappeared. I still noticed many empty cells caused by the words *and* and *just*, which I coded separately.

Consequently, I tried to differentiate preceding nasals in the word *and* as opposed to other tokens and the preceding sibilant /s/ in the word *just* as opposed to other tokens.

So, I created two new levels within the preceding segment hierarchy to separate preceding /n/ from /n/ in the word *and*, and /s/ for sibilant fricatives from /s/ in *just*. The problem persisted, so I ended up excluding *and* and *just*. The token number for the study

decreased greatly, but I still had over a thousand tokens of CSD for NS of NZSED and over three thousand tokens for AMs, to test my hypothesis.

After making all the needed amendments, any asymmetries left in the data sets are the result of natural data asymmetries rather than data-related problems.

3.6.4 Initial modelling

After cleaning the data sets in these ways, I started initial modelling in R by conducting non-parametric statistical analyses to find out the possible significant constraints that condition variation on CSD and (ING) in the speech of NS of NZSED and AM speakers of NZE.

As mentioned above, non-parametric tests (here, random forests and conditional inference trees) provide insight into the predictors (constraints) that influence a response (in this case variant choice) without showing the direction of this effect or its statistical significance and they are, therefore, important for initial modelling. These methods also help to check if the constraints that were found to be significant in previous studies have genuine effects on variation, especially social constraints. These non-parametric tests also allow the visualization of predictors working in tandem and thus provide a better understanding of variation.

The results from the non-parametric tests formed the basis for the models for NS of NZSED and AMs for the two variables of the study. The next step was for me to test the variables' significance in R, since non-parametric modelling does not provide information about statistical significance and/or its direction. I chose to build the model manually in R rather than going straight to Rbrul as this allows me to better explore the details of my datasets.

I started with the NS group and tried to find the statistical significance level for each of the variables that were indicated to be important for the variable CSD and (ING). I started by trying every single constraint with the response for each of the variables at a time and excluded all constraints that had a p-value greater than 0.05.

I then utilised tried and true measures of goodness of fit convenient for different types of models and different types of relationships among models. These include nested models

where one model is embedded in a larger one, nested here in that all predictors of a smaller model occur in a larger model, and I used Chi-square tests to decide which model was the best fit to the data. Non-nested models are ones where it is not possible to derive one from the other either by means of parametric restriction or limiting process (R-bloggers 2017).

For example, when I compared form-based models with function-based models of CSD among NZSED speakers, these two models had different assumptions and different predictors. I considered other measures of goodness of fit for non-nested models like log likelihood, AIC and Fisher scores. I also considered the amount of variation explained by a model by considering the value of the R squared (R2).

The last step was to try and avoid Type I (and Type II) errors. In statistical hypothesis testing, a Type I error is the incorrect rejection of a true null hypothesis, for instance, by exaggerating the effect of a constraint. For example, reporting falsely that a social factor like gender or age effects variation (when this is perhaps an artefact of one or two speakers of that gender or age group making very high or very low use of one variant). A Type II error is basically the opposite: a true effect is rejected, incorrectly retaining a false null hypothesis, e.g. denying a social effect that truly exists. I have tried to avoid such errors by coding speakers and word as random effects by using mixed-effects modelling.

3.7 Chapter summary

In conclusion, I have chosen statistical methods that allowed me to identify, quantify and tease apart (i) best models among several competing models, (ii) significant factor groups, (iii) interactions or collinearity. In this chapter, I have outlined the process of preparing the data for analysis and then the steps undertaken to analyse the two datasets using tools provided by freely available software packages. I have highlighted some of the hazards associated with analysing naturally occurring data sets and outlined the measures I have taken to avoid the major pitfalls, such as over-reporting the significance of effects, and under-estimating the independence of different predictors.

This lays the ground for the next chapter, where I present the results of my analysis of the two variables under investigation, and ultimately will compare the constraints, and constraint hierarchies operating on the variables in two groups of speakers. This will enable me to be confident when making claims that AM speakers of New Zealand English have or have not acquired patterns of variation associated with two stable variables in NS speech.

Chapter 4: Results

The purpose of this chapter is to present the results of the analysis of the two variables under investigation; CSD and (ING). I will illustrate how NS of NZSED behave (same/different) compared with other NS of other varieties of English. I will also show that the AM corpus provides evidence for a reliance on natural, phonetic tendencies and for partial replication of the NS norms especially for the first-generation of AMs.

Young AMs (1.5 generation) patterns of variation may be described as a combination of NS norms and first-generation AMs norms as there seems to be evidence of both the acquisition of local patterns of NS variation and the maintenance of older AMs patterns. Before moving on to discuss the results for my research project, I believe it is very helpful to explain the process involved in the exploratory analysis that I have conducted.

The structure of the chapter is as follows: First, I introduce the results for variation patterns found among NS of NZSED for the variable CSD. I then introduce the results for variation patterns found among AMs of New Zealand English for the variable CSD. I first try to provide results for AMs of all age groups, I then demonstrate that it is better to discuss AM results under two sub-groups: old and middle-aged AMs versus young AMs because the feedback from regression analysis, recursive partitioning and qualitative data regarding the social factors of the AMs revealed that although all AMs respondents belong to an Arab migrant community of practice, they also belong to two second language communities of practice. I then provide a summary of the results for CSD among NS and AM groups.

In the second section of this chapter, I introduce the results for variation patterns found among NS of NZSED for the variable (ING). I then introduce the results for variation patterns found among AMs of New Zealand-English for the variable (ING), and I provide further evidence that AMs results are best explained under two sub-groups: old and middle-

aged AMs versus young AMs . I then provide a summary of the results for (ING) among NS and AM groups. Finally, I provide a chapter summary.

As noted in the last chapter, a crucial point to emphasise here is that introducing the results for non-parametric tests followed by the results for parametric regression (for each of the variables among the three groups of speakers) does not reflect the order in which these tests were conducted. I did not follow a linear order where I applied non-parametric tests and then followed them with logistic regression. The process was cyclical and mutually constitutive throughout.

4.1 Results for CSD among NS of NZSED

I started the data analysis process for the variable CSD for NS in the NZSED corpus by trying to conduct an initial modelling phase that will help me downsize the number of linguistic and non-linguistic constraints on variation for CSD and (ING) among NS of English as suggested in the literature. One of the advantages of applying non-parametric statistical test like random forest techniques is its ability to include many constraints in a regression model without the model crashing or producing faulty results. Applying this method has also enabled me to test hypotheses about the role of other constraints that may have not been studied extensively before.

4.1.1 Results for non-parametric regression for CSD among NS of the NZSED

The non-parametric regression analysis formula was fed all possible constraints that may be responsible for conditioning variation in CSD among NS of NZSED. These constraints were chosen based on the literature provided for the variable in different varieties of NS English, as well as constraints found in ethnolects and other contact induced-varieties as well as constraints found in non-native speaker varieties like learners' interlanguage and language-abroad settings (see chapter 3, methodology).

Figure 4.1 was created in R statistical analysis software to represent the output of the random forest analysis. As discussed in chapter 3, random forests output is usually visualized using the plot function in R, representing all significant constraints[11] in a decreasing order of potential effects on a variable. The values provided in Figure 4.1 that was run for the variable CSD among NS of NZSED are reported in terms of the mean decrease in accuracy (mean decrease Gini); the mean decrease in Gini coefficient is a measure of how each constraint

---

[11] Random forests and conditional inference trees introduces significant factor groups in terms of distribution and not model fit.

contributes to the homogeneity of the nodes and branches in the resulting random forest (R Development Core Team, 2013).

Each time a constraint is used to split a node, the Gini coefficient for the child nodes are calculated and compared to that of the original node. The Gini coefficient is a measure of homogeneity from 0 (homogeneous) to 1 (heterogeneous), and the changes in Gini are summed for each constraint and normalized by converting these values into zero scores at the end of the calculation.

In other words, the values introduced as the Mean Decrease Gini are normalized into Z-scores for easier comprehension and comparison; constraints with a large mean decrease in accuracy are usually more important in explaining the data (R Development Core Team, 2013).

The output of this non-parametric regression analysis is easy to interpret. For example, the split in the potential effect of the constraints is very visible as the following segment, the grammatical category, the preceding segment and age have the strongest effect sizes that may or may not reach statistical significance in predicting the CSD variation by NS in NZSED. Other constraints cluster with weak effects.

```
Figure 4.1: Variable importance for constraints conditioning CSD among NS
of NZSED
Following Segment                                                    ●
Grammatical Category                                            ●
Preceding segment                                          ●
Age                                                  ●
Realization-Of-prior-token                     ●
Style                                   ●
Distance-between-tokens                 ●
Gender                              ●
Number-of-morphemes-of-
current-token                       ●
Task-type                           ●
                          0  10  20    30    40    50    60
                        Mean decrease Gini
```
Italicised constraints are different measures of persistence
Bold constraints are more important than others

Figure 4.1 illustrates that there seems to be a persistence effect, since all constraints pertaining to various explanations about the role of persistence are shown in the R figure. Nevertheless, neither the direction nor the level of statistical significance can be derived from the mean decrease values produced here.

The output from the random forest model supports what we already know about CSD being a stable variable with consistent constraints that condition its variation across varieties of English. The fact that age emerges as an important predictor is interesting. An age effect could indicate a stable, age-graded variable, or a change in progress. I will examine age in more detail shortly.

I decided to use conditional inference trees after applying the random forest function to check for interactions that might be missed when using other regression methods. Conditional inference trees produce nodes that represent the most significant constraints that collectively create a model that best fits the distribution of the data, each as a separate node. These nodes are then attached to each other via branches that represent interactions amongst constraints. An average of the application value of the response (CSD/ deleted realisations) is then produced for the interactions.

Figure 4.2 represents the condition tree produced in R for the variable CSD for NS, the abbreviation *ps* stands for preceding segment, and the abbreviation *fs* stands for the following segment while *pos* stands for part of speech or grammatical categories. Looking at Figure 4.2, we notice that there is a dichotomy in respect of the effect of a preceding nasal, and other preceding consonant within the preceding segment constraint (node 1 labelled *ps* in the figure).

In the following segment constraint (node 2 labelled *fs*), we notice that glides and nasals are treated alike by NS of NZSED, and that pauses and vowels cluster, distinct from

- 105 -

other consonants. There is also a generational effect among NS, younger and middle-aged

speakers cluster and differentiate their variation of CSD for verbal categories (node 5 labelled

pos) depending on the following pause or vowel. Following pauses favour retention whereas

following vowels favour deletion. Older people, on the other hand, favour retention in both

these following environments.
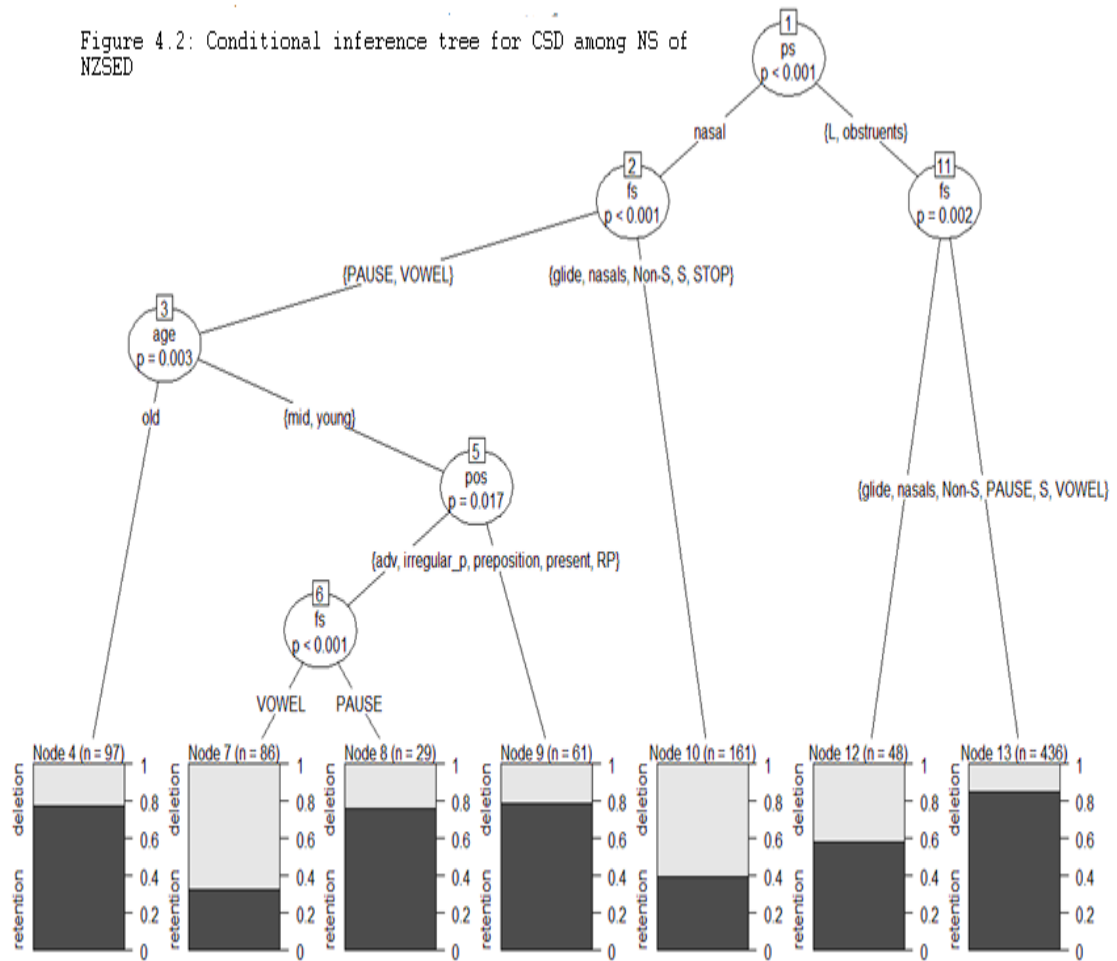
Figure 4.2: Conditional inference tree for CSD among NS of NZSED

The output from the conditional inference tree reiterates our findings from the abovementioned random forest's regression analysis. Phonological and grammatical constraints together with the social constraint of age are the most consistently selected factors conditioning variation in CSD.

The conditional inference tree is also able to illustrate interactions amongst constraints that cannot be captured through contingency tables (such as cross-tabulations in Rbrul) or by exploring the interaction of constraints in regression analysis in R. For example, a quick glance at the tree illustrates how phonological constraints interact, and how age and grammatical category also interact, and that grammatical category interacts with the following segment.

In conclusion, the output from the non-parametric methods suggests that the constraints that might be significant in conditioning variation on CSD by NS of NZSED are phonological (the preceding and the following environment), grammatical and social (generational effect). The output also suggests that social factors like gender and style are not as consistently selected for this variable for this sample. Based on this primary regression analysis, I will conduct parametric analysis to double check the output presented here.

4.1.2 Results for parametric regression for CSD among NS of NZSED

The first step for feeding a mixed model that explains CSD in NZSED among NS was to consider the output from the non-parametric regression for the variable as the baseline for choosing constraints for mixed modelling in Rbrul. This was to check if the constraints that are always reported to condition the CSD variable have a genuine effect on it. This is important because many of researchers who studied CSD seem to have replicated the same models for this stable variable, although some suggest that grammatical category, which is usually reported to be significant in varieties of native-speaker English, may not be a genuine effect on CSD (see Temple, 2003; Tagliamonte & Temple, 2005; Hazen, 2011) .

The next step was to conduct a detailed analysis of all potential constraints on CSD by trying out one constraint at a time to test their significance and then excluding any constraint that had a p-value greater than 0.05. Eventually, my model included the preceding segment constraint, the following segment constraint, the grammatical category constraint as well as age, gender and style-shifting as social constraints.

A point to reiterate before I start presenting the results for this first model is that, as I have mentioned previously, the coding of the grammatical category constraint as a conditioning factor on CSD in English differed in the literature. Therefore, I tried building two competing models for CSD by NS of NZSED to understand the two competing hypotheses about the grammar underlying the variation of CSD.

To this end, I coded the grammatical category in two ways; the first coding schema was designed to examine a functional hypothesis and the second to test a structural hypothesis. Given that uneven distribution of tokens is the norm rather than the exception in sociolinguistic studies, the alternate ways of coding of grammatical category/part of speech allow me to explore this claim a little further.

After running the two different models and based on the criteria provided in the methodology for choosing the best model fit[12], the function-based model was found to be the best statistical model for CSD among NS of NZSED which illustrates that they share an underlying grammar for variation on CSD that is built on a functional notion of grammatical category, i.e., its role in speech.

Based on this result, I carried out an investigation for the role of persistence within the framework of the functional notion of grammatical category as opposed to persistence within the framework of a form-based notion of grammatical category and this constraint failed to show statistical significance in either model.

The results of regression modelling in Rbrul report the statistical significance of constraints, constraint hierarchies, and factor weights. A constraint is statistically significant if it is smaller than the 0.05 cut-off value and a constraint's strength is usually reported in a unit called the range.

The range is calculated by subtracting the weakest category's factor weight from the strongest category's factor weight. The calculated range is then compared to the ranges of the remaining significant constraints. The highest figure or widest range defines the strongest constraint. This information enables us to position constraints with respect to each other according to their relative strengths (cf. Tagliamonte (2006)).

---

[12] The significance of the function-based model was determined by taking a chi-squared value as twice the difference in loglikelihood (=20) with degrees of freedom as the difference between the models =4), which results in a p-value of $p < 0.01$.

Comparison between function-based and form-based models for CSD among NS of NZSED					
Non-nested models: see methodology for best fit criteria					
function-based model of CSD (5 levels of grammatical categories)	N 935	Df 19	**LL** **– 440**	AIC 917	**R2** **total** **0.49**
Form-based model of CSD (two levels of number of morphemes per token)	N 935	Df 15	LL **– 450**	AIC 919	R2 total 0.44

A constraint hierarchy lists levels within a constraint according to their factor weights from highest to lowest. If a factor weight, which is a measure of the probability of the variant under investigation in a specific context, is anything over 0.50, it generally favours the application of the rule, while it disfavours it when it is below the 0.50 level.

However, for a comparative variationist exercise, it is more useful to understand the relative factor weights and their relative position within the constraint ranking. Because we are interested in understanding the underlying grammars that govern variation across NS groups as well as how far AMs tend to have a similar/different underlying grammar and not the exact values of factor weights.

My results from studying NS constraints on CSD in NZSED, based on my analysis of casual speech (Warren, 2002) generated the table 4.1 (the output of a parametric logistic regression analysis in Rbrul).

Table 4.1: Summary of best mixed-effects model for CSD among NS of NZSED. Application value: deleted /t,d/				
Degrees of freedom			15	
Cantered input probability			0.19	
Overall proportion of application value			0.32	
Log-likelihood			−321	
AIC			673	
R2 total			0.48	
Constraints	Log odds	N	Proportion of application value	Centered input probability
Following segment				
Stops	0.79	55	46%	0.69
Non-sibilant-fricatives	0.32	73	34%	0.58
Glides	0.27	48	33%	0.57
Nasals	0.16	39	39%	0.54
Sibilant-fricatives	0.14	66	36%	0.54
Vowel	−0.79	264	32%	0.31
Pause	−0.89	101	19%	0.29
				Range = 40
Preceding segment				
Nasals	1.1	334	46%	0.75
/l/	−0.18	161	21%	0.46
Stops	−0.38	94	18%	0.41
Fricatives (sibilant and Non-sibilant)	−0.54	57	11%	0.37
				Range = 38
Grammatical category				
Nouns and base-form-verbs	0.54	361	39%	0.63
Adjectives	0.09	138	32%	0.52
Irregular-past-and -regular past	−0.63	147	18%	0.35
				Range = 28
Age				
Young	0.57	171	43%	0.64
Old and middle-aged	−0.57	475	28%	0.36
				Range = 28
Speaker random	**Std = 0.87**			
Word random	**Std = 0.94**			
[gender, style-shifting, persistence]				

 Non-significant constraints on CSD are in square brackets at the end of the table.

Table 4.1 represents the results from running a logistic regression analysis with speaker and word considered as random effects. The constraints listed in the table are all statistically significant, and those that were found to be not significant for this group of NS of NZSED are in square brackets at the end of the table.

The results illustrate that when holding the speaker and word as random effects, the constraints that condition variation on CSD by NS in NZSED are phonological and grammatical in nature. The following segment constraint has the strongest effect on CSD in NZSED by NS; this result was also noted in earlier studies on CSD in NZE (Bell, 1977; Holmes & Bell, 1994), as well as in other NS varieties of English (Guy, 1980; Neu, 1980; Labov, 1989; Guy, 1991; Roberts, 1997; Tagliamonte & Temple, 2005; Smith et al., 2009; Hoffman & Walker, 2010).

The constraint hierarchy of the following segment constraint seems to follow other varieties of NS English where consonants favour deletion and vowels favour retention. My results suggest that stops and non-sibilant fricatives favour deletion the most followed by glides. Nasals and sibilant fricatives slightly disfavour deletion while vowels and pauses disfavour deletion the most.

Earlier findings regarding the following segment constraint in NZE listed similar results, whereby obstruents favour deletion the most, followed by liquids and glides and finally vowels and pauses disfavour deletion the most (Holmes & Bell, 1994).

My results also suggest that the preceding segment constraint has the second strongest effect size in predicting variation on CSD, whereas this constraint is usually reported to have low, if any, effect on CDS as suggested by the previous results reported for NZE (Bell, 1977; Holmes & Bell, 1994).

Holmes and Bell (1994) noted that both sibilant and non-sibilant fricatives behaved alike when they excluded the word *just* (which had the highest rate of deletion) from the analysis. Stops and nasals slightly favoured retention, while the /l/ liquid sound favoured retention the most. They suggested that the result of the liquid sound being the most favouring context of retention may be a result of the vocalisation of /l/ especially before /t/, a phenomenon that stops the /l/ from being a consonant, thereby neutralizing the consonant cluster reduction context.

The results from this study indicated a pattern going in the other direction from previous findings for NZE. In the NZSED data, a preceding nasal favours deletion the most, although words like *and* and negated forms like *don't* were excluded from the final regression analysis to avoid skewing the data. Preceding /l/ also favours deletion and stops and fricatives favour deletion the most. The range for this constraint is 38, making it the second strongest set of factors constraining CSD for these speakers.

As for the grammatical category constraint, my results support that nouns favour deletion, but slightly less than adverbs. Base form verbs seem also to favour deletion. Adjectives have a neutral effect on deletion, while Irregular past tense forms and regular past tense forms favour deletion the least with a slight difference in factor weights, with irregular past forms disfavouring deletion slightly less than regular past tense forms. Therefore, both past tenses were combined in latter analyses.

A similar result regarding the similar effects of regular and irregular past tense forms among NS of NZE was reported by Holmes and Bell (1994) who also noted that past tense forms, both regular and irregular, disfavour deletion with a slight difference in factor weight values. Guy (1991) also reported an almost identical effect for regular and irregular past tense forms, both disfavouring deletion in US dialects.

Age, in this study, was found to have an effect size the same as that of the grammatical category constraint on CSD, both with a range of 28. There seems to be an age-grading effect where young NS favour deletion at a higher rate than middle-aged and old NS who seem to cluster and behave differently from younger speakers.

This result is supported by evidence from running non-parametric conditional inference trees. This effect, nevertheless, interacts with the following segment constraint; older speakers favour deletion whether the following segment is a pause or a vowel. Younger and middle-aged speakers favour deletion if the following segment is a pause and the grammatical category of a token is verbal, whereas they favour retention if the following sound is a pause and the token is nominal. Unlike the findings regarding CSD in NZE in (Holmes & Bell, 1994), gender was not found to be statistically significant in predicting variation in CSD in NZSED.

4.1.3 Competing hypotheses and multiple models

It is worth mentioning before I move on to discuss the results for AMs for the variable CSD to illustrate that there were some competing hypotheses that I was trying to resolve other than the form-function distinction of the grammatical category that I have discussed earlier. For example, the literature on the variable CSD has recently added some new perspective on the constraints on variation aside from the well-attested phonological and grammatical constraints mentioned earlier: the effect of persistence and ways to measure it which have varied in the literature.

For example, Schleef et al. (2011) mention that what appears to be a persistence effect may be attributed to intra-personal variation and may not reflect genuine persistence. Starting from this note, I tried to figure out an explanation for persistence across variants, I, therefore, tried to test several operationalisations of the constraint.

First, I replicated Abramowicz's (2007) definition of persistence as a persistence effect across close tokens within the framework of function-based categorisation of grammatical category. I have also coded for Tamminga's (2014) psycholinguistic model where she reframes the concept of "persistence" and expands it to suggest that underlying forms at specific linguistic levels prime each other. But neither hypothesis was found to be significant among NS of NZSED.

The reason could lie in my operationalization of the theories, or in the fact that the evidence for persistence noted by Abramowicz (2007) and Tamminga (2014) is only valid in varieties of US English. Having noticed that, I tried to look for other plausible explanations within the framework of the broader concept of pragmatic and thematic coherence. One way to explain the patterns of "persistence" of some variants was obviously that they are linked to a given topic. Hence, if a speaker is talking about a specific topic, it seemed logical to expect him/her to reuse words and reuse specific variants.

With that in mind, I tried to code across various stages of a conversation, for formality levels (formal/ informal stances) and for topic typology (see Labov 2001). The topic seemed like a promising avenue when I first started coding, but as the data expanded and I started to handle four different corpora, the idea of consistency seemed impossible, meaning that the comparability of any analysis based on specific topics would be jeopardized[13].

I resolved to another way for coding thematic and prosodic coherence as the level of involvement, and interest respondents displayed when interviewed. But this coding may have caused a collinearity with formality and style-shifting, since they are basically measures of the same thing, how at ease is a person during talking. Running these two constraints in separate models and together with a coding for interaction revealed that formality, as well as topic, do not seem to explain much of the variation.

Part of the problem almost certainly lies in the nature of the data available to me. Not being a sociolinguistic corpus, the NZSED conversation component was restricted in time (about ten minutes) and to one topic (drink driving). It seems reasonable to conclude that these facts made it difficult for other typical sociolinguistic variables, such as formality and attention paid to speech, to emerge.

Even when looking at these results against earlier findings about the role of style-shifting in New Zealand English as a factor affecting the CSD variable, I must say that the comparison Holmes and Bell (1994) made between spoken corpora and the news is probably not the ideal way of operationalizing style-shifting, mainly because the news is a register in its own right; the news is written language, read-aloud and, therefore, arguably not best

---

[13] An interesting finding that arose from this attempt is that formality level, topic and style shifting do not seem to necessarily have a one-to-one correspondence. For example, for NS of NZSED, a conversation that is casual, according to the Labovian typology (2001), is usually associated with an informal topic like drink-driving ads, but the respondents were not always at ease. The resulting combination for this context was casual, informal topic, but not at ease.

regarded as on the same style continuum as formal to casual speech. A point to mention though is that CSD is seldom found to be as sensitive to style-shifting as some other variables, such as (ING).

I finally resolved to look for prosodic features to explain the persistence of one variant over another. These mainly included auditorily identifiable prosodic features like stress, intonation, and pitch. I noticed that the simple distinction between emphatic/no-emphatic under the constraint prosody seemed to explain a lot of the variation. When people are very involved in a conversation and/or have strong feelings about a topic, they tend to be more auditorily emphatic.

This is especially so when discussing something like drink-driving ads which carry a heavy cultural load in New Zealand, with strong attitudes at two extreme ends. I tried to ensure consistency and replicability in operationalizing this constraint. The effect of emphatic versus non-emphatic turned out to be categorical in the sense that all emphatic stances had released /t/, or /d/, whereas other stances were reduced, so I dropped this constraint from further analysis.

4.2 Results for CSD among AMs

Following the same methodology that I have applied to the NS of NZSED, I conducted non-parametric regression analyses for the AMs of my sample to be able to get a clear picture of the patterns of variation existing for them.

4.2.1 Results for non-parametric regression for CSD among AMs: all age groups

Figure 4.3 represents the visualisation of random forests illustrating variable importance, reported in units of mean decrease in accuracy (mean decrease Gini). The figure suggests that phonological constraints and individual speakers are probably significant in conditioning CSD variation among NNS. Grammatical category seems also to play a role together with persistence. By contrast, NNS-specific factors seem to have very weak effects on the conditioning of CSD.

```
Figure 4.3: Variable importance for constraints conditioning CSD
among AMs: all age groups
Following Segment                                                    ●
Preceding-segment                                          ●
Speaker                                                    ●
Grammatical-                                               ●
category
Realization-of                              ●
-prior-token
Style                                    ●
Number-of-                               ●
morphemes of-
prior-token
Self-reported-                           ●
proficiency
Number-of-                ●
morphemes for-
current-token
NNS-specific-            ●
constraints
             0  10  20  30  40  50  60  70  80  90 100 150
                       Mean decrease Gini
```

Bold and italicised constraints are different measures of persistence

I was unable to produce a conditional inference tree to account for the variation patterns of AMs as a collective group with three age cohorts. Failing to conduct non-parametric regression analysis to account for CSD by AMs motivated me to look for possible reasons for that. One possibility is that there might be some interactions across the constraints, hence R's inability to partition constraints in a regression analysis.

4.2.2 Results for parametric regression for CSD among AMs: all age groups

I started this investigation by trying to conduct a parametric regression for all groups of AMs. Nevertheless, the regression analysis conducted in R and Rbrul revealed many warnings like convergence warnings. Sometimes the regression was interrupted, and the software would crash, this is usually the result of too many convergence problems and interactions. I, accordingly, tried to combine internal levels within various constraints. Nevertheless, the rank ordering of the constraints and their associated p-values mismatched.

Table 4.2: Summary of best mixed-effects model for CSD among AMs: all age groups. Application value: deleted /t,d/

Degrees of freedom			12	
Centered input probability			0.11	
Overall proportion of application value			0.36	
Log-likelihood			-1511	
AIC			3046	
R2 total			0.24	

Constraints	*Log odds*	*N*	*Proportion of application value*	*Centered input probability*
Following segment				
Nasals	0.56	136	47%	0.64
Obstruents	0.41	964	42%	0.6
Glides	-0.03	304	33%	0.5
Vowels	-0.41	957	31%	0.4
Pauses	-0.52	459	28%	0.37
				Range= 27
Preceding segment				
Nasals	0.84	65	40%	0.70
Fricatives	0.45	345	39%	0.61
Liquids	0.075	172	31%	0.52
Stops	-0.003	75	29%	0.50
				Range = 20
Networks				
Non-Arab	0.60	2271	40%	0.65
Arab	-0.40	549	17%	0.35
				Range = 30
Speaker random	**Std = 0.74**			
Word random	**Std = 0.86**			
[sex,				
Grammatical-				
category,				
style-shifting,				
persistence]				

Non-significant constraints on CSD are in square brackets at the end of the table.
Bold and italicized range values indicate convergence issues in the model.
Age was not run in this model because it interacts with networks.

A closer look at the Rbrul output reveals some interactions in the model. Based on the confusing Rbrul output in table 4.2, the rank ordering, as suggested by the range, is confusing since the preceding segment constraint has a range of 20 and the networks constraint has a range of 30, the later with a higher range than the range allocated to the following segment constraint which was listed by Rbrul as the constraint with the strongest effect on CSD among AM speakers of NZE.

To understand this confusing output, I started a primary investigation in Rbrul via cross-tabulating. I cross-tabulated each age-group alone with other constraints, that I believe may affect variation on CSD for AMs, and I noticed that young AMs cluster and display patterns different from those who are older and/or middle-aged. At this point, it seemed rational to try and sub-classify the respondents into separate groups based on (i) the data distribution as suggested by non-parametric and parametric methods, and (ii) NNS-specific social factors.

Young AMs arrived in New Zealand at 6 or less years old, they have non-Arab networks and describe themselves as NS of NZE, and they also identify themselves as Kiwis. Moreover, younger speakers seemed more NS-like to me when I met them, and they may have a different underlying grammar from the old and middle-aged AMs.

Middle-aged and old AM groups acquired English integratively as adults in New Zealand, only two middle-aged women received some formal introduction to the English language at school and even these two stated that they learned English in New Zealand. This group also have self-reported proficiency levels ranging between good to excellent, but they never describe themselves as NS-like.

This group also seem to identify themselves as migrants to New Zealand rather than New Zealanders. The older speakers keep an only Arab network, and middle-aged speakers prefer a mixed network of Arabs and Kiwis. This difference in network preference probably

arises because middle-aged AMs have been working in New Zealand, whereas old AMs either have their own businesses or do not work at all.

4.3 Results for CSD among separate age-groups of AMs

Based on this rationale for splitting AM age groups into two subgroups, I reran the non-parametric regression for old and middle-aged speakers once and then for young AMs . I found that I did not face any of the statistical issues I faced earlier when I was trying to provide a model that accounts for all AM-age groups collectively (i) convergence warnings, (ii) mismatch between proportion of application value and the associated factor weight and (iii) rank ordering incompatible with p-values of constraints in a model.

4.3.1 Results for non-parametric regression for CSD among old and middle-aged AMs

Figure 4.4 represents the conditional inference tree for old and middle-aged AMs, and the general pattern is to retain the final consonants. The figure also reveals some interactions amongst constraints; the following segment constraint is the most significant for older and middle-aged AMs, and it interacts with self-reported proficiency and time spent in New Zealand.

Figure 4.4: Conditional inference tree for CSD among old and middle-aged NNS of NZE
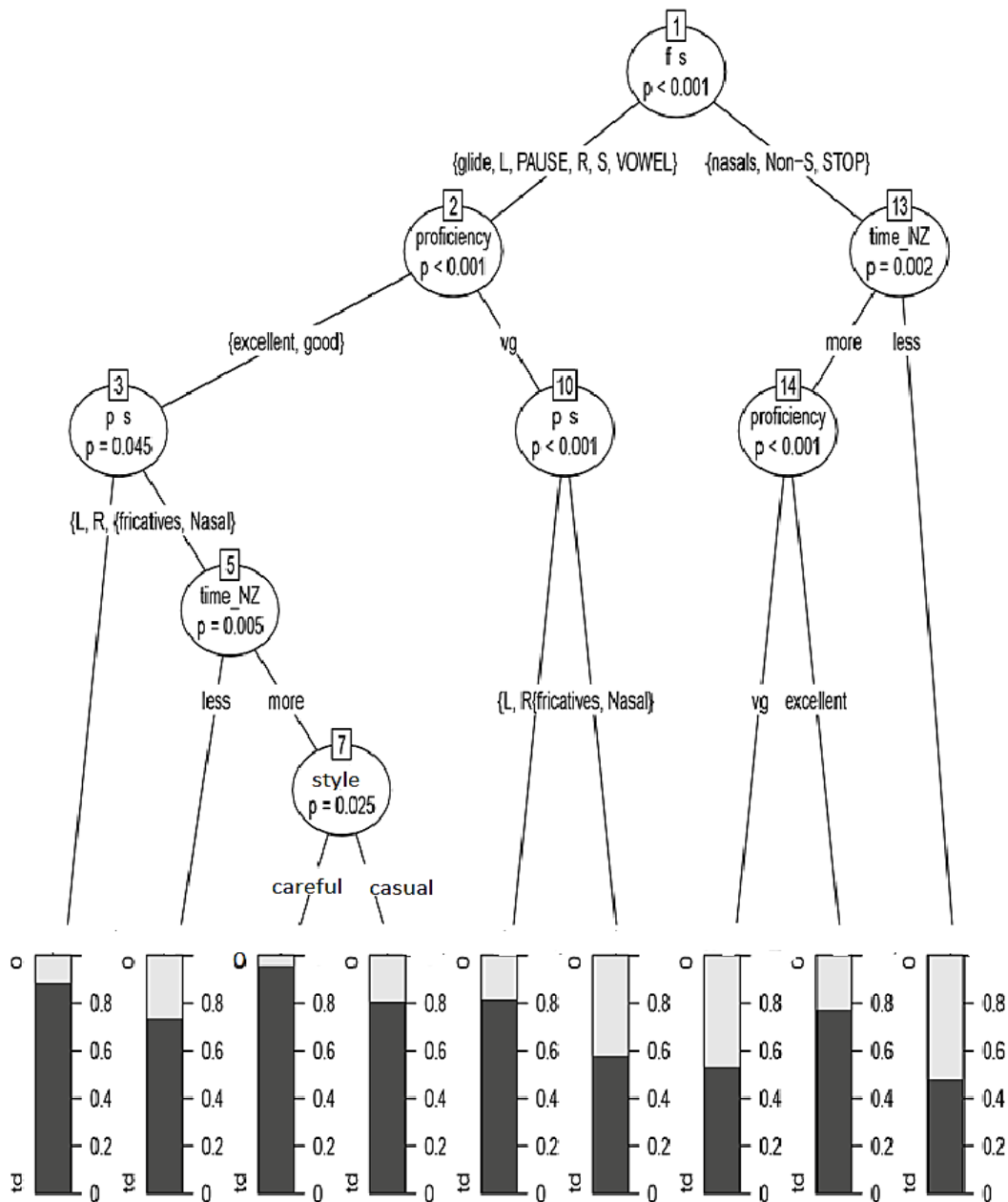
Figure 4.4: Conditional inference tree for CSD among old and middle-aged AMs

4.3.2 Results for non-parametric regression for CSD among young AMs

Figure 4.5 illustrates that the preceding segment seems to be the most significant constraint that conditions CSD variation (node 1 labelled *Ps)* among young AMs . Preceding nasals interact with the grammatical category constraint (node 2 labelled *grammatical category*); verbal constructs favour retention, whereas nominal ones interact with persistence represented by two constraints: the realisation of the previous token and the morphological status of the previous token.

If the grammatical category is nominal and the realization of the previous token is a maintained /t,d/ or was a neutralization context, then deletion is favoured. But if the previous realization was a deleted consonant, then the morphology of the previous token is important. If the previous token is monomorphemic and has a deleted /t,d/, then the current token will most likely have a deleted stop. When the previous token is bimorphemic (past tense verbs, derived nouns), then the current token has an even stronger probability of having a deleted stop.

When the preceding segment is a non-nasal consonant, there is an interaction with the following segment constraint (node 9 labelled Fs). Vowels and /r/ are treated separately from other following consonants in that they highly favour retention. This observation further advances the proposition that young AMs are more NS-like, since they are non-rhotic. Other following consonants favour retention to a lesser degree compared with following vowels.

Figure 4.5: Conditional inference tree for CSD among young NNS of NZE
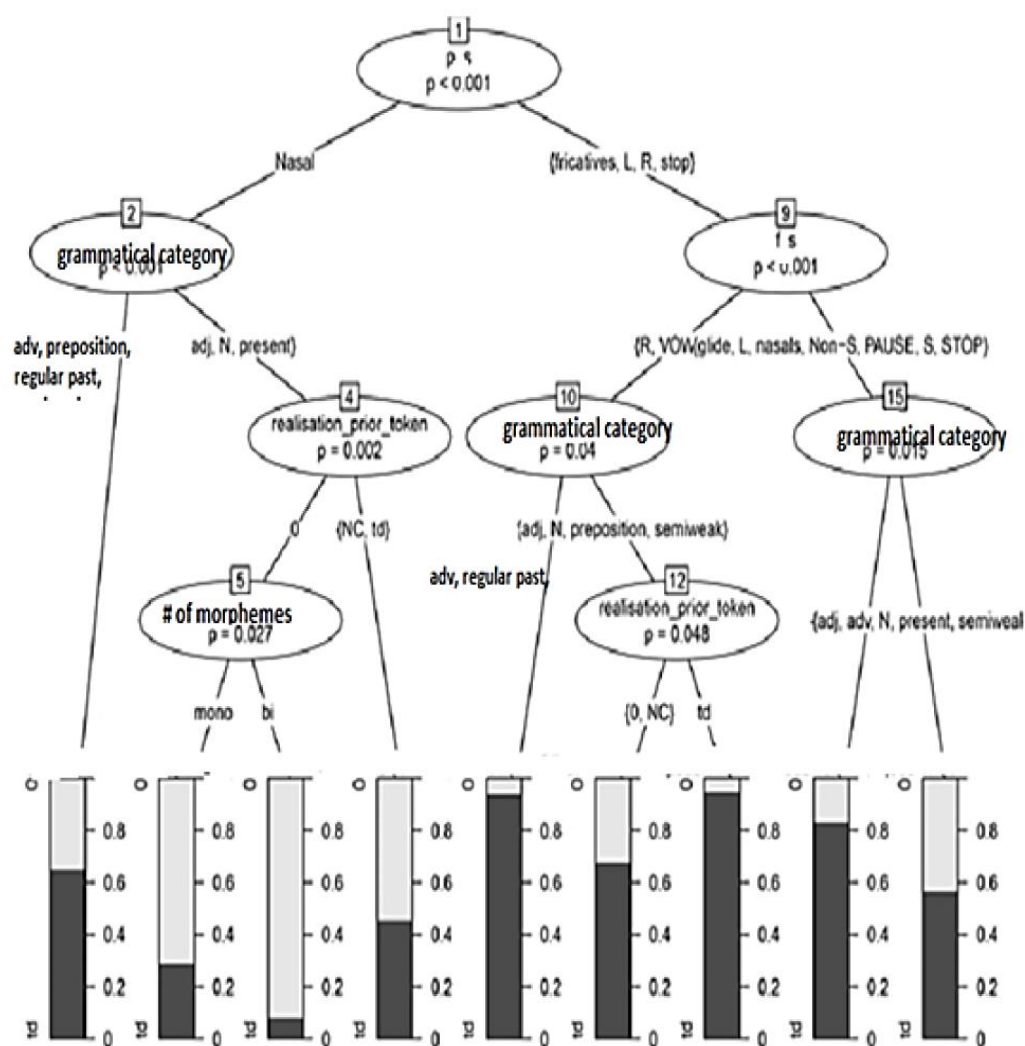
Figure 4.5: Conditional inference tree for CSD among young AMs

4.3.3 Results for parametric regression for CSD among old and middle-aged AMs

After running competing models and based on the criteria provided in the

methodology for choosing the best model fit, a function-based model was found to be the best

statistical model for CSD by AM speakers of NZE[14]. This means that AMs share an

underlying grammar for variation on CSD that differentiates parts of speech based on their

grammatical function, not their morphological status.

The results for the old and middle-aged AMs model seem to be more accurate than

the full model in describing AM patterns of variation for CSD. The methodology of figuring

this out is by conducting a chi-square test of model goodness, since the full model and the old

and middle-age-only model are nested. The results indicate that the smaller nested model, i.e.,

the old and middle-aged model, is significantly better than the full one with a p-value of

3.88e-292[15].

---

[14] These two models are not nested since one of them has a constraint that represents number of morphemes and the other has a constraint that represents grammatical categories, to see whether NNS have an underlying grammar that pays attention to word class or morphology. The AIC number for the function-based model denotes that this model is a better fit to the data.

[15] These two models are nested, so I used chi-square test to find the best fit. Chi-square = 1342, df = 2, p = 3.88e-292

Table 4.3: Summary of best mixed-effects model for CSD among AMs: old and middle aged. Application value: deleted /t,d/			
Degrees of freedom		10	
Centered input probability		0.20	
Overall proportion of application value		0.28	
Log-likelihood		-840	
AIC		1699	
R2 total		0.32	

Constraints	Log odds	N	Proportion of application value	Centered input probability
Following segment				
Nasals	0.56	32	38%	0.64
Obstruents	0.49	587	36%	0.62
Glides	-0.15	187	23%	0.46
Vowels	-0.36	505	22%	0.41
pauses	-0.54	293	21%	0.37
				Range= 27
Preceding segment				
Nasals	0.84	65	40%	0.70
Sibilant and non-sibilant-fricatives	0.45	345	39%	0.61
Stops	0.075	172	31%	0.52
Liquids	-0.003	75	29%	0.50
				Range = 20
Speaker random	**Std = 0.87**			
Word random	**Std = 0.86**			
[gender, grammatical category, style-shifting, persistence]				

Non-significant constraints on CSD are in square brackets at the end of the table.

Age was not included because it was controlled for by having two separate AM age groups.

Table 4.3 shows the output for parametric logistic regression analysis in Rbrul. It illustrates that the constraints on CSD among old and middle-aged AMs are the following segment constraint and the preceding segment constraint. But grammatical category and gender do not condition variation on CSD for this subgroup of AMs. The rate of CSD reported here, 28%, is less than that reported for NS of NZSED as well as that reported for young AMs and NS of NZE.

The results illustrate that old and middle-aged AMs patterns for the following segment seem to make a dichotomous split between consonants which favour deletion, and vowels and pauses which favour retention. Glides are treated more like vowels, slightly favouring retention. The preceding segment constraint is the second strongest constraint conditioning CSD for this subgroup of AMs as the preceding sonorant nasals and fricatives favour deletion whereas preceding stops and liquids have a neutral effect on CSD.

The grammatical category constraint was found to be non-significant for this subgroup of AMs. This result might indicate that old and middle-aged AMs' patterns are primarily phonologically driven, and they are steadily treating past tense forms as unanalysed lexical items.

4.3.4 Results for parametric regression for CSD among young AMs

The results for the young AM's model seems to be more accurate in describing patterns of variation for CSD among young AMs than a combined model with all age groups. The methodology of figuring this out is by conducting a chi-square test of model goodness, since the full model and the young AM's model are nested. The results indicate that the smaller nested model, for young AMs, is significantly better than the full one with a p-value p < 0.001[16].

The results indicate that young AM patterns are a combination of old and middle-aged AMs' patterns, and NS patterns. As table 4.4 suggests, young AMs share the same significant constraints found among NS of NZSED, although there is a different rank ordering and internal hierarchies as well as different effect size of constraints. Young AMs' rate of deletion is comparable with but slightly higher than that of NS: 36% versus 31%.

---

[16] Chi-square = 167.468, df = 7, p = 8.58e-33

Table 4.4: Summary of best mixed-effects model for CSD among AMs: young. Application value: deleted /t,d/				
Degrees of freedom			13	
Centered input probability			0.21	
Overall proportion of application value			0.36	
Log-likelihood			-1403	
AIC			2833	
R2 total			0.37	
Constraints	*Log odds*	*N*	*Proportion of application value*	*Centered input probability*
Preceding segment				
Nasals	0.64	1231	47%	0.65
Sibilant and non-sibilant-fricatives	0.17	574	29%	0.54
Stops	-0.62	582	24%	0.44
Liquids	-0.55	229	23%	0.38
				Range = 27
Following segment				
Nasals	0.52	130	47%	0.63
Obstruents	0.45	888	43%	0.61
Glides	-0.03	282	33%	0.49
Vowels	-0.50	888	31%	0.40
Pauses	-0.44	428	29%	0.38
				Range = 25
Grammatical category				
Nouns				
Base-form-verbs	0.25	867	38%	0.54
Irregular past	0.17	1282	37%	0.54
Regular past	-.001	176	35%	0.50
	-0.41	291	21%	0.41
				Range = 13
Speaker random	**Std = 0.91**			
Word random	**Std = 0.81**			
[gender, style-shifting, persistence]				

Non-significant constraints on CSD are in square brackets at the end of the table.

Young AMs diverge from NS in the ordering of constraints. The preceding segment constraint is the strongest constraint for this sub-group of NNS, whereas the following segment constraint had the strongest effect for NS of NZSED. Regarding the internal hierarchy of the preceding segments, young AMs are more like NS in that preceding nasals favour deletion most, but they otherwise differ in their ranking of other levels.

For young AMs, preceding nasals favour deletion the most and other consonants, excluding fricatives, favour retention, with liquids being the least favouring of deletion. A point to mention is that the strength of preceding segment constraint seems to be only slightly stronger than the following segment as a young-AM constraint, but young AMs have lower effect size for these two constraints compared with those NS of NZSED display.

The following segment constraint has the second strongest effect on CSD among young AMs with an internal hierarchy slightly different from that of NS of NZSED; following consonants favour deletion and vowels and pauses favour retention. As we saw in Table 4.3, older and middle-aged AMs treat following glides like following vowels; there seems to be continuity between old and middle-aged AMs and younger ones, with internal orderings in this constraint the same for both age groups. Like NS of NZSED, young AMs separate the following glides from the following vowels, with glides having a neutral effect on CSD.

Like NS of NZSED, young AMs do show an effect for the grammatical category; it is the third constraint conditioning variation on CSD. However, the range for this constraint was much greater for NS 28 versus 13 for young AMs. This result may shed light on the underlying grammar of young AMs.

The internal hierarchy of the grammatical category constraint is slightly different from that of NS; monomorphemic and bimorphemic nouns, as well as monomorphemic base form verbs, favour deletion, whereas regular past tense forms favour retention the most.

Irregular past tense forms have a neutral effect on CSD diverging from NS norms, for whom, irregular past forms disfavour deletion.

4.4 Summary: CSD findings among NS of NZSED and AMs

In the first part of this chapter, I introduced the results of parametric and non-parametric regression analysis for the variable CSD. I illustrated how I implemented non-parametric tests in R for initial modelling of the variable CSD; I also used the output from the non-parametric tests to justify some of the unusual findings for my data sets.

I started with the results for NS of NZSED then for AM speakers of NZE. I also demonstrated that it is better to separate the analysis of the AMs into two age groups, old and middle-aged and young AMs and I have provided a detailed rationale for this decision.

I also tried to introduce a comparison between the NS and the two AM groups in terms of the patterns of variation prevailing for them. The results indicated that AMs seem to acquire the main significant constraints that condition variation on CSD. AMs, however, seem to be inclined to exhibit "transformation under transfer", a finding that will be discussed in depth in the following chapter.

There are also differences between the old and middle-aged AMs and young AMs which suggests that old and middle-aged speakers seem to be driven by phonological constraints, whereas young AMs seem to display grammatical constraints that are similar but also diverging from that of NS of NZE. Table 4.5 illustrates the general patterns for the three groups, highlighting points of convergence.

Table 4.5: Summary of best mixed-effects model for CSD for NS of NZSED, old and middle-aged AMs and young AMs living in Wellington. Application value = deletion of /t,d/		
NS	Old and middle-aged AMs	Young AMs
1.Following segment Stops Non-sibilant fricatives Glides Nasals Sibilant fricatives Vowels Pauses Range=40	*1.**Following segment*** Nasals obstruents Glides Vowels Pauses Range=27	*1.Preceding segment* Nasals Fricatives Liquids Stops Range= 27
*2.**Preceding segment*** Nasals /l/ Stops Fricatives Range= 38	*2.**Preceding segment*** Nasals Fricatives Stops Liquids Range= 19	*2.Following segment* Nasals Obstruents Glides Vowels Pauses Range=25
3.Grammatical-category Nouns and base-form-verbs Adjectives (neutral) Regular-and-irregular Past Range = 28	NA	*3.Grammatical category* Nouns Base form verbs Irregular past tense Regular past tense Range = 13
4.Age Young Old and middle-aged NS		
[gender, persistence, style-shifting]	**[gender, persistence, style-shifting, grammatical-category]**	**[gender, persistence, style-shifting]**

Bold and *italicized* values indicate the points of similarities and/or difference across the three Groups of speakers.

In the second part of this chapter, I will discuss the results for the second variable, (ING), and I will use the same methods that I have used to present the results for the variable CSD: (i) by introducing exploration and initial modelling techniques through the implementation of two non-parametric tests in R, and (ii) by undertaking a full parametric regression analysis. I will start by presenting the data for the NS of New Zealand English. I will then return to presenting the results for the AM speakers of New Zealand English and demonstrate, once more, that it is more accurate to discuss the results of AMs under two separate age-groups: old and middle-aged versus young AM speakers of NZE.

4.5 Results for (ING) among NS of NZSED

After testing competing models for (ING) in varieties of NS English, I reached the conclusion that a model inclusive of the constraints, previously found to be significant across varieties of NS English, with speaker and word accounted for as random effects provided the best fit to the data.

4.5.1 Results for non-parametric regression for (ING) among NS of NZSED

Figure 4.6 reveals that the speaker constraint probably has the strongest effect on (ING) variation among NS of NZSED, another reason to include speakers as a random effect in the parametric regression analysis. The grammatical category seems to be significant as well, followed by the following segment. The preceding segment, age and gender seem to have a reduced effect compared with the previous three constraints, but they do seem to have a role in conditioning (ING) among NS of NZSED.

```
┌─────────────────────────────────────────────────────────────────┐
│ Figure 4.6: Variable importance for constraints conditioning (ING)│
│ among NS of NZSED                                                 │
├───────────────────────────────────────────────────────────────────┤
│ Speaker                                                         ● │
│                                                                   │
│ Grammatical-                                                      │
│ category                                            ●             │
│                                                                   │
│ Following-segment                            ●                    │
│                                                                   │
│ Preceding-segment                  ●                             │
│ Age                                ●                             │
│ Gender                       ●                                   │
│              0  10  20  30  40  50  60  70  80  90  100  150      │
│                        Mean decrease Gini                         │
└───────────────────────────────────────────────────────────────────┘
```
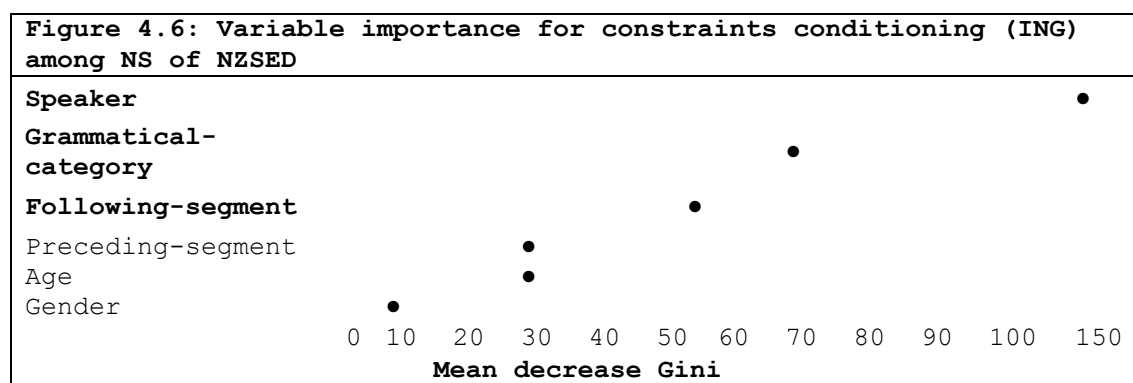
The next step was to create a conditional inference tree for NS of NZSED for the variable (ING) to locate possible interactions amongst constraints. Figure 4.7 illustrates that there seem to be interactions between the grammatical category, the following segment and age constraints.

More specifically, the grammatical category constraint (node 1 labelled *GC*) separates discourse markers and pronouns from other categories. These two groups of nominals seem to favour a velar realization whereas the rest of the categories interact with the following segment constraint.

If the following segment (node 3 labelled *fs*) is a pause, other grammatical categories slightly favour a velar variant, but if the following segment is another consonant or a vowel, then the alveolar nasal is favoured by all speakers, but especially so by middle-aged and older speakers.

Figure 4.7: Conditional inference tree for (ING) among NS of NZSED
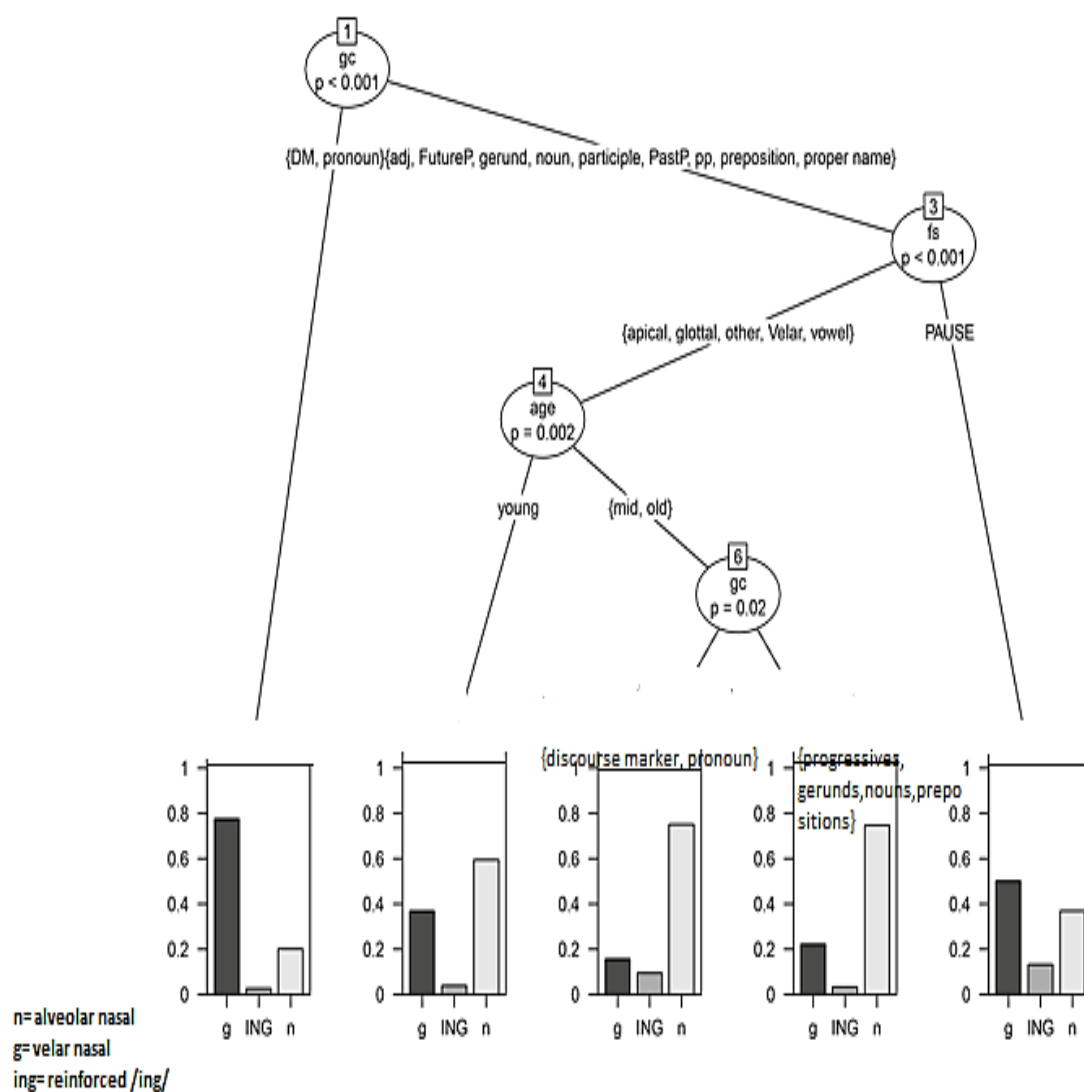
Figure 4.7: Conditional inference tree for (ING) among NS of NZSED

4.5.2 Results for parametric regression for (ING) among NS of NZSED

The next step was to conduct a regression analysis for the variable (ING) to determine

the constraints that are statistically significant and those which are non-significant among

NS of NZE. The results are displayed in table 4.6.

Table 4.6: Summary of best mixed-effects model for(ING) among NS of NZSED. Application value = alveolar nasal				
Degrees of freedom				12
Centered input probability				0.60
Overall proportion of application value				0.63
Log-likelihood				-404
AIC				832
R2 total				0.40
Constraints	*Log odds*	*N*	*Proportion of application value*	*Centered factor weight*
Grammatical-category				
Progressive	0.86	332	76%	0.70
Nominal gerunds	0.73	192	71%	0.68
Participles	0.68	76	70%	0.66
Adjectives	0.56	25	69%	0.64
Nouns	0.50	36	56%	0.62
Discourse markers	-1.25	36	22%	0.22
Pronouns	-2.1	64	19%	0.11
				Range = 59
Following segment				
Apical	1.17	126	83%	0.76
Velar	0.44	26	73%	0.61
Other consonants	0.03	227	63%	0.51
Vowels	-0.27	281	62%	0.43
Pauses	-1.37	101	38%	0.20
				Range = 56
Speaker random	**Std =0.92**			
Word random	**Std =0.006**			
[age, gender, preceding segment]				

Non-significant constraints on(ING) are in square brackets at the end of the table.

The results for NS of NZSED illustrate that only two linguistic constraints are significant on (ING): grammatical category and the following segment. Grammatical constraints on the variable (ING) has the strongest effect on (ING) by NS of NZSED, but the constraint strength is very close to the following segment constraint as suggested by the range value for both constraints (59, 56).

The results for the grammatical category constraint do not exactly replicate the Labovian nominal-verbal continuum; (Labov, 2001) notes that nominal categories favour the velar nasal variant the most and more verbal categories favour the alveolar nasal variant (progressive – gerundial participle - adjectives - gerunds and nouns). The results for NZSED are best explained in terms of a split between discourse markers and pronouns (which favour velar nasals) and all other grammatical categories (which favour the alveolar).

These results suggest that (adverbial) nominal gerunds and verb participles are treated alike by NS of NZSED. This conclusion, combined with the fact that adjectives and nouns, which are usually reported to favour the velar nasal, also slightly favour the alveolar nasal, suggests the nominal-verbal distinction may be fading in this corpus of NZE speakers.

These results slightly diverge from previous results for (ING) variation in NZE. Bell and Holmes (1992) found that the results for grammatical category followed the Labovian continuum with nominal categories favouring the velar nasal and verbal categories favouring the alveolar nasal. Discourse markers and pronouns were reported to be the most favouring of the velar nasal realization with a big gap between them and other grammatical categories. My findings also contrast with Schleef et al. (2011) findings in Scotland where these categories bracketed the variation space for (ING). This means that the ranking of different grammatical categories shown here is apparently dialect-specific.

The following phonological context is the second strongest constraint on (ING) for NS of NZSED with following vowels and pauses favouring velar nasals. It is worth

mentioning here that this constraint's strength is dialect-specific. For example, in southern dialects of the United States, the following phonological segment has a strong effect (Houston, 1985), whereas, in other dialects of the U.S., this constraint seems to have "no strong phonological conditioning before the following velars or apicals"(Labov, 2001:87). The dialect-specific nature of the following context will be important when I discuss the acquisition of variation of (ING) by AMs, as we will want to see whether AMs acquire these local constraints on (ING).

The preceding phonological context, as well as the social constraints that I have studied for NS of NZE, were found to be non-significant for NS of NZSED. In contrast with previous results for NZE (Bell & Holmes, 1992), gender and age are non-significant constraints on (ING). This result is probably due to the model including speaker and word as random intercepts, so it avoids Type I errors such as those that result in exaggerating social factor effects.

Style-shifting, which across many varieties of NS English constraint has been found to be a significant factor conditioning (ING), was also found to be non-significant. This might be because the NZSED conversations do not contain tasks of clearly different formality levels. Therefore, it would be hasty to state that style-shifting is no longer a significant constraint on (ING) in NZE, and more research needs to be conducted to explore this.

4.6 Results for (ING) among separate age-groups of AMs

Following the same methodology that I outlined earlier, I conducted non-parametric analyses for the NS of my sample to gain a clear picture of the patterns of variation influencing their use of (ING).

4.6.1 Results for non-parametric regression for (ING) among AMs: all age groups

The results for the importance of different variables shown in Figure 4.8 are reported in units of mean decrease Gini. It suggests that the speaker constraint is the most significant for conditioning (ING) variation among AMs, another reason why speaker should be accounted for as a random effect. The grammatical category and phonological constraints seem to play a role together with various measures of persistence (in bold italics). NNS-specific factors seem to have little or no effect on (ING) constraint because they were not returned by R as shown in Figure 4.8.

The next step was to produce a conditional inference tree to visualize potential interactions amongst constraints conditioning (ING) for AMs. The grammatical category seems to be the most significant constraint with progressive and adverbs patterning differently from the rest of the grammatical categories. When the grammatical category is progressive, and/or an adverb and the token was not the initial production of an (ING) word, the alveolar nasal is favoured.

Other grammatical categories, when preceded by a prior reinforced [ɪŋg] favour reinforced [ɪŋg]. If the previous token had an apical realization and is verbal, an alveolar realization is favoured. If the previous realization is velar and the token is nominal, the velar realization is favoured. A reinforced (ING) which is referred to as a velar nasal plus stop, i.e., [ɪŋg] (Wells, 1982) describes the variant where words like *during* and *watching* are realised with the same pronunciation as *finger* and *linger,* i.e., with a velar stop after the nasal.

```
┌─────────────────────────────────────────────────────────────────────────┐
│ Figure 4.8: Variable importance for constraints conditioning (ING) among │
│ AMs: all age groups                                                       │
├─────────────────────────────────────────────────────────────────────────┤
│ Speaker                                                                   │
│ Grammatical-                                                              │
│ category                                                    ●             │
│ Following-segment                                  ●                      │
│ Preceding-segment                  ●                                      │
│ Realization-of-                                                           │
│ prior-token                        ●                                      │
│ Prior-                                                                    │
│ grammatical-                   ●                                          │
│ category                                                                  │
│ Distance-between-                                                         │
│ tokens                     ●                                              │
│             0 10  20  30  40  50  60  70  80  90  100    150              │
│                      Mean decrease Gini                                   │
└─────────────────────────────────────────────────────────────────────────┘
```
Bold and vitalised constraints are different measure of persistence

4.6.2 Results for parametric regression for (ING) among AMs: all age groups

Table 4.7 shows the results of the Rbrul regression analysis of the data for all AMs. The constraints listed in Table 4.7 are all statistically significant, and those that were found to be not significant for this group of AMs are in square brackets at the end of the table.

Based on the results of the non-parametric regression analysis for (ING) by AMs of New Zealand English, I have attempted to include persistence in the model and results from the regression analysis in Rbrul suggests that the input from the non-parametric regression were accurate as persistence is found to be statistically significant in conditioning (ING) among AMs.

Table 4.7: Summary of best mixed-effects model for (ING) among AMs: all age groups. Application value = alveolar nasal

Degrees of freedom				16
Centered input probability				0.56
Overall proportion of application value				0.55
Log-likelihood				-735
AIC				1503
R2 total				0.47

Constraints	*Log odds*	*N*	*Proportion of application value*	*Centered factor weight*
Grammatical-category				
Progressive	0.86	565	74%	0.70
Participles	0.28	35	***46%***	***0.67***
Nominal-gerunds	0.15	305	***54%***	***0.66***
Discourse-markers	-0.152	37	***29%***	0.64
Nouns-and-pronouns	-0.39	342	***33%***	0.62
Adjectives	-0.75	62	***32%***	0.22
				Range = 48
Preceding segment				
Liquids	0.56	82	***60%***	0.64
Velars	0.38	200	***72%***	0.60
Other-consonants	-0.33	722	54%	0.42
Apicals	-0.61	342	50%	0.35
				Range = 29
Persistence				
Prior-alveolar	0.43	751	65%	0.61
Prior-reinforced (ING)	-0.05	336	47%	0.50
Prior-velar	-0.38	259	43%	0.41
				Range = 20
Following segment				
Pause	0.16	201	***55%***	0.54
Apical	0.12	169	***62%***	0.53
Other-consonants	0.11	581	***57%***	0.53
Vowels	-0.39	395	40%	0.40
				Range = 14
Speaker random	**Std = 0.93**			
Word random	**Std = 1.2**			
[age, gender]				

Pronouns are very few so were combined with nouns.
Bold and italicized values for the proportion of application value and the corresponding factor weights suggest interactions in the model.
non-significant constraints on (ING) are in square brackets at the end of the table.

The grammatical category constraint has the strongest effect on (ING) by AMs with progressive structures favouring the alveolar nasal the most. Participles and gerunds act alike and favour alveolar nasals as do discourse markers and other nominal categories. Only adjectives favour the velar nasal.

The preceding segment constraint has the second strongest effect on (ING) by AMs. Liquids and velars favour the alveolar nasal the most, whereas apicals favour velar nasals the most.

Persistence constraint has a strong effect on (ING). When the realization of the previous token was an alveolar nasal, the current token favours an alveolar realization; if the realization of the previous token is a velar realization, the velar realization is favoured in the current token. The following segment constraint has the weakest conditioning effect on (ING) among AMs, with following vowels favouring velars.

Based on results for other AMs in the acquisition of (ING) variation (e.g. Schleef et al. 2011), the full model for all AMs produced expected results. However, a closer look at it reveals some problems such as the lack of correspondence between the proportion of application value and associated factor weight. These are highlighted in italics in Table 4.7. This suggests that there are untested interactions between factors in the model.

I have started with a full model for the AMs which includes all age-groups. Later we will see that it is better to present the AMs data in sub-groups of old and middle-aged and younger ones (as we also saw for CSD). I explored two separate models for AMs by splitting the AMs into two sub-groups: old and middle-aged versus young AMs. I then ran chi-square test to figure whether the sub-models are significantly better than the full one for AMs of NZE.

4.6.3 Results for parametric regression among old and middle-aged AMs

The results indicated that a separate old and middle-aged model is significantly better than a model that includes all NNS of different ages together[17]. As noted when I introduced separate models of CSD for young AMs and old and middle-aged AMs, making this split in the AMs corpus seems logical if we want to explore the acquisition of variation. Table 4.8 represents the results for the old and middle-aged AMs.

Table 4.8: Summary of best mixed-effects model for (ING) among AMs: old and middle-aged. Application value = alveolar nasal				
Degrees of freedom			9	
Centered input probability			0.45	
Overall proportion of application value			0.49	
Log-likelihood			-477	
AIC			971	
R2 total			0.44	
Constraints	*Log odds*	*N*	*Proportion of application value*	*Centered factor weight*
Grammatical category				
Progressive	0.93	332	66%	0.72
Nominal-gerunds	0.27	226	51%	0.57
Nouns-and-pronouns	-0.26	235	27%	0.44
Adjectives	-0.34	26	26%	0.23
				Range = 49
Preceding segment				
Liquids	0.62	61	64%	0.65
Velars	0.36	123	56%	0.59
Other-consonants	-0.43	421	46%	0.40
Apicals	-0.56	214	44%	0.36
				Range = 29
Speaker random	**Std = 0.92**			
Word random	**Std = 1.01**			
[gender, following segment, persistence]				

Non-significant constraints on (ING) are in square brackets at the end of the table.

---

[17] Chi-square = 517.262, df = 7, p = 1.56e-107

The results in Table 4.8 indicate that the grammatical category has the strongest effect size on (ING) among old and middle-aged AMs . Progressives favour the alveolar nasal the most with a factor weight of 0.72. Nominal gerunds favour alveolar nasals too but to a lower extent, whereas, nouns and pronouns favour the velar nasal and adjectives favour the velar nasal the most.

The constraint with the second strongest effect size is the preceding segment with preceding liquids favouring the alveolar nasal the most followed by velars while preceding apicals favour the velar nasal the most. The following segment, persistence and gender are not significant constraints among old and middle-aged AMs .

4.6.4 Results for parametric regression for (ING) among young AMs

The next step was to make sure that a model for young AMs is also a better fit to the data by following the same methodology implemented for the old and middle-aged AMs . Since the complex model for all age groups has this young AMs model nested in it, it was appropriate to check if the smaller model is significantly better than the full one by implementing a chi-square test. The results indicate that the smaller model is indeed significantly better than the full more complex one for young AMs[18].

---

[18] Chi-square = 1033.342, df = 7, p = 7.52e-219

Table 4.9: Summary of best mixed-effects model for (ING) among AMs: young. Application value = alveolar nasal				
Degrees of freedom			9	
Centered input probability			0.70	
Overall proportion of application value			0.71	
Log-likelihood			-219	
AIC			455	
R2 total			0.70	
Constraints	*Log odds*	*N*	*Proportion of application value*	*Centered factor weight*
Grammatical-category				
Progressive	1.47	233	85%	0.81
Nominal-gerunds	0.46	79	61%	0.61
Nouns-and-pronouns	-0.61	36	35%	0.35
Adjectives	-1.32	107	48%	0.21
				Range = 60
following segment				
Apicals	0.71	67	90%	0.67
Pauses	0.30	53	72%	0.57
Velars	0.19	13	77%	0.55
Other-consonants	-0.07	183	74%	0.48
vowels	-0.91	152	60%	0.29
				Range = 38
Speaker random	**Std = 0.69**			
Word random	**Std = 1.77**			
[gender, preceding segment, persistence]				

Non-significant constraints on (ING) are in square brackets at the end of the table.

The results from Table 4.9 illustrate that the grammatical category constraint has the strongest effect size on (ING) among young AMs. Progressives favour the alveolar nasal the most followed by nominal gerunds, while nouns and pronouns favour the velar nasal and adjectives favour the velar nasal the most.

The constraint with the second strongest effect is the following segment; apicals favour alveolar nasals the most followed by pauses whereas vowels favour the velar nasal the most. The preceding segment, persistence and gender are not significant constraints on (ING) among young AMs.

These results are interesting because it seems that young AMs are at a cross-point between NS of NZSED and old and middle-aged AMs. Although young AMs pattern like NS in respect of some constraints on (ING), specifically, the significant constraints and their rank ordering, they still display some NNS-specific norms that are not significant for NS of NZE.

For example, within the grammatical category constraint, NS did not have a canonical verbal-nominal split, but old and middle-aged AMs did. Table 4.9 shows that young AMs still seem to have the nominal-verbal split of the older AM peers at the same time as they have relinquished preceding segment (the other significant AM's constraint in Table 4.8) and have introduced a significant effect for the following segment, like NS have. It is worth noting that the constraint effect allocated to the grammatical category is very high (as shown by the range of 60).

Moreover, young AMs seem to be taking the grammatical category constraint further than NS of NZSED by introducing further levels. This might reflect what Schleef et al. (2011) refer to as supra-local constraints on variation when AMs draw on the constraints they have acquired from other varieties rather than the local variety. This is usually the result of formal education in English as well as passive language contact through the media and social media channels.

A point to consider here is that the parents of this subgroup of AMs insisted on getting extra English classes for their children as well as enrolling them at schools that provided ESOL classes. Consequently, these children might have been introduced to grammatical distinctions that their NS peers have not.

4.7 Summary: (ING) findings among NS of NZSED and AMs

In the second part of this chapter, I introduced the results of parametric and non-parametric regression analysis for the variable (ING). I started with the results for NS of New Zealand Spoken English Database then looked at AMs who spoke a version of New Zealand English. I showed that it is better to separate the analysis of the AMs into two age groups: old and middle-aged AMs, and young AMs. I also tried to introduce a comparison between the native speakers and the two non-native speaker groups in terms of the patterns of variation that are significant for them.

The results indicated that AMs seem to acquire the main significant constraints that condition variation on (ING). AMs, however, seem to exhibit "transformation under transfer" which I will return to in the next chapter. There are also differences between the old and middle-aged AMs and young AMs. Whereas young AMs seem to display grammatical constraints that are more similar to NS constraints on (ING) variability than the older AMs do, but young AMs also diverge from the NS of NZSED. Table 4.10 illustrates the general patterns for the three groups, highlighting points of convergence.

Table 4.10: Summary of best mixed-effects models for (ING) among NS of NZSED, old and middle-aged AMs and young AMs. Application value: alveolar nasal		
NS of NZE	**Old and middle-aged AMs**	**Young AMs**
Grammatical category	*Grammatical category*	*Grammatical category*
Progressive	*Progressive*	*Progressive*
Nominal-gerunds	*Nominal-gerunds*	*Nominal-gerunds*
Participles	*Nouns-and-pronouns*	*Nouns-and-pronouns*
Adjectives	*Adjectives*	*Adjectives*
Nouns		
Discourse-markers		
Pronouns		
Range = 59	*Range = 44*	***Range = 60***
Following segment	**Preceding segment**	**Following segment**
Apicals	Liquids	**Apicals**
Velars	Velars	Pause
Other-consonants	Other-consonants	Velars
Vowels	Apicals	Other-consonants
Pauses		Vowels
Range = 56	***Range = 29***	*Range = 38*
[age, gender, persistence, preceding segment]	**[sex, persistence, following segment]**	**[sex, persistence, preceding segment]**

Non-significant constraints on (ING) are in square brackets at the end of the table.
Bold and italicized values indicate points of similarities and/or difference across the three Groups of speakers.

4.8 Chapter summary

In this chapter, I have presented the results for running non-parametric and parametric regression in R and Rbrul for two variables; CSD and (ING). I first presented the results for the CSD variable for NS of NZSED, then I introduced the results for AMs, both old and middle-aged and younger groups. I then summarized the major findings. I suggested that AMs have a tendency towards "transformation under transfer" in the acquisition of constraints on variation for the variable CSD, with differences between the patterns amongst the two sub-groups of AMs.

In the next chapter, I will highlight implications of the major findings of this study for both NS and AM groups. I will also provide an updated account for CSD and (ING) among NS of NZSED as a representation of NZE in the early 2000s. I then provide argumentation about the patterns of "transformation under transfer" as a major tendency for AMs as well as for other NNS groups in similar contact settings.

Chapter 5: Discussion

This chapter will discuss the findings reported earlier for coronal stop deletion (CSD) and the alternation between coronal and velar realisation of (ING) within the framework of the questions that this dissertation investigates. The first part of this chapter discusses the linguistic and the social constraints that condition variation on CSD among NS of NZSED, illustrating similarities and differences between the NZSED norms and other NS English varieties and the implications of these results.

The second part of the chapter discusses the linguistic and the social constraints found among AMs for CSD, illustrating major patterns of acquisition of variation as well as the implications of the findings for the understanding of the process of acquisition of variation in contact settings by first and 1.5 generations of AMs.

I then move on to discuss the results for the social and the linguistic constraints on (ING) among NS of NZSED and conclusions are again drawn to illustrate how far NZSED is similar/different from other NS English varieties. This is followed by the last part which introduces a discussion of the implications of the patterns of variation found among first and 1.5 generations of AMs. the variable (ING) is discussed, drawing on findings for CSD, to create a broader understanding of the acquisition of variation. I try to highlight how different contact settings, as well as NNS-specific social constraints, affect the differential acquisition of variation.

5.1 Discussion of the patterns of variation for the variable CSD among NS of NZSED

This part of chapter five discusses the findings reported in chapter four along the lines of the following questions:

1. Do NS of NZSED share the same constraints on CSD with other varieties of NS English, i.e., phonological constraints and grammatical category constraints and/or morphological status?

2. If the answer is yes, do NS of NZSED display constraint rankings like those found in other NS varieties of English?

3. Do the significant constraints found in NZSED display similar internal hierarchies as are found in other NS varieties of English?

4. What are the constraints that are non-significant in NZSED, and are these the same as constraints found to be non-significant in other NS varieties of English?

5.1.1 The three lines of evidence outlined for CSD among NS of NZSED

The results for CSD for speakers of NZSED illustrate that when speaker and word are modelled as random effects, the constraints that condition variation on CSD are the same constraints found to be significant in other varieties of NS English.

The rate of deletion (frequency of application value) reported for NZSED speakers, 32%, is within the same range of deletion rate found in other NS English varieties; Porirua corpus 39% (Holmes et al. , 1991) and Northern American dialects 33% (Guy, 1980). This rate is, nevertheless, less than rates of deletion reported for ethnolects like AAVE which has a deletion rate up to 44% (Poplack & Tagliamonte, 2001), Hispanic-English with a deletion rate of 50% (Santa Ana , 1996) and English-based creoles like Jamaican creole 70 % (Patrick, 1991) and Caribbean creoles 50-59 % (Poplack & Tagliamonte, 2001).

NZSED speakers, however, diverge from other NS English speakers in terms of (i) the strength of effects of the significant constraints and (ii) the internal hierarchies within these constraints. Specifically, the following segment constraint has the highest effect size on CSD in NZSED with a range of 40, followed by the preceding segment with a very close range of 38. The grammatical category constraints and age are last with a range of 28 each.

These two points are very important because I propose that the strength of the effect of the significant constraints do not only reflect dialect-specific aspects of variation among constraints, that are significant cross-dialectally, but also that these values may reflect a change in progress.

In the case of NZE, a change in progress may display itself as a divergence from the norms of variation transformed from more dominant dialects like British English and Australian English. Moreover, the effect size of significant constraints may reveal the role they have versus being the result of data-related asymmetries, especially in the case of CSD where some dialects have no grammatical constraints and CSD is mainly phonologically driven.

The second point, regarding the internal hierarchy of significant constraints on CSD, is very important to establish for two reasons. First, the internal hierarchies are assumed to reflect the underlying grammar; this makes it easier for us to determine what direction contemporary NZE is taking away from or towards other dialects. This is also crucial for understanding the acquisition of variation by AMs in New Zealand, because we need a clear understanding of the assumed target variety instead of assuming, a priori, that NZE illustrates the same constraints on CSD as other NS English varieties.

A point to mention here is that I sometimes report on values from regression analyses based on more elaborate coding for distinctions across levels within the internal hierarchy of a significant constraint. This is to illustrate acute differences in the behaviour of single levels.

These results are, nevertheless, the same as the ones reported in chapter four where hierarchal levels were regrouped, when appropriate.

5.1.1.1 The following segment constraint on CSD among NS of NZSED

The following segment constraint in NZSED displays a consonant-vowel continuum where non-sonorant consonants[19] favour deletion of the final consonant stops, whereas sonorant consonants, i.e., nasals and glides, have relatively neutral effects on CSD. Following vowels and pauses favour retention the most.

Specifically, non-sonorant stops favour deletion the most with a factor weight of 0.69, non-sibilant fricatives also favour deletion but with a reduced factor weight of 0.58. Sonorant nasals and glides slightly favour retention. Whereas vowels, the most sonorant, favour retention followed by pauses that favour retention the most.

The role of following pauses, as suggested earlier in the results chapter, is dialect-specific for NS of English; following pauses and following vowels have different factor weights cross-dialectally, but they are always the least favouring of deletion.

Another point to raise here is regarding the role of the following glides. Following glides behave differently from following vowels, as they are separate phoneme sets among speakers of NZSED. This fact will be informative when the AM results are discussed; the way AMs treat vowels and glides, in the target language, will inform our understanding of the overall process of acquisition of variation.

Specifically, if vowels and glides are treated as two separate sets of phonemes, we would assume that the AMs have already acquired the phonetic differences in the target language and are capable of targeting variation in the L2, since linguistic competence is a

---

[19] Sonority is defined as the amplitude of a sound or its resonance, in relation to other sounds. A sonority scale positions the sonority of a sound in relation to other sounds.

prerequisite for sociolinguistic competence. If AMs treat vowels and glides alike, this might suggest that AMs are still in the process of acquiring the phonetic inventory of the L2, which is a very complex process given their L1 has a completely different vowel system and a range of possible glides, and that their underlying grammars are not developed enough to start targeting L2 variation.

This discussion cannot be complete, however, without highlighting NNS-specific social constraints that may hinder or support the development of AM's-specific grammar, like networks, attitude towards the host country and its variety of English and proficiency in the L2.

One anomaly presents itself for NZSED speakers within the following segment constraint; sibilant-fricatives favour retention at a higher rate than nasals and liquids. But what we already know from other studies of CSD among NS of English is that sibilant-fricatives are less sonorant than following nasals and liquids, and we would, therefore, assume that sibilant fricatives would favour deletion at a higher rate than following nasals and liquids. I, consequently, tried to understand this anomaly by cross-tabulating the response (deleted /t, d/ versus retained /t, d/) with grammatical category when the following segment is a sibilant fricative.

The resulting table 5.1 illustrates that the context of the cross-tabulation returned the tokens with the following grammatical categories: adjectives, irregular past and regular past tense.

```
┌─────────────────────────────────────────────────────────────┐
│ Table 5.1: Cross-tabulation of the response (t, d/0) with grammatical │
│ category when the following segment is a sibilant fricative.  │
│ Native speakers of NZSED. The application value is deleted/t, d/ │
├─────────────────────────────────────────────────────────────┤
│ Following segment = Sibilant fricative                        │
│                                                               │
│           Grammatical category                                │
│ /t, d/-0   adj IrrP  PNoun present RP total                   │
│ 0          7   0     16    2       1   27                     │
│ TD         18  2     14    8       11  53                     │
│ total      25  2     30    10      12  80                     │
└─────────────────────────────────────────────────────────────┘
```

This result is expected from a functional semantic prescriptive (Nichols, 1984, Hengeveld et al. 2010); tokens representing adjectives, irregular past and regular past, carry meaning that would otherwise be jeopardised if the last consonant was deleted. Another piece of information drawn from the cross-tabulation concerning the grammatical category of nouns supports my explanation. Based in functional semantics, monomorphemic nouns delete at a higher rate than bimorphemic nouns, probably because the information contributed by affixation will be jeopardised if final stops are deleted, and it is therefore retained.

Holmes and Bell (1994) investigated CSD in the Porirua corpus (PC) which is also known as the Wellington Social Dialect Survey (Holmes et al., 1991). Their results presented similar findings to this sample of NZSED; the following segment constraints was found to have the strongest effect on CSD with a range of 47. And the internal hierarchy of the following segment constraint from most favouring of deletion to least favouring of deletion was reported as follows: obstruents > liquids > glides > vowels > pauses. These results illustrated that non-sonorant consonants favour deletion the most, whereas following vowels favoured deletion only 26% of the time (Holmes & Bell, 1994).

The results of the current project are also in line with results from studies conducted on CSD in other NS varieties of English. The following segment constraint was found to have

the strongest effect on CSD in the United States (Fasold, 1972; Labov, 1975; Guy, 1980, Neu, 1980; Labov 1989; Guy, 1991; Guy & Boberg, 1997; Roberts, 1997), Canada

(Hoffman & Walker, 2010)[20], and Britain (Tagliamonte & Temple, 2005; Smith et al., 2009).

The internal hierarchy for this constraint, as suggested by my results, also seems to follow other varieties of NS English where consonants favour deletion the most. Vowels and pauses cluster, and favour retention the most.

Many explanations have been suggested to account for the hierarchy of the following segment constraint. For example, the concept of linguistic universals has been deployed to explain the tendency of final stops in a stop to delete before following sounds. One operationalisation of universal constraints was to think in terms of unmarked canonical syllable structures (CVC).

Another account that tackles the concept of universal constraints was suggested by Guy and Boberg (1997) in terms of the obligatory contour principle (OCP), where adjacent segments are not allowed to be alike. This explains the difference between the following consonant and following vowels with the later favouring retention; when a consonant is followed by another consonant with which it shares many features, deletion is preferred, whereas consonants and vowels do not share features, so retention is favoured.

Another interesting account derived from universal constraints is introduced within the framework of Optimality theory (OT). The concept of "licensing by cue" to phonological neutralization was suggested as a framework to explain stop reduction. The assumption is that

---

[20] The results reported here are for first-generation Irish/British migrants to Canada (Hoffman and Walker 2010):
  1. Following segment. Range = 48.
  2. Morphological status. Range = 45.
  3. Preceding segment. Range = 12.

final stops are perceived at varying degrees based on acoustic cues; these are formant transition and release of the stops in the case of CSD. These acoustic cues form linguistic contexts, and the context that forms the most recognizable clusters licenses the release of that last consonant, hence the term, licensing by cue (Wheelerm, 2005).

For example, when a final stop is followed by another consonant, none of the linguistic cues is achieved, and therefore deletion is preferred. But when a stop is followed by a vowel, both formant transition and stop release are well-perceived, and therefore retention is favoured.

Another possibility is the re-syllabification of final consonant stops with adjacent sounds (Guy, 1980); if the final stop could be regrouped with the following segment to create an acceptable syllable in English, the consonant is retained, if not the final stop is deleted. (Labov, 1997) experimented with the concept of re-syllabification, as well, but he eventually suggested that a sonority-based account will better inform our understanding of the role of the following segment as a constraint conditioning variation on CSD, at least in Philadelphian English.

A second look at the internal hierarchy for the following segment constraint among speakers of NZSED presented in table 5.2 illustrates that the sonority hierarchy seems to best explain the internal hierarchy of the following segment in NZSED; least sonorant sounds favour deletion the most and more sonorant sounds favour retention.

```
Table 5.2: Extract from the regression analysis for CSD among NS of
NZSED. The following segment constraint. The application value is deleted
/t,d/.
```

Following segment	log odds	tokens	0/0+TD	centred factor weight
Stops	0.642	73	0.521	0.655
Non-sibilant fricatives	0.144	101	0.415	0.536
Glides	-0.257	43	0.372	0.436
Nasals	-0.340	75	0.293	0.416
Sibilant fricatives	-0.413	80	0.338	0.398
Vowels	-1.311	356	0.275	0.212
Pauses	-1.349	133	0.180	0.206

The extract is from a regression of more detailed levels (internal hierarchy) than the regression reported in the results where levels were regrouped.

5.1.1.2 The preceding segment constraint on CSD among NS of NZSED

The preceding segment constraint has the second strongest effect on CSD in NZSED with a range of 38, not much lower than that of the following segment constraint. The results for the preceding segment constraint diverge from previous findings for NZE, in terms of strength of effect and the internal hierarchy; the preceding segment constraint was reported to have minimal effects on CSD compared with the following segment and the grammatical category in NZE (Bell, 1977; Holmes & Bell, 1994; Schreier, 2005)[21].

Nevertheless, the results for NZSED regarding the strong effect of the preceding segment constraint may provide support to the proposition that CSD may be mainly conditioned by phonological constraints and that morphological constraints have no genuine effects.

---

[21] Bell's data reported on findings from news reading which provided a very formal register that is probably not to be compared with naturalistic data.

The results for NZSED also suggest that the internal hierarchy for the preceding segment has a pattern moving in the other direction to that reported by Bell (1977) and Holmes and Bell (1994) for NZE. They reported that fricatives, both sonorant and non-sonorant, favoured deletion the most, followed by stops that had a neutral effect on CSD, then nasals and liquids favoured deletion the least.

The internal hierarchy for preceding segment among NZSED presented by table 5.3, from most favouring of deletion to least favouring, suggests the exact opposite pattern from that reported above; that nasals favour deletion the most with a factor weight of 0.75 although words like *and* and negated forms like *don't* were deleted to avoid skewing the data. Liquids also favour deletion with a p-value of 0.55, while stops and sibilant fricatives have similar effects on CSD, both slightly disfavouring deletion and non-sibilant fricatives disfavour deletion the most with a p-value of 0.27.

There is also a gap in values of factor weight allocated to nasals versus other consonants, and there is also an interesting cut-off point for less-sonorant sounds; non-sibilant fricatives favour retention the most breaking from other fricatives (nasals > liquids > stops > sibilant-fricatives > non-sibilant fricatives) reported earlier for NZE (Holmes & Bell, 1994).

Table 5.3: Extract from the regression analysis for CSD among NS of NZSED. The Preceding segment constraint. The application value is deleted clusters of CSD.

Preceding segment				
factor	log odds	tokens	0/0+TD	centred factor weight
nasal	1.214	393	0.468	0.771
/l/	0.199	96	0.188	0.55
stop	-0.185	104	0.173	0.454
Sibilant -fricatives	-0.238	207	0.217	0.441
Non-sibilant fricatives	*-0.989*	*66*	*0.091*	*0.271*

The extract is from a regression of more detailed levels (internal hierarchy) than the regression reported in the results where levels were regrouped.

Although my results contradicted Holmes and Bell (1994) results, a more recent study of CSD in NZE supports my findings. It suggested that for preceding environments, Pakeha New Zealand English adheres to the patterns familiar in U.S. varieties, in that laterals and nasals favour deletion more than most fricatives and plosives (Schreier, 2005). This suggests two possible interpretations: either that NZESD represents a variety of NZE that is different from the one reported for the Porirua corpus, or that NZSED patterns represent a change in the constraints on CSD. The second rationale is plausible since the NZSED reported on represents Pakeha speech in Wellington in the 21st century, a decade after the Porirua data was collected. It could also be that the inclusion of Maori speakers in the Porirua corpus have camouflaged our understanding of white-NZE. Either way, the results are intriguing for such a stable variable like CSD.

Many explanatory accounts were presented for the role of the preceding segment constraint. One account was based on the concept of a sonority hierarchy, for example, data from York English suggested that less sonorous segments tend to favour deletion (Tagliamonte & Temple, 2005). This account is supported by (Patrick, 1991), who proposed the following sonority hierarchy for the preceding segment from most favouring to least favouring of deletion: Stops > Fricatives > Sibilants > Nasals > / l /. This explanation obviously does not explain the current project's findings.

Guy and Boberg (1997) tried the concept of the obligatory contour principle (OCP) as a framework for the role of the preceding segment constraint. We can accordingly postulate that the more distinctive features the final stop shares within the stop, the more likely deletion is to occur. The distinctive features defining /t/ and /d/ are [+coronal, -sonorant, -continuant]. Therefore, Guy and Boberg (1997) originally presented the hierarchy, from most favouring to

least favouring of deletion as follows: Sibilants > Stops > Nasals > Non-sibilant-fricatives and Laterals, but this account also fails to explain the results of my data.

Labov (1989) summarized the preceding segment constraint from most favouring to least favouring of deletion: sibilants > stops > nasals > fricatives > / l /. The hypothesis proposed is that more sonorous preceding segments are more likely to incur deletion. Nevertheless, this suggestion fails to explain why stops favour deletion more than nasals.

Santa Ana (1996) also suggested that the more sonorous the preceding segment within the stop the more likely deletion is to occur. I believe that a sonority-based account from most to least sonorant, may explain the internal hierarchy found for NZSED speakers,The fact that there are many contradicting accounts that explains the role of the preceding segment constraint on CSD across many NS English varieties suggests that the ordering of internal hierarchy for this constraint is probably dialect-specific. Not only do I propose that the internal hierarchy of the preceding segment constraint is dialect-specific, but I also suggest that the strength of the effect of the constraint is dialect-specific as well.

This piece of information is very relevant when I discuss the results for AMs to judge whether they are sensitive to dialect-specific constraints on variation. Evidence to support this proposition can be drawn from varieties of English that illustrate differential levels of strength of effects for the preceding segment constraint. For example, Tagliamonte and Temple (2005) investigated one variety of English in the UK, York English, to find that the preceding segment has a stronger effect on CSD than grammatical category.

In another study of CSD in a small town in Scotland, the preceding environment is found to be as strong a constraint as the other two classical factors as suggested by the close range-values allocated to the constraints on CSD. Smith et al. (2009) also demonstrated similar results for Buckie town in Scotland. Hispanic English in the United States (Santa Ana, 1996) has also illustrated a strong effect of the preceding segment constraint.

5.1.1.3 The grammatical category constraint on CSD among NS of NZSED

The grammatical category is also a statistically significant constraint for NS of NZSED with a range of 28, with the following constraints hierarchy: nouns and base form verbs > adjectives > irregular past > regular past. Nouns and base form verbs favour deletion the most with a factor weight of 63%.

A note to mention here is that a regression analysis conducted for a more detailed coding of grammatical categories illustrated that adverbs favour deletion the most probably because other studies did not include a fine distribution of grammatical categories and therefore, the grammatical category *Nouns* would usually be reported as the most favouring of deletion. Or because the grammatical category, *adverbs* can be re-analyzed as adverbial prepositions which are usually excluded from analysis.

My results support the notion that nouns favour deletion, but slightly less than adverbs. Adjectives have a neutral effect on CSD and irregular past and regular past tenses favour retention the most with a slight difference in factor weights.

Diverging from my findings, Holmes and Bell (1994) reporting on the Porirua data suggested that the grammatical category constraint has a stronger effect on CSD than the preceding segment. It could be that they did not consider speaker and word as random effects, or it could be that they did not conduct a cross-tabulation among grammatical category and the following segment constraints and therefore did not report any potential for the grammatical constraints to be an artefact of data and have no genuine effects.

The internal hierarchy reported in my study is, nevertheless, the same as that reported by them. Monomorphemic nouns favour deletion at a higher rate than monomorphemic irregular past tense and bimorphemic past tense. Holmes and Bell (1994:69) reported that

"these two sets of past tense morphemes pattern almost identically in disfavouring CCR [stop reduction]". That past tense forms, both regular and irregular, disfavour deletion with a slight difference in factor weights was also established for Pan-American dialects (Guy, 1991).

The reduced strength of the effect of the grammatical category constraint reported for NZSED could support the proposition that the role of the grammatical constraint may be the result of unevenly distributed data. It is worth mentioning, however, that when reporting a role for the grammatical category, Functional semantics is sometimes used as a framework to explain it. Nominal tokens favour deletion because the loss of a final sound in a cluster would not distort meaning, whereas deleting final sounds in verbs may distort meaning because they distort the aspect of tense marking (Guy, 1980; Neu, 1980).

Another explanation was suggested by Guy (1991) within the framework of lexical phonology. Guy suggested that the addition of morphemes to roots is cyclical in nature, meaning that the addition of affixes is achieved at several levels of phonological and morphological building. Stops that occur at the first level of phonological word-building are deleted at an exponentially higher rate than those achieved by applying multiple morphological and phonological processes.

This explanation may be plausible when discussing the acquisition of past tense marking by NS children or second language learners, as the marking of tense (or its absence) represents their developing grammars. But I do not find this framework to clearly justify stop reduction by adult NS, especially since semi-weak verbs in the NZSED are deleted at a rate very close the deletion rate of regular past tense forms. This means that the double marking of the irregular verbs, first by altering the vowel and then adding a past tense marker, does not make irregular past tense more prone to deletion than regular past tenses.

It is possible that CSD among NS of English is primarily phonologically conditioned, this explains the strong effects of the following segment constraint cross-dialectally. It also

explains the persistence of the preceding segment constraint with a strength of effect that is dialect-specific.

Some researchers even argue, that CSD is only phonologically conditioned and that the role of the grammatical constraint is merely an artefact of the uneven distribution of tokens across grammatical categories (see Temple, 2003; Tagliamonte & Temple, 2005; Hazen, 2011). Hazen (2011:105) suggested that:

> "This finding [for Appalachian English] is in stark contrast with some other vernacular varieties and suggests that apparent morphological influences are actually phonological influences that present themselves as morphological trends."

I have tried to experiment with the hypothesis that grammatical constraints on CSD may not exist and that this variable is mainly phonologically conditioned for NZSED. To do that, I created a cross-tabulation between the grammatical category and the following segment. As table 5.4 illustrates, all grammatical categories favour deletion when the following segment is a consonant, 64% of the time. Even past tense categories, that are usually found to be the least favouring of deletion, seem to favour deletion when the word is followed by a consonant. Irregular and regular past categories favour deletion at a rate of 63% and 73%, respectively.

```
Table 5.4: Cross-tabulation of the grammatical category with the
following segment. NS of NZSED. The application value is deleted /t,d/.
```

	Counts							
TD_0 = 0								
Following	**Grammatical category**							
Segment								
	adj	adv	IrrP	N	preposition	base-form	RP	total
Glide	4	3	2	12	0	1	0	22
Nasals	7	1	1	3	2	1	3	18
Non-sibilant fricatives	8	13	2	11	10	6	2	52
PAUSE	7	1	0	10	1	0	6	25
Sibilant-fricatives	7	1	0	16	0	2	1	27
STOP	12	3	2	11	2	6	4	40
VOWEL	***9***	***4***	***4***	***70***	***2***	***5***	***6***	***100***
total	54	26	11	133	17	21	22	284
f-cons	43	22	7	63	15	16	16	184
deletion%	79%	84%	63%	55%	88%	76%	73%	64%
TD_0 = TD								
Following	**Grammatical category**							
Segment								
Glide	17	1	3	12	0	6	14	53
Nasals	13	1	2	7	1	1	3	28
Non-sibilant-fricatives	5	2	2	20	6	12	18	65
PAUSE	19	9	1	52	2	2	26	111
Sibilant-fricatives	18	0	2	14	0	8	11	53
STOP	13	2	3	7	1	3	7	36
VOWEL	***33***	***24***	***13***	***99***	***11***	***21***	***68***	***269***
total	*118*	*39*	*26*	*211*	*21*	*53*	*147*	*615*
f-VOWEL	***33***	***24***	***13***	***99***	***11***	***21***	***68***	***269***
retention%	28%	62%	50%	47%	52%	40%	46%	44%

The extract is from a regression of more detailed levels (internal hierarchy) than the regression reported in the results where levels were regrouped.

I, nevertheless, cannot be entirely sure about the accuracy of suggesting that morphological conditioning is an artefact of unevenly distributed data among NZSED speakers because when I ran this constraint as a possible individual factor on CSD in R, it returned a p-value < .0001 (see Table 5.5 is an extract from the R coding), which R also returned for the following segment and the preceding segment constraints, both of which turned out significant in the R and Rbrul regression analyses.

Table 5.5: Extract from R code. NS of NZSED.CSD: following segment, preceding segment and grammatical category

```
## to check which factor should be kept in the model for NZSED

## pos: p<.0001 (GOOD, should include in model)
ns.csd.glm <- glmer(TD_0 ~ pos+(1|speaker), data = data, family =
"binomial")
summary(ns.csd.glm)
mixed (TD_0 ~ pos+(1|speaker), family=binomial, data=data, method =
"LRT")

## ps: p<0.0001 (GOOD, should include in model)
ns.csd.glm <- glmer (TD_0 ~ ps+(1|speaker), data = data, family =
"binomial")
summary(ns.csd.glm)
mixed (TD_0 ~ ps+(1|speaker), family=binomial, data=data, method =
"LRT")

## fs: p<0.0001 (GOOD, should include in model)
ns.csd.glm <- glmer (TD_0 ~ fs+(1|speaker), data = data, family =
"binomial")
summary(ns.csd.glm)
mixed (TD_0 ~ fs+(1|speaker), family=binomial, data=data, method =
"LRT")
```

(#) symbol is used in R, to insert comments and notes, that will not be read by R as a code to run.
pos **is a code for grammatical category constraint, fs is a code for the following segment constraint, and** *ps* **is a code for the preceding segment constraint**
we can fit models with each variable as an individual predictor. This variable is of interest if a p <0.05, which is the norm in sociolinguistics (standard was chosen significance level = 0.05 cut-off).If the p-value is less than the chosen significance level, the observed data is sufficiently inconsistent with the null hypothesis, and the null hypothesis can be rejected. When the p-value is calculated correctly, this test guarantees that the Type one error rate is at the chosen significance level. For typical analysis, using the null hypothesis is rejected when p<.05 and not rejected when p>.05.

5.1.1.4 Other constraints on CSD among NS of NZSED

I have attempted to experiment with potential persistence effects on CSD among NZSED speakers, and I considered different measures of persistence each of which has specific underlying assumptions.

The first measures of persistence, coded for, was adapted from Abramowicz (2007) who defined persistence within the framework of recency and shared morphological status of successive tokens; he found evidence for a tendency of the persistence of one variant across the same morphological class when the tokens are close enough.

In my study, I coded for this measure of persistence by creating two constraints (i) the morphological structure of words (number of morphemes) and (ii) the distance between consecutive tokens (whether the tokens were within the same breath unit or not).

I also considered persistence by operationalizing Tamminga's (2014) psycholinguistic framework of persistence. Her hypothesis, simply put, suggests that a simple persistence effect occurs within the same morphological class. I consequently created a constraint for the number of morphemes of the current token and another for the preceding one. I also added another constraint to represent the realization of the prior token; whether the final consonant was deleted or retained, since it seemed only logical to account for the previous realisation of CSD when trying to account for any role of persistence.

I found only a weak effect for persistence summarised in a mild effect that the realisation of the previous token illustrated. This result was only achieved when I ran logistic regression analysis in R. Although this result is non-significant , it was moving in the expected direction; if the previous word has a deleted consonant, the current one would most likely have a deleted one as well. Moreover, the mere existence of a priming effect for a

variable that is not discourse-related is an interesting finding, especially when I try to understand the differences in the role of persistence between NS and AMs.

Age is the only social constraint that is significant for CSD among NZSED speakers. Young NS of NZSED favour deletion with a factor weight of 0.64, whereas both middle-aged and older speakers favour retention with a factor weight of 0.36. This result may indicate a generational effect whereby old and middle-aged NS cluster and use the informal variant (alveolar nasal) less than young NS. Such a generalization is to be taken with high caution since the coding for age cohorts in the NZSED may obscure what is really going on age-wise.

To elaborate, the stratification of age cohorts in the NZSED was data-led; 18-30 were coded as young, 31-45 were coded as middle-aged, and 46-60 were coded as old (Warren, 2002). I would assume, nevertheless, that 30-year olds would behave like 31-year olds, the same goes for 45-year olds who would behave like 46-year olds, thus camouflaging an age affect if any.

Style-shifting was found to be non-significant for CSD among speakers of NZSED. A point to mention, however, is that CSD does not generally have strong overt attitudes. Deletion is, nevertheless, favoured more in casual speech. Another related point is that the NZSED was based on conversations that did not introduce potential contexts for style-shifting to occur; most speakers knew each other and/or were colleagues and the only scenario, I found to be more formal, was when some speakers were not very comfortable in front of a researcher with an audio recorder.

The most reported social factors to usually influence CSD, i.e., gender and style-shifting, were found to be non-significant for this sample. The reason why two of the most reported social constraints have been found to be non-significant for this sample of NZSED speakers, might be due to the use of mixed-effects modelling, where the intra-speaker

variation is accounted for in order to avoid creating Type I statistical errors that may result in reporting social effects that do not necessarily exist.

5.2 Summary of the discussion of results for CSD among NS of NZSED

I discussed the results for speakers of the NZSED for the variable CSD in section 5.1 of this chapter. I now present a summary of these results alongside the questions that I have raised for this sample of the project.

Speakers of NZSED do illustrate the same constraints that are found to be significant by other NS of English; preceding and following segment constraints and grammatical category. They, however, diverge from other varieties of English in the rank ordering and the internal hierarchies of the constraints conditioning CSD.

The following segment constraint has the strongest effect on CSD in all NS English varieties that I have considered for this project. The internal hierarchy for NZSED is also compatible with other varieties of English where following consonants favour deletion and following vowels and/or pauses favour retention.

The role of the following pause is dialect-specific (Guy, 1980), but it is not necessarily a regionally-motivated dialectal difference. For example, NZSED has an internal hierarchy for the following segment that is like Buckie-town Scottish dialect (Smith et al., 2009), York English (Tagliamonte & Temple, 2005) and AAVE in Philadelphia (Cohen & Labov, 1967; Fasold, 1972).

The effect size of the preceding segment constraint and its internal hierarchy seem to be dialect-specific as well. NZSED display a strong effect size of the preceding segment with an internal hierarchy compatible with the sonority hierarchy as presented by (Labov, 1997).

Morphological conditioning is most likely to be a result of the uneven distribution of data across grammatical categories. This places NZSED amongst York dialect (Tagliamonte

& Temple, 2005) and Appalachian English (Hazen, 2011). NZSED also diverges from other varieties of NS English in that some social constraints including gender and style-shifting are non-significant in NZSED probably because I used mixed modelling.

In the next section of this chapter, I discuss the results for the linguistic and the social constraints that condition variation on CSD by AMs.

5.3 Discussion of the patterns of variation for the variable CSD among old and middle-aged AMs

As mentioned in chapter 4, I started with a AM model that included all age groups: old, middle-aged and young because I started with the assumption that membership in an AM community of practice would entail membership in a second language speaker community of practice. This assumption turned out to be inaccurate though, as quantitative data (together with information about the social factors of the AMs) indicated that old and middle-aged AMs belong to one second language community of practice and young AMs belong to another one.

Specifically, old and middle-aged AMs had experiences in English and network patterns that set them apart from young AMs . Old and middle-aged AMs received direct instruction to English in New Zealand as a preliterate group and I was personally involved with one of the NGOs responsible for linguistically empowering newcomers. Young AMs , on the other hand, although almost NS-like, have received special attention at schools by hving to attend special ESOL classes. Their mothers have also informed me that they took their children to Sunday schools in English which turned out to be play dates with NS of NZE whose mothers volunteered to help the newcomers.

This resulted in different experiences in Engish for the two AM groups, nevertheless, both groups experienced some direct prescriptive instruction to English that I discuss shortly

as the potential source for the groups' tendency to apply supra-local constraints on the grammatical category on (ING).

The two groups also have different network patterns; young AMs are fully integrated in the NZ community but they still hold very tight connections with their culture and ethnic identity as Arabs. The young AMs' friends are shared, they also went to the same kindergardens, schools, colleges and eventually universities.

The data from the regression analysis for the AM groups also supported this qualitative observation. The initial model that included all age groups of AMs had many convergence warnings that compelled me to conduct a cross-tabulation for the social factors of the AMs and I found that there are some severe interactions and knockouts. I, therefore, tried different possible groupngs of the AMs that complied with my observations and the qualitative data I achieved and eventually decided that it is definitely better to consider that AM as two separate second language communities of practice.

The findings from the detailed coding (see section 3.3.2 page 80-81) helped me to infer what the AMs' inventories are, based on the constraints on the variation they demonstrate.that old and middle-aged AMs do indeed have different L2 phonetic inventories than young AMs which also proves that they belong to different second language communities of practice. Moreover, the resuts from the non-parametric test provided in sections (4.2.1, 4.3.1, 4.3.2. for CSD and sections 4.6.1 for (ING)) have also supported this decision to split the AM's group.

Once I created separate models the constraint networks, had to be taken out of the model as there are extreme interactions between age and networks; old speakers had mixed networks, middle-aged speakers had non-Arab networks, and younger ones had non-Arab networks.

```
Table 5.6: Cross-tabulation of age with networks among AMs. The
application value is deleted /t,d/
                              Counts
           Age
Networks  middle-aged  old     young total
Arab           0       569       0     569
Non-Arab      834      403     1001   2238
total         834      972     1001   2807
```

5.3.1 The three lines of evidence outlined for CSD among old and middle-aged AMs

In this section, I discuss the results for old and middle-aged AMs within the framework of the questions that I have raised for this sub-group of AMs. Specifically:

1. Do old and middle-aged AMs illustrate the same constraints on CSD that are found in the NZSED?

2. If the answer is yes, do old and middle-aged AMs display rank ordering of constraints like those found in NZSED for the variable CSD?

3. Do the constraints found significant for old and middle-aged AMs have similar internal constraint hierarchies as those found in NZSED for the variable CSD?

4. What are the constraints that are non-significant for AMs? And are there any AM-specific constraints?

5. Are there any implications of "transformation under transfer" (Meyerhoff, 2009a; Meyerhoff & Schleef, 2012)? If the answer is yes, are there any signs of weak transfer, strong transfer and calquing?

6. Are the patterns of variation for CSD among old and middle-aged AMs found among other first-generation NNS of English who live in similar contact settings?

Chapter 4 illustrated that the acquisition of constraints on linguistic variation for CSD is systematic and stable for old and middle-aged AMs, in terms of significant constraints, rank ordering and internal hierarchies. Moreover, the rate of CSD, reported for old and middle-aged AMs of 28%, is also close to the deletion rate found in NZSED, 32%. Old and middle-aged AMs, nevertheless, do not seem to have reached target-like use of CSD patterns in terms of the strength of the effect of the two significant phonological constraints and their internal hierarchies.

The strength of effect of the significant constraints among old and middle-aged AMs is reduced compared with strength of effects displayed for the same constraints among speakers of NZSED; NZSED had a range of 40 for the following segment constraint followed by a range of 38 for the preceding segment constraint while old and middle-aged AMs had a range of 27 for the following segment constraint and a range of 19 for the preceding segment constraint.

Old and middle-aged AMs also diverge from NS norms for CSD because they do not have the grammatical category among the significant constraints on CSD, probably, as a result of three possible scenarios. A point to mention here is that the suggested reasons, for the inquisition of grammatical constraints on CSD by this group of AMs, may be working in tandem and the truth of one suggested scenario does not dispute the possible role of the others.

The first is that NS norms for CSD might be phonologically conditioned and any apparent grammatical constraints reflect data asymmetries typical of sociolinguistic data. Consequently, there is no grammatical category constraint for AMs to acquire. In this case, NZSED will be like other NS varieties of English like York English (Temple, 2003; Tagliamonte & Temple, 2005) and Appalachian English (Hazen, 2011).

The second scenario is based in that first-generation AMs have developing linguistic competence. If we assume that grammatical category is a significant constraint on CSD among NS of NZSED, then the lack of a grammatical category constraint may be a result of the developing acquisition of the L2 grammar by old and middle-aged AMs. And since the acquisition of linguistic competence is a prerequisite for the acquisition of sociolinguistic competence, defined as the ability to target NS-variation patterns, old AMs and middle-aged AMs do not target L2 variation.

The lack of a significant grammatical category for old and middle-aged AMs may also reflect how first-generation AMs perceive the importance of constraints in the L2 and how they reinterpret them. For example, (Wolfram, 1984) suggested that NNS seem to target the acquisition of the L2 grammar when they first arrive into a host country; linguistic competence. They, nevertheless, shift their attention to local norms after spending a temporal threshold in the host country, Wolfram suggests it is three years after their arrival, provided that they have social networks with the NS of the L2. Only then, NNS starts to target sociolinguistic competence including the acquisition of variation.

Specifically, Wolfram (1985) considered the intersection between phonological and grammatical processes for marking regular past tense forms by learners of English and found that past tense forms were usually marked when in a cluster followed by a consonant, especially for those with length of residence (LOR) of 4–7 years and longer LOR of 4–7 years deleted in monomorphemes more often than in regular past tense forms, like patterns found among NS of English. In contrast, the group with the shorter LOR of 1-3 years were more likely to delete in past tense forms than in monomorphemes. This led Wolfram to state:

> "While the role of the phonological process cannot be ignored at any stage
>
> of acquisition, it appears that it takes on an increased role vis-à-vis the
>
> grammatical process as overall proficiency increases. In the incipient stages,

the processes may converge, with the grammatical process taking

precedence, while in the latter stages the phonological process may take on

increased significance." (p. 235).

Relatedly, Schleef (2017a), Drummond (2011) and Sharma and Sankaran (2011)

illustrated that migrants who have spent two to three years in a host country become more

likely to acquire stylistic variants and their associated social indices.

The NNS-specific constraints for this subgroup of AM can provide further support to

the last proposition by considering the cross-tabulation of social constraints of the time spent

in New Zealand and the social network patterns (maintaining Arab only networks as opposed

to mixed networks) as presented in table 5.7.

```
Table 5.7: Cross-tabulation of age, network and time spent in NZ
old and middle-aged AMs. The application value is deleted /t,d/
Counts
timing = less than three years
          age
networks  mid-aged  old  young total
Arab         0        0     0     0
Non-Arab   702        0     0    702
total      702        0     0    702


time_NZ = more than three years
          age
networks  miaged   old   young total
Arab         0     550      0    550
Non-Arab   128     395    1001   1524
total      128     945    1001   2074
```

I propose that old AMs' acquisition of the L2 linguistic competence seems to have been frozen because of their network involvement and lifestyle. Specifically, migrants to New Zealand are required to enrol in TESOL courses to help them become qualified for work and life in the host country and they lose all their benefits from the government if they do not attend these courses.

I have been personally involved in teaching English for AMs in Wellington, and I observed that many people, especially old AMs are usually present at these courses because they are afraid to lose their governmental benefits. Once they get a job, they usually stop attending their classes, because they have a regular source of income and do not mind losing their benefits.

All older AM women, in my sample, are stay-at-home-mums, whose networks are mainly Arab and therefore they do not need to speak English at all. The older AM men also have very restricted networks, and this might have made their linguistic development freeze at some point, for example, after they finished the obligatory English courses, or got jobs, they stopped attending any more TESOL classes.

Old AMs seem to have a restricted acquisition of the L2 where their linguistic competence has been frozen. Accordingly, they never go beyond linguistic competence and are, therefore unable to achieve sociolinguistic competence that is usually associated with the acquisition of variation in the L2.

Middle-aged AMs, on the other hand, might go beyond the three-year threshold for the acquisition of linguistic competence and it would be interesting if I get to interview them again after some time (two to three years) to see how far their competencies develop as they have different lifestyles and network involvement from old people who live and work within closed Arab networks.

5.3.1.1 The following segment constraint on CSD among old and middle-aged AMs

Old and middle-aged AMs internal hierarchy for the following segment constraint seem to illustrate a continuum of sounds where consonants favour deletion and vowels and pauses favour retention. This internal hierarchy is very similar to that found in NZSED; obstruent sounds favour deletion, glides have a neutral effect on CSD while vowels and pauses favour retention.

Old and middle-age AMs diverge from the NS internal hierarchy in two respects: the first is in terms of the rank order of the NS levels, as AM levels are assigned different factor weights compared with NS internal hierarchy. This factor weight difference may reflect L1 transfer in the old and middle-aged AMs' patterns of variation.

For example, following glides behave differently from following vowels among speakers of NZSED as they are two different phonemes in NZSED. Whereas old and middle-aged AMs seem to treat the two sets of phonemes alike, favouring retention. This suggests that old and middle-aged AMs are still in the process of acquiring the phonetic inventory of the L2, which is a very complex process especially given that their L1, Arabic, has a different vowel system and a range of possible glides.

This also suggests that their underlying grammars are not developed enough to start targeting L2 variation. Had they illustrated the distinctions between the two morpheme sets, we would assume that the old and middle-aged AMs have already acquired the phonetic differences in the target language and are capable of targeting variation in the L2, but this is not the case.

Middle-aged AM have networks with NS. They, nevertheless, have been in New Zealand for less than three years except for one male respondent who owns an Arab shop and maintains mixed networks, and therefore he behaves more like old AM.

Old and middle-aged AMs do not display NS-like variation patterns because old AM lack the L2 exposure needed to achieve sufficient competence in the L2, and middle-aged AM are more likely to target the linguistic competence in the L2, and to start acquiring local constraints after spending three years in New Zealand.

The second point of divergence from NS norms is the role of the following nasals that have a neutral effect on CSD among NS of NZSED, whereas following nasals are the most favouring of deletion among old and middle-aged AMs.

Relatedly, an interesting aspect of old and middle-aged group's internal hierarchy; nasals > obstruents > glides > vowels and pauses, is that it cannot be explained by a sonority-based account that well-explains NZSED patterns. Labov (1997) suggested that a sonority account will better inform our understanding of the role of the following segment as a constraint on CSD, where least sonorant sounds favour deletion the most and more sonorant ones favour retention. This account does not, however, justify why obstruents favour deletion more than glides and vowels among old and middle-aged AMs.

Another account to explain the role of the following segment constraint was suggested by Guy and Boberg (1997) utilising the concept of the obligatory contour principle (OCP); from most favouring of deletion to least favouring: Sibilants > Stops > Nasals > Non-sibilant-fricatives and Laterals. This account also fails to explain the hierarchy found for old and middle-aged AMs: nasals > obstruents > glides > vowels and pauses.

Since the traditional accounts fail to explain the internal hierarchy of the following segment constraint among old and middle-aged AMs, another account that considers NNS-specific features for this AM group, like proficiency in the L2 and their L1, may better inform our understanding of the acquisition of variation among them.

Old and middle-aged AMs have developing grammars. Therefore, the internal hierarchy of the following segment constraint for them might be a result of the re-

syllabification of final stops with following sounds in accordance with Arabic allowable syllable structures. For example, Levantine Arabic allows /tn/, /dn/, /tm/, /dm/, /tb/, /db/, /tj/, /tw/, /dj/, /dw/ are plausible syllables in Arabic.

A final note is that old and middle-aged group's internal hierarchy seem to have acquired dialect-specific aspects of variation on CSD; they displayed a factor weight of the following pauses that mirrors that found in NZSED.

5.3.1.2 The preceding segment constraint on CSD among old and middle-aged AMs

Old and middle-aged AMs seem to have acquired the NS rank ordering and to have partially acquired the internal hierarchy of the preceding segment constraint; they seem to have acquired that preceding nasals favour retention more than other consonants. Old and middle-aged AMs, nevertheless, reverse the internal order of the remaining levels.

Unlike NZSED speakers, this AM sub-group illustrate that preceding fricatives favour deletion more that preceding liquids while stops maintain the same position found among NZSED. One way of understanding this divergence from NZSED norms is what has been established about old and middle-aged AMs; that they are still learning to differentiate the phonemes of the L2. Liquids are obviously a source of confusion to them because of the differences between Arabic and NZSED phonetic systems.

Although OCP (Guy & Boberg, 1997) is a plausible explanation for the preceding segment hierarchy among old and middle-aged AMs, the previous line of reasoning compels me to suggest an explanation for the preceding segment hierarchy as a re-syllabification process. Because it explains not only the data but also provides support for my assumptions about this group. For example, Arabic allows preceding obstruents to form syllables with /t,d/ resulting in clusters like /bt/, /bd/.

Although old and middle-aged AMs acquired the rank order for the preceding segment constraint, they seem to illustrate a reduced strength of effect for this constraint, as suggested by the low range value. A point to mention, however, is that this constraint might pose further challenges for old and middle-aged AMs because, as suggested earlier, the internal hierarchy and the strength of the effect of the preceding segment constraint are dialect-specific for NZSED, the assumed target group.

Old and middle-aged AMs, like their NZSED peers, found style-shifting, gender and persistence to be non-significant constraints on CSD.

5.4 Conclusions for CSD among old and middle-aged AMs

The results discussed above based on the comparison between NS of NZSED and old and middle-aged AMs patterns of variation on CSD reveal that AM illustrate a tendency of "transformation under transfer". Meyerhoff (2009a) identifies three main tendencies that of "transformation under transfer";

1. Weak transfer occurs when the same constraints are significant constraints on a variable in the model (source) and in the replica varieties.

2. Strong transfer occurs when the same constraints are significant in both model and replica, and the ordering of these constraints is the same in both model and replica.

3. Calquing occurs when the same constraints are significant in both model and replica, and the ordering of these constraints is the same in both model and replica, and the levels within the significant constraints have the same rankings in model and replica.

According to the definitions above, old and middle-aged AMs display instances of strong transfer of NS constraints on CSD, whereby they replicate the following segment and the preceding segment constraints maintaining NS rank orderings and very similar internal hierarchies. One could even argue for calquing for these phonological constraints because the AMs seem to have acquired the main NS distinctions between following consonants and vowels for the following segment constraint. They also acquired a very similar internal hierarchy for the preceding segment, separating preceding nasals from other consonants.

Nevertheless, old and middle-aged AMs do not replicate other distinctions of internal hierarchies for the phonological constraints probably because of their developing phonetic system in the L2.

The finding that first-generation AMs have illustrated strong transfer of NZSED constraints on CSD is unexpected, as we usually see that strong L1 transfer and low-proficiency in the L2 grammar usually make first-generation migrants less prone to acquire

NS-like constraints. But a reconsideration of the situation may yield new insights into the acquisition of variation by migrant groups.

It is possible that calquing and/or strong transfer occurs for this subgroup of AMs because phonological constraints have universal aspects to them rather than language-specific constraints. Another more plausible way to put this is that CSD is mainly conditioned by phonological or "articulatory" constraints (Labov, 1989). NNS may more readily acquire articulatory constraints since these constraints as suggested by results for child NS (Labov, 1989; Roberts, 1997), and other NNS of English (Bayley 1996) for the variable CSD, compared with stylistic and dialect-specific ones,

It could also be that old and middle-aged AMs perceive CSD not as an example of variation in the L2 but as a lenition process. Old AMs cannot perceive local constraints on variation in general because of the lack of exposure to NZE. It could also be that middle-aged AMs are also incapable of perceiving local norms of variation, at this point, because they have not yet passed the three-year threshold after which local norms start to be targeted.

5.5 Placing old and middle-aged AMs in the context of other first-generation migrants acquiring variation in contact settings

In this section, I try to place old and middle-aged AMs patterns of acquisition of variation amongst other studies of CSD in different contact settings. I try and compare patterns of variation found among first-generation NNS with my results for first-generation AMs, showcasing these tendencies for each constraint at a time. The general pattern that seems to be persistent is that first-generation NNS with different first languages have acquired the following segment and the preceding segment constraints on CSD to varying degrees, and some showed acquisition of grammatical constraints as well. First-generation NNS, regardless of their first language and their level of proficiency in the L2, seem to be attentive to the phonological constraints on variation.

Other first-generation NNS illustrated similar results to those found among old and middle-aged AMs, for the variable CSD, including the acquisition of NS rank order and internal hierarchy-structures for the following segment constraint illustrating an ability to acquire dialect-specific aspects of variation in an L2 (since we have already established that rank orderings and internal hierarchies may be dialect-specific across NS varieties).

For example, Chinese migrants learning English in the United States displayed the following segment constraint as significant with the strongest factor effect on CSD. The internal hierarchy for this constraint also resembles that of old and middle-aged AMs, whereby consonants favour deletion the most while pauses and vowels favour retention (Bayley, 1996).

Another example illustrates that two first-generation migrant groups to Canada (Chinese and Italian) displayed the following segment as significant with the strongest effect on CSD. Consonants favoured deletion while vowels favoured retention with pauses favouring retention at a lower rate than vowels (Hoffman & Walker, 2010). The role of the

following pauses is dialect-specific (Guy 1980), and we can conclude, accordingly, that these first-generation NNS are attentive to dialect-specific aspects of variation on CSD.

Moreover, Mandarin speakers learning English display the following segment as having the strongest effect on CSD, where consonants favour deletion and vowels favour retention, for this sample, following pauses favoured retention at a higher rate than vowels (Edwards, 2011). Again, showcasing how first-generation NNS are sensitive to dialect-specific constraints on variation.

The results from the preceding segment constraint among other first-generation NNS do not have consistent results like those presented earlier for the following segment constraint; in terms of statistical significance, the strength of effect and internal hierarchy structures displayed.

These seemingly inconsistent results do, nevertheless, represent very similar underlying tendencies across different first-generation NNS for the variable CSD and should not, therefore, blind us from making inferences about the tendencies of "transformation under transfer" typical amongst NNS.

More specifically, since the preceding segment constraint, unlike the following segment constraint, is more likely to host L1 transfer for CSD, as I have proposed earlier (section 2.3.1.1), we will see different representations of L1 transfer amongst first-generation NNS resulting in a less consistent account for tendencies for the preceding segment among first-generation NNS.

Moreover, any inconsistencies in statistical significance, the strength of effects and the internal hierarchies of CSD among different migrant groups is, in fact, showcasing how first-generation NNS are attentive to NS-dialect-specific aspects of variation, whereby different NNS groups acquire dialect-specific aspects of target varieties that have different local norms.

The strength of effect reported for the preceding segment constraint on CSD is different across groups of first-generation NNS. For example, first-generation Italian migrants to Canada illustrated that the preceding segment has the weakest effect on CSD (Hoffman and Walker 2010). Another example illustrates that Chinese speakers learning English in the U.S.A were found to have the preceding segment constraints as the second most significant constraints after the following segment (Bayley, 1996).

Interestingly, the internal hierarchy for the preceding segment constraint remained consistent across different groups of first-generation NNS. This pattern replicates the findings for my sample of old and middle-aged AMs. For example, Italian migrants to Canada illustrate an internal hierarchy where nasals favour deletion the most, followed by fricatives then stops and laterals that favour deletion the least (Hoffman & Walker, 2010). Chinese learners of English in the U.S illustrated the same internal hierarchy where a split exists between nasals and other consonants; the preceding nasal favour deletion the most followed by other consonants (Bayley, 1996).

If we consider the re-syllabification account we adopted earlier for AM, then the different internal hierarchy structures for the preceding segment constraint are easily explained as typical contexts of first language transfer. And since the first-generation migrant groups have different L1, it is only logical to expect different acceptable syllables and therefore different internal hierarchies.

For example, first-generation of Italian migrants to Canada applied paragoge and/or re-syllabification on target words (Hoffman & Walker, 2010)). Young-Scholten & Archibald (2000) suggest that learners of English with romance languages may be inclined to transfer first language processes that preserve sounds rather than those that simplify or delete clusters.

Mandarin native speakers deleted more coronal stops syllable finally, if preceded by an /n/ due to forming acceptable singleton codas allowable in mandarin (cf. , 1996; Lim & Guy, 2005; Hoffman & Walker, 2010; Edwards, 2011), a similar tendency was found amongst native speakers of Singapore English (Lim & Guy, 2005).

The preceding segment constraint's hierarchy structure is not only subject to L1 transfer but may also be subject to developmental stages of learners' grammars in the L2. For example, Young (1991) documented changes in the internal ordering of the preceding segment constraint in the acquisition of English past tense morphology among Chinese learners of English because of increased proficiency in the L2.

This reordering of internal hierarchy structure is, nevertheless, sometimes dependent on NNS-specific constraints like the level of proficiency. For example, highly advanced Irish learners of French do not seem to change the hierarchy structure even after moving into the target language country, what changes is mainly the rate of a variant (Regan 1996).
The results for the role of grammatical constraints cannot be explained without an excellent understanding of what a NNS-group consider their target group. So, I can only speculate about the reasons behind patterns of acquisition for this constraint. For example, Chinese learners of English who were taught English formally illustrated the grammatical category as a significant constraint on CSD with a hierarchy like that of NS of English; monomorphemes favour deletion and past tenses disfavour deletion (Bayley 1996).

The story told across generations of migrants to Canada is as interesting as it is revealing of the role of the grammatical category constraint on CSD. First-generation of Italians and Chinese migrants to Canada were reported to have the grammatical category as the constraint with the strongest effect on CSD. These migrants, nevertheless, seemed to delete past tenses at a rate as high as monomorphemes (Hoffman & Walker, 2010). These migrants' developing grammars may have been frozen because of low interaction (network

involvement) with NS and social isolation. Their low-linguistic performance, accordingly, reflects their developing L2 grammars as they treat past tense forms as unanalysed lexical items.

Second-generation Italian-Canadians and Chinese-Canadians illustrated a decrease in the factor strength of the grammatical category; both migrant groups had the grammatical category as the least significant constraint. There was also a change in the internal hierarchy, compared with first-generation migrants of the same ethnic groups, whereby monomorphemes favoured deletion more than past tenses for second-generation migrants. This can be explained in terms of the intensified exposure these two groups had with NS of English, which enabled them to acquire NS-like norms.

Second-generation canadians seem to have acquired linguistic competence and are, therefore, ready to acquire sociolinguistic competence. A point to mention is that this research is assuming that all three generations of ethnic minorities share the same NS-target group which might not be entirely true because sometimes people acquire the norms of a target group that is most revealing to them; be it to acquire an identity they admire or even a social affiliation (see Mendoza-Denton, 2008; Podesva, 2007).

What is interesting about these results is that, like old and middle-aged AMs, first-generation migrants display strong transfer of constraints from English. This result might be surprising at first, but some consideration of NNS-specific social constraints may inform our understanding of the acquisition of variation by NNS.

These first-generation migrants had low if any exposure to English before leaving their countries. Their proficiency in English did not change much upon arrival because of maintaining social networks with other migrants as well as illustrating social isolation from NS (see Hoffman and Walker 2010 for a detailed account of accommodation preferences, neighbourhoods and jobs that migrants keep in Canada). The low proficiency in English

means that NNS have no source for innovation to affect NS constraints except for L1 transfer. This means that NNS of this type would be more likely to accept NS constraints at face value with similar rank orderings, internal hierarchies and frequencies of occurrence.

5.6 Discussion of the patterns of variation for the variable CSD among young AMs

In this section of the chapter, I will discuss the results for young AMs within the framework of the questions that I have raised for this sub-group of NNS. Specifically;

1. Do young AMs illustrate the same constraints on CSD that are found significant in the NZSED?

2. If the answer is yes, do young AMs display rank orderings of significant constraints like those found in NZSED for the variable CSD?

3. Do the constraints found significant for young AMs have similar internal constraint hierarchies as those found in NZSED for the variable CSD? What are the constraints that are non-significant for them? And are there any young-AM-specific constraints?

4. Are there any implications of "transformation under transfer" (Meyerhoff, 2009a; Meyerhoff & Schleef, 2012)? If the answer is yes, are there any signs of weak transfer, strong transfer and calquing?

5. Are the patterns of variation for CSD among young AMs found among other 1.5 and 2nd generations who live in similar contact settings?

Chapter 4 illustrated that the acquisition of constraints on linguistic variation patterns for CSD is systematic and stable for young AMs only in terms of significant constraints as they diverge from NS norms in the rank ordering of significant constraints, their internal hierarchies as well as associated strength sizes.

5.6.1 The three lines of evidence outlined for CSD among young AMs

The results for the young AMs revealed that the rate of deletion is much higher than that reported for old and middle-aged AMs group and for NS of NZSED, 36%. I do not think, however, that this may reflect age-grading among Arabs where young AMs delete at a higher rate of old and middle-aged NS because unlike old and middle-aged AMs, young ones are more NS-like.

Young AMs illustrate that they share the same constraints found for NS of NZSED, they also illustrate that the constraints found to be non-significant among NS were also found to be non-significant for them; gender and style-shifting.

Young AMs diverge from NS, however, in the rank ordering of the significant constraints, their internal hierarchies and strength of effect. For example, young AMs display that the preceding segment constraint has the strongest effect followed by the following segment constraint and the grammatical category constraint. The difference in constraint effect for the phonological constraint is small, with the same range difference between the following segment and the preceding segment found among NS of NZSED. But young AMs have reduced strength allocated to these constraints and the grammatical category constraint has a much lower effect.

This pattern diverges from that found for NS of NZSED whose results illustrate that the following segment has the strongest effect size on CSD, followed by the preceding segment and the grammatical category constraints with a gap between the effect size of the phonological constraints, that have close ranges and effect sizes, and the grammatical ones.

Young AMs illustrate that the internal hierarchy for the preceding segment is relatively different from that of NS. Although preceding nasals represent the context most favouring of deletion for both young AMs and NS, the two groups have different factor weights allocated to this level of the preceding segment constraint. Young AMs illustrate that

preceding fricatives, both sibilant and non-sibilant have an almost neutral effect on CSD with a factor weight of 0.54. Liquids disfavour deletion followed by stops that disfavour deletion the most. On the other hand, NS of NZSED illustrate that liquids have a neutral effect on CSD, whereas stops and sibilant fricatives slightly favour retention and non-sibilant fricatives favour retention the most.

Several accounts were suggested to explain the role of the preceding segment constraint like the sonority hierarchy, by which we would expect less sonorant sounds to favour deletion and more sonorant ones to favour retention. This account does not, however, explain the internal hierarchy observed for young AMs.

Another account goes along the lines of the obligatory contour principal (OCP) as suggested by Guy and Boberg (1997), by which no two-adjacent sounds can be alike. According to this principle, we would expect coronal sounds including nasals, stops and fricatives to be deleted at high rates because of their likeness to /t/ and /d/ sounds. This may hold true for young AMs as nasals favour deletion the most followed by fricatives, stops and liquids with very close factor weights.

What is interesting about the internal hierarchy for the preceding segment constraint among young AMs is that it reflects the status of young AMs' norms as a mixture of old-and-middle-aged-NNS norms with NS ones on variation. Specifically, NS of NZSED illustrate a clear cut-off point between preceding nasals that favour deletion with a factor weight of 75, and other consonants that collectively disfavour deletion to varying degrees.

Old and middle-aged AMs , on the other hand, seem to have captured the nasal-other consonants dichotomy but had L1 transfer affect the internal structure of the preceding segment due to re-syllabification tendencies as expressed earlier. So, fricatives favoured deletion while stops and liquids had neutral effects on CSD.

Young AMs illustrate that preceding nasals favour deletion the most, unlike NS of NZSED and old and middle-aged AMs , fricatives, both sibilant and non-sibilant have a neutral effect on CSD, while stops and liquids favour deletion the most following the internal order of old and middle-aged AMs but with heightened factor weights.

A point to mention here is that as I have suggested earlier, phonological constraints may have room for first language transfer that may or may not become markers of ethnicity or identity at large, and since the preceding segment is especially prone to reflect ethnolectal and identity marking, then the behaviour of younger Arabs may reflect ethnolectal marking and not L1 transfer that was applicable for old and middle-aged AMs .

The following segment constraint has the second strongest effect on CSD among young AMs with consonants favouring deletion and vowels and pauses favouring retention. Younger NNS seem to have acquired the following segment constraints with a different internal hierarchy of that of NS of NZSED. Young AMs do, nevertheless, acquire the dialect-specific aspect of CSD variation as they have replicated NS patterns of the following pause favouring retention the most.

NS of NZSED have the following hierarchy stop < non-sibilant fricatives < nasals < glides < sibilant fricatives < vowels < pauses and seem to follow a sonority hierarchy moving from least to most sonorant. Whereas young AMs displayed the same internal hierarchy of that of old and middle-aged AMs with very close values of factor weights as well.

The following nasals and obstruents favour deletion the most with very similar factor weights. like old and middle-aged AMs , who treat glides and vowels alike favouring retention, young AMs illustrate a similar tendency, in this case, surely not as a reflection of developing acquisition of the L2 phonetic inventory. Young AMs have acquired the linguistic constraints of English having arrived in New Zealand no later than the age of five. Therefore,

this internal hierarchy could be a manipulation of constraints that defines their identities as Arab-New Zealanders.

The factor weight of the following nasal for young AMs diverges from a sonority hierarchy explanation, and a re-syllabification-based explanation might be a better one here too by which 1.5 generation migrants mark their identities.

Unlike old and middle-aged AMs , young AMs display the grammatical category as a significant constraint on CSD; monomorphemes favour deletion and regular past tense favours retention with irregular past tense having a neutral effect on CSD, unlike NS for whom irregular past disfavour deletion[22].

---

[22] A point to mention is that the data did not present enough tokens for the grammatical category discourse marker, and therefore, I cannot comment whether this subgroup of speakers treat discourse markers differently from other categories.

5.7 Conclusions for CSD among young AMs

The discussion presented above based on the comparison between NS and young AMs patterns of variation on CSD reveal that young AMs illustrate a tendency of "transformation under transfer". More specifically, they display instances of weak transfer whereby they replicate the preceding segment constraint, the following segment and the grammatical category constraints with rank ordering and internal hierarchies different from the ones applicable by NZSED speakers.

What is interesting about this line of evidence is that it is unexpected in the sense that we would assume young AMs to illustrate the same constraint ranking and internal hierarchies as found among NS of NZSED. What adds to the peculiarity of the situation is that old and middle-aged AMs who are expected to diverge sharply from NS norms have illustrated strong transfer.

One way of understanding this case is by considering the levels of proficiency and the types of proficiency achieved by NNS groups. Old and middle-aged AMs , as suggested by their results, have developing grammars of NZE. They are, therefore, still building their linguistic competence and are probably have not started to approach local norms, sociolinguistic competence. They are, therefore, more likely to readily accept NS-norms, as is, with no reordering of constraints except because of L1 transfer; that is **reordering for old and middle-aged AMs is not stylistic.**

Young AMs, on the other hand, seem to have a full acquisition of the linguistic competence of the L2 and are probably stylistically deploying the constraints to display their sociolinguistic competence in their own way. Schleef et al. (2011) suggested that migrants seem to innovate constraints that are most meaningful to them, and the constraints of young AMs may indeed be an illustration of relevant social meanings conveyed as a reordering of

NS-constraints; whereby **1.5 generation of Arabs stylistically use what was formerly an L1 transfer effect to signal social meanings that are meaningful for them.**

This tendency could also be an aspect of an emerging ethnolect, but any further discussion of the possibility of an emerging ethnolect calls for more research across a wider range of Arab New Zealanders as well as comparing different Arab groups based on levels of integration in the New Zealand community, like the Lebanese, who are very integrated, versus newly arriving Syrian refugees.

It is also possible that young AMs are not entirely NS of NZE; although they are very proficient in the L2, they are still illustrating L1 transfer effects, probably as an effect of interacting with their parents mostly in Arabic as explicitly mentioned by them.

5.8 Placing young AMs' patterns of variation in CSD amongst other 1.5 and Second-generation migrants in contact settings

Other 1.5 generation migrants in contact settings were harder to locate, but I believe using second-generation migrant groups will suffice. For example, second-generation migrants to Canada from different ethnic backgrounds (Italian, Chinese) illustrated that the preceding segment had the strongest effect on CSD with nasals favouring deletion the most followed by fricatives, stops and laterals.

Interestingly, third-generation, Canadian-born Chinese and Italian migrants, who are reported to have ethnolects of their own, have also displayed the preceding segment as the constraint with the strongest effect on CSD.

This suggests to me that the preceding segment may, indeed, host what starts as an L1 transfer for first-generation of migrants, then becomes a stylistic resource for host-country born migrants and then becomes an aspect of ethnolectal marking among native speakers of the resulting ethnolect.

Other examples may be illustrated from well-established ethnolects like AAVE and Chicano English. Studies of CSD variation revealed that AAVE illustrated the preceding segment constraint as having stronger effects than the following segment constraint, probably as a window for ethnic marking (Cohen & Labov, 1967; Wolfram, 1969; Fasold, 1972) and Chicano English revealed a similar pattern (Santa Ana , 1991).

Like young AMs, the following segment constraint, seem to have the second strongest effect on CSD by second and third generation migrants with an internal hierarchy similar to that found for young AMs in NZ; following consonants favour deletion and following vowels/paused favour retention. But the role of the grammatical constraints, although significant for some second-generation migrants, has different weightings by third

generation-migrants, who are also NS of ethnolects (see Hoffman and Walker (2010): table 4 for details).

These findings support my conclusions for young AMs; second-generation migrants have full linguistic competence in the L2, and they reorder phonological constraints, either because they still display some L1 transfer effects due to interaction with first-generation migrants who are their parents and grandparents or because they are using these constraints as a stylistic resource.

Third-generation migrants either accommodate completely with NS of an L2 and display the same constraint orderings, internal hierarchies, significant and non-significant constraints and rate of application value for a variable. Or they illustrate the emergence of ethnolects, where what started as L1 transfer amongst their grandparents, has become a stylistic marking typical of ethnolects.

To summarise, first-generation migrants illustrate L1 transfer that result in internal hierarchies that are different from those of NS of NZSED. They also have developing linguistic competence in the L2 that are also partly responsible for their unique internal hierarchies. Whereas the 1.5 generation of AMs, who have mastered linguistic and sociolinguistic competence of the L2 illustrate weak transfer from the L2 probably because they use their knowledge stylistically and reorder constraints to convey a social meaning that is of value to them. It is also possible that the results for young AM illustrate traces of an ethnolects of their own but I can only speculate about this.

5.9 Discussion of results for (ING) among NS of NZSED

This part of chapter 5 discusses the linguistic and the social constraints that condition variation on (ING) among Native speakers of New Zealand Spoken English Database, within the framework of the following questions:

1. Do NS of NZSED share the same significant constraints on (ING) with other varieties of NS English, i.e., phonological constraints and grammatical category constraints?

2. If the answer is yes, do NS of NZSED display constraint rankings as found in other NS varieties of English?

3. Do the significant constraints found in NZSED have similar internal hierarchies as are found in other NS varieties of English?

4. What are the constraints that are non-significant in NZSED, and are they the same as constraints found non-significant in other NS varieties of English?

5.9.1 The three lines of evidence outlined for (ING) among NS of NZSED

The results from chapter 4 illustrate that when the speaker and word are considered as random effects, the significant constraints on (ING) among NS of NZSED are the same as those found in other varieties of NS English. The results, however, illustrate that NZSED have strength of effects and internal hierarchies that differ from other varieties of NS English.

The frequency of the alveolar variant in NZSED (Wellington data) is 60%. This result is much higher than that reported for working class Pakeha speakers in the Porirua corpus; 30% (Bell & Holmes, 1992).

I can only speculate why the rate in my data is much higher because I have not coded for social status and I have not tried to figure out the social salience of the alveolar variant and any associated social judgements, but I can suggest that the frequency of the alveolar nasal seems to be dialect-specific.

For example, Labov (2001) illustrates that southern states English, Northern English in England and Scots in the UK use the alveolar variant almost exclusively across styles and registers, resulting in high frequencies of the application of this variant. Laturnus et al. (2016) illustrate that southern U.S states have higher rates of /n/ compared with northern dialects. African American communities are also reported to have high frequencies of the alveolar variant (Wolfram & Christian, 1979; Kendall, 2010).

Not only are the rates of the alveolar nasal dialect-specific but also are the social evaluations associated with this variant (See Campbell-Kibler (2012)). Although I have not considered attitudes and perceptions regarding the social meaning of the variable (ING), it does not seem to me that the variable /n/ carries as much social stigma as it does in some parts of the United States, where only a few tokens of the alveolar variant can trigger negative social judgements (Labov et al., 2006).

As for the rank order of the significant constraints in NZSED, my results suggest that the grammatical category constraint has the strongest effect on (ING) with a range of 59, followed by the following segment constraint with a close range of 56. Bell and Holmes (1992) reporting on the Porirua data suggested similar results; the grammatical category constraint has the strongest effect on variation with a rate of 50, followed by stress patterns on the syllable preceding the (ING) with a range of 37 and the following segment constraint has the least effect on (ING) with a range of 24.

Therefore, it is plausible to suggest that the different strength of effect reported for the following segment constraint on (ING) may be dialect-specific. For example, some British dialects (Houston, 1985) and American dialects (Abramowicz, 2007) illustrate that the following phonological constraint has a strong effect on (ING). Whereas other dialects of the U. S illustrate that this constraint seems to have "no strong phonological conditioning before the following velars or apicals" Labov (2001:87) and other dialects show that the following

segment is not significant at all (Hazen 2006). This is relevant when I discuss the acquisition of variation of (ING) by NNS, to see whether NNS acquire dialect-specific constraints as well as other constraints on (ING).

Grammatical constraints on the variable (ING) in NZSED have the strongest effect size, but the constraint's strength is very close to the following segment constraints as suggested by the range value for both constraints (59 and 56).

The internal hierarchy for the grammatical category constraint, unlike the internal hierarchy reported by (Bell & Holmes, 1992) for the Porirua data, does not completely follow the Labovian nominal-verbal continuum, whereby nominal categories favour the velar nasal variant the most and more verbal categories favour the alveolar nasal variant as follows: progressive – gerundial participle - adjectives - gerunds and nouns (Labov, 2001).

The internal hierarchy for the NZSED illustrates that progressives, nominal gerunds, participles and adjectives favour the alveolar nasals, while discourse markers and pronouns disfavour the alveolar nasals the most with a huge range gap.

An interesting discussion arises about this pattern; nominal gerunds and progressives that are verbal in function are treated alike by NS of NZSED; in the sense that they both favour the alveolar nasal variant. Participles and adjectives that have an adjectival function are treated alike as well, even *-ing* ending nouns are treated like them favouring the alveolar nasal. Discourse markers and pronouns are at the other end of the continuum, clearly favouring velar nasals as suggested by the factor weights of 22 and 11, respectively.

A possible explanation may be that the nominal-verbal distinctions are fading out among NS of NZSED who do not seem to perceive nominal-verbal distinctions as much as NS of English elsewhere. Probably because of diachronic change of the variable (ING) as

Bell and Holmes (1992) noted that variation in (ING) might have roots in syntactic and geographic differences which originate in old and Middle English[23].

Progressives and gerunds introduce similar action-related meanings, the sentence "he was *walking*" and "*walking* is fun" are semantically very similar, referring to the action of moving on foot. Participles and adjectives provide similar adjectival functions, so why would NS of English, whose grammatical knowledge is below the level of consciousness, treat two structures that provide similar functions separately. Similar findings were reported in the United States of America where verbs, gerunds, adjectives and nouns favoured the alveolar nasal to varying degrees and the Labovian nominal-verbal continuum was not validated (Forrest, 2015).

Another reason for this internal hierarchy may be rooted in methodology and the underlying assumptions of coding. For example, most studies code gerunds as discrete noun phrases whereas this assumption may be mistaken to start with (for a detailed discussion see Houston 1985). Another example is the coding of adjectives and verbal participles which functionally have adjectival functions.

The last point to mention about the internal hierarchy of the grammatical constraints among NZSED is that discourse markers and pronouns are the most favouring of the velar nasal realisation with a big gap existing between them and other grammatical categories in terms of factor weights.

---

[23] Labov (1989) suggested that the influence of the grammatical category constraints may be traced back to old English –*inde* for verbal participles and –*inge* for verbal nouns.

An important point to mention in this context is that the coding I used may have caused discourse markers and pronouns to take an extreme end, highly favouring the alveolar nasal, therefore creating an internal hierarchy that differs from the results reported for NZE and other NS varieties of English for this variable.

A closer look at this result, however, reveals that these discourse markers are *-thing* compounds as in *and everything*, *or anything*, and all *–thing* compounds were previously reported to favour the velar nasal in NZE (Bell and Holmes 1992). This explains why discourse markers seemingly appear to favour the velar nasal because of its high level of nominalization, where in fact it is because discourse markers involve the use of *–thing* compounds.

These results about the behaviour of *-thing* compounds compel me to support Gordon (1998), who suggested that *-thing* compounds include *-thing* morphemes and not variants of (ING). In other words, *-thing* morphemes and (ING) have intersecting variants that should probably be separated in future research.

The Porirua data (Bell & Holmes, 1992) reported an internal hierarchy for the following segment constraint that is different from that illustrated for the NZSED, in terms of the order of the internal levels and their associated factor weights. The Porirua corpus illustrates that alveolar sounds favour the alveolar nasal the most with a factor weight of 60. Vowels had a neutral effect on (ING) and velars and pauses disfavoured the alveolar nasal the most.

This pattern was explained along the lines of place of articulation as a source of assimilation; alveolar sounds share the same place of articulation with the alveolar nasal and therefore favour it the most, whereas velars have a different place of articulation and disfavour the alveolar nasal the most, while the role of the following pause seems to be dialect-specific (Bell & Holmes, 1992).

The internal hierarchy for the following segment constraint on (ING) among NZSED speakers diverges from the abovementioned pattern. NZSED illustrates that alveolar sounds and velar sounds favour the alveolar nasal the most with factor weights of 76 and 61, respectively. Whereas vowels slightly disfavour it and pauses disfavour the alveolar nasal the most.

Obviously, an explanation of this internal hierarchy based in place of articulation assimilation across adjacent sounds does not work, because following alveolar and velar sounds favour the alveolar nasal the most. It could be that the alveolar nasal dissimilates in the manner of articulation with following oral sounds because vowels do not have the same characteristics as consonants. This explanation is supported by the data as I noticed that the following vowels have a neutral factor weight, neither favouring velar or on alveolar nasal variants.

Neither the preceding phonological constraint nor the social constraints included in this study were found significant amongst NS of NZSED. Unlike findings from the Porirua corpus (Bell & Holmes, 1992), gender is not a significant constraint on (ING) probably due to using speaker and word as random effects.

Style-shifting, which is usually a persistent constraint reported to condition (ING) across varieties of NS English was also found to be non-significant across speakers of NZSED. This might be because the conversations that I have analysed are casual and I have not included other tasks, like sentence reading or minimal pair reading, that may be perceived as having varying formality levels by NS of NZSED.

It would, therefore, be hasty to state that style-shifting is non-significant in NZSED as more research needs to be conducted including different tasks. Age was also found to be non-significant; it is worth mentioning, nevertheless, that the non-parametric regression illustrated that age interacts with other constraints and its effect in predicting variation on (ING) is,

therefore, to be considered in tandem with the grammatical category and the following segment constraints.

Some dialects of English, both in the United States and the UK seem to have statistically significant persistence effects (Abramowicz, 2007; Schleef et al., 2011; Drummond, 2012; Tamminga, 2014; Laturnus et al., 2016), whereas this constraint was found to be non-significant for NS of NZSED. This suggests to me that persistence, as a constraint conditioning variation in (ING), might be dialect-specific. For example, neighbouring dialects within the UK illustrate not only different ranking order for persistence but also different internal hierarchies and factor weights (see Schleef et al. (2011) and Schleef 2017a for NS of London and Edinburgh).

Table 5.8 is an extract from regression analysis in Rbrul, it illustrates that although persistence was found to be non-significant for NS of NZSED, it patterns in the expected direction. If the prior token had an alveolar realisation, the current one would slightly favour the alveolar nasal, and when the prior token is nominal or verbal, the alveolar variant is favoured for the current token. If the prior token was a discourse marker, then the current token would favour the velar nasal.

Table 5.8: Extract from regression analysis for(ING). NS of NZSED. Application value is the alveolar nasal			
Realisation_prior_token			
	logodds	tokens	n/n+g+ING centred	factor weight
Alveolar nasal	0.223	510	0.673	0.555
Reinforced (ING)	-0.058	42	0.500	0.485
Velar nasal	-0.165	304	0.484	0.459
prior-grammatical-category			
	logodds	tokens	n/n+g+ING centred	factor weight
nominal	0.144	364	0.604	0.536
verbal	0.092	450	0.609	0.523
Discourse-Marker	-0.236	42	0.405	0.441

This pattern seems to comply best with a simplistic definition of persistence effect as "the idea that the realisation of one (ING) can then affect the realisation of a subsequent (ING), favouring similarity" (Drummond (2012:107). It seems that the tendency to repeat variants is a "mechanical production-related phenomenon" when a following word is close enough and is of the same word class.

Whether some abstract structure stays in the memory of speakers, below their level of consciousness, and triggers a repetition of variants, is also a reason for the surface result of persistence effect, needs more investigation that was beyond the scope of my current research. I did not find evidence that persistence arises when the prime and target belong to the same morphological category as was suggested by Tamminga (2014).

5.10 Conclusions for (ING) among NS of NZSED

NS of NZSED share the same constraints on (ING) with other NS varieties with different strengths of effect and internal hierarchies, NZSED also does not display significant social factors that other varieties have as significant probably because word and speaker were coded as random effects.

The grammatical category constraint has an internal hierarchy that does not entirely match the nominal-verbal continuum that Labov (2001) suggested, but this pattern was reported across some NS varieties of English. The internal hierarchy for the following segment is best explained along the lines of dissimilation of the manner of articulation across adjacent sounds.

5.11 Discussion of the results of (ING) among old and middle-aged AMs

This part of chapter 5 discusses the linguistic and the social constraints that condition variation on (ING) by old and middle-aged AMs within the framework of the following questions:

1. Do old and middle-aged AM illustrate the same significant constraints on (ING) that are found in the NZSED? If the answer is yes, do old and middle-aged AMs display rank ordering of constraints like those found in NZSED?

2. Do the constraints found significant for old and middle-aged AMs have similar internal hierarchies as those found in NZSED for the variable (ING)?

3. What are the constraints that are non-significant for old and middle-aged AMs? And are there any AM-specific constraints?

4. Are there any implications of "transformation under transfer" (Meyerhoff, 2009a; Meyerhoff & Schleef, 2012)? If the answer is yes, are there any signs of weak transfer, strong transfer and calquing?

5. Are the patterns of variation for (ING) among old and middle-aged AMs found among other first-generation migrants who live in similar contact settings?

I started with a model that included all age groups for NNS; old, middle-aged and young. This collective model, however, had many convergence problems and as outlined earlier, many reasons compelled me to create two separate models for old and middle NNS and young AMs.

A persistence effect was detected for the collective model; when a previous token has an alveolar realisation, the current token favours an alveolar realisation as well, and if a previous token has a velar realisation, the current one is more likely to have a velar realisation. This constraint became non-significant, however, when I created two AM groups.

5.11.1 The three lines of evidence outlined for (ING) among old and middle-aged AMs

The model created for old and middle-aged speakers indicate that the grammatical category and the preceding segment are the only two significant constraints on (ING) amongst old and middle-aged AMs with a gap in terms of the strength of effect for each constraint; 49 and 29. The following segment constraint was found to be not significant for this subgroup of NNS, and the rate of the alveolar nasal is 50 %.

The grammatical category constraint has the strongest effect on (ING), as progressives favour the alveolar nasal the most with a factor weight of 0.72, followed by gerunds with a factor weight of 0.57. Nouns and pronouns slightly favour the velar nasal while adjectives favour the velar nasal the most with a factor weight of 0.23.

What is interesting about the internal hierarchy for this constraint is that old and middle-aged AMs seem to treat progressives and gerunds like NS of NZSED do, probably because they provide the same information; progressives and gerunds provide very similar information.

The way this subgroup of AMs handles *-thing* compounds may inform our previous conclusions about *-thing* compounds among NS of NZSED. I have included *-thing* compounds in the coding of (ING) although they are not a part of the historical development of (ING) in English for NS because I wanted to create a benchmark to be able to see how differently/similar AM would handle this grammatical category. Interestingly, first-generation migrants to NZ seem also to treat *-thing* compounds differently which illustrates their capacity to distinguish variants of (ING) against variables with intersecting variants like *-thing* compounds.

As for adjectives, old and middle-aged AMs seem to have developing grammars, and they cannot yet associate adjectives and participle as conveying similar adjectival functions.

The constraint with the second strongest effect is the preceding segment. Liquids favour the alveolar nasals the most with a factor weight of 0.65. Velars also favour the alveolar nasal with a factor weight of 0.59, whereas apicals disfavour the alveolar nasal the most with a factor weight of 0.36.

One way to understand the internal hierarchy for the preceding segment is by cross-tabulating it with the grammatical category to see if there are any interactions, i.e., combinations that may have surfaced as a phonological effect. Nevertheless, the resulting table was not informative.

So, I resorted to another explanation that may justify the internal hierarchy for the preceding segment among Old and middle-aged AMs which is re-syllabification. As I proposed earlier for CSD, Old oOld and middle-aged AMs seem to be re-syllabifying preceding sounds to create acceptable L1 syllable structures. This proposition is also plausible in the case of (ING) since Old and middle-aged AMs may also be re-syllabifying the preceding segment with the variant of (ING), influenced by their L1 to create acceptable L1 syllable structures. It is also interesting if we consider that this group of AMs may be handling (ING) as two variables; a vowel and a nasal variable but this aspect of variation was not carried through I this thesis.

5.12 Conclusions for (ING) among old and middle-aged AMs

Old and middle-aged AMs seem to share only one significant constraint with NZSED; grammatical category. They, however, do not allocate the same strength of effect found among NZSED speakers for this constraint although they seem to have acquired the same rank order. This task is complex especially that the grammatical category constraint has a strength of effect that is almost equally as significant as the following segment constraint among NS of NZSED.

Old and middle-aged AMs have also partially acquired the internal hierarchy for this constraint compared with NS of NZSED. This suggests that old and middle-aged AMs are sensitive to dialect-specific constraints on (ING) variation; since the internal hierarchy, as argued earlier, is dialect-specific.

Old and middle-aged AMs acquired the rule that progressives and gerunds behave alike and favour the alveolar nasal, while nominal categories favour the alveolar nasal less. They have partially acquired the linguistic competence of NZE, their knowledge is, nevertheless, still developing. Therefore they are incapable of completely replicating NZSED patterns.

Old AMs have had no exposure to English before arrival to Wellington, and their acquisition of English is mainly integrative. As mentioned above, migrants to New Zealand are required to take English courses, this may explain how old AMs have partially acquired the grammar of NZE. But since old AMs seem to maintain Arab only networks, their interlanguage freezes at some point, especially since they can quit English classes when they find a job.

Middle-aged speakers are also the ones who have been in NZ for less than three years, and as Wolfram (1984) suggests, NNS who spend a period longer than three years in a host country are more likely to target linguistic competence at this point. But unlike old AM, middle-aged ones would probably carry on developing their interlanguage because they all have jobs in NZ and they maintain mainly Non-Arab networks.

Once more, the results suggest that old and middle-aged AMs have a tendency for "transformation under transfer". Although (ING) is categorized as a morpho-phonemic variable, that is more complex than phonological variables like CSD, this group of AMs seem to have a weak transfer of NZSED constraints on (ING). This result is interesting because it

is very unexpected, especially with old AMs' no previous knowledge of English and middle-aged speakers' less than three years residency in NZ.

What is interesting is that old and middle-aged AMs seem to have created a preceding segment constraint on (ING) most likely as an L1 transfer effect whereby they tend to re-syllabify the preceding segment with the following (ING).

Schleef et al. (2011) suggested that NNS tend to create constraints that are meaningful to them. These constraints could be a reinterpretation of a NS-constraint, or they could be an innovated NNS-specific constraint. I can only speculate about the source of the preceding segment constraint, but I believe it would be interesting to test if NNS perceive (ING) as a vowel followed by an [ng] sound cluster. Only then, one may study the role of co-variation of the vowel and the following nasal.

5.13 Placing old and middle-aged AMs patterns of (ING) amongst other first-generation migrants in contact settings

Placing the results for (ING) variation patterns for this subgroup of NNS amongst other studies of the acquisition of the same variable by first-generation NNS groups is not as straightforward as the discussion I presented earlier for CSD in various contact settings. Specifically, different ethnic groups with different L1s seem to have differential tendencies in the patterns of variation for (ING). Nevertheless, the possible reasons for this result may be informative.

I, therefore, resolved to discuss the patterns of (ING), one group at a time, illustrating that NNS seem to acquire constraints on (ING) depending on (i) the range of constraints available in the target variety and (ii) some NNS-specific constraints. First-generation migrants, sometimes, create constraints on (ING) that are important to them either because (i) these constraints are the result of first language transfer or because (ii) they are driven by NNS-specific constraints, like networks.

The first example is drawn from a study of Cambodian and Vietnamese migrants to Philadelphia and Washington DC. These students replicated phonological and stylistic constraints on (ING) as found in the NS varieties. They, nevertheless, did not replicate the NS rank ordering and did not display the preceding segment and the grammatical category constraints as significant (Adamson & Regan, 1991).This NNS pattern can be explained as a result of developing L2 grammars; the NNS are study-abroad students who are actively acquiring the L2 linguistic competence.

These students were, nevertheless, able to acquire dialect-specific constraints, namely: style-shifting and the frequency of the alveolar nasal; female NNS exactly replicates female NS rates, whereas male NNS exceeds the male NS rates. A very important point to mention here is that NNS exact replication of NS frequencies may be (i) a reflection of the acquisition

of NS frequencies, thus illustrating perfect cases of accommodation with target-group norms; (ii) a reflection of the acquisition of NS style constraints, thus replicating a variant in specific contexts, like the use of unmarked variables by female NNS that most likely replicates female NS norms; (iii) reflect the acquisition of NS gender norms, whereby NNS males target NS males, and NNS females target NS females.

But what we mostly observe for NNS is a "reversal" of NS style-shifting norms (except in the case of female NNS replicating the same female NS frequencies as illustrated in table 5.9), a mismatch between frequencies in NS speech and that of NNS, and possibly a reinterpreted NNS-understanding of NS gender roles.

The first step to understand what seems to be the NNS males' inability to achieve target-group norms is to acknowledge that similar findings were reported across many other male NNS studies for the style-shifting constraint. The authors of such studies have referred to the seemingly unexpected behaviour of male respondents as (i) limitations to the accommodation of target group variation (Adamson & Regan, 1991; Howard et al., 2006; Meyerhoff & Schleef, 2012), (ii) the reversal of NS style-shifting norms (Adamson & Regan, 1991; Uritescu et al. , 2002; Mougeon et al., 2004; Howard et al., 2006), (iii) a reversal of typical NS gender norms (Drummond, 2011, 2012; Schleef, 2013a, 2013b), or (iv) the invention of NNS-specific social factors, gender (Schleef et al., 2011).Table 5.9 illustrates several examples where NNS illustrate a tendency to reverse NS style-related norms.

Table 5.9: Examples of reversal of style among NNS				
Study	Variable	Respondents	Type of respondent	Patterns of variation in style
Adamson and Regan 1991	(ING)	Cambodian and Vietnamese students in the U.S. A	Student Not advanced	Style: reversed by male students; more alveolar variants even in formal styles. Female students replicated the exact frequency of female NS
Uritescue, Mougeon and Handouleh 2002	Schwa deletion			Style: reversed; more deletions even in formal styles. But overall frequency less than NS Students who had family stays produced more deletions
Mougeon et al. 2004	Schwa deletion	Instructed students in French immersion programmes		Style: reversed, more deletion even in formal styles
Howard et al.2006	/l/ deletion	Advanced Irish learners of French		Style: reversed, more deletions even in formal styles
Drummond 2011	/t/ release and /t/ glottalling Word medially and word finally	Polish migrants to Manchester		Reallocation of gender effect but reversed (women produced more glottalized /t/, the author explains that women create prestige variants not only follow them.
Drummond 2012	(ING)	Polish migrants in Manchester	Low proficiency	Reallocation gender Reversal women more alveolar nasals
Schleef 2013 a	(ING) and t glottalling	Polish migrants in London and Edinburgh	Beginners	London /t/ glottalling Ing Creation of gender Reversed gender boys produced more velar nasals
Schleef 2013 b	T-glottalling word finally	Polish migrants in London	Three phases, same respondents	Style: reallocation Reversed effect more glottalling even in formal styles
Schleef 2017 a	T-glottalling word finally	Polish migrants in London	Three phases, same respondents	Style: reallocation Reversed effect more glottalling even in formal styles

It is possible that male NNS, in the studies from table 5.9, may be sensitive to peer pressure and they, therefore, end up overproducing what they perceive as the acceptable male NS norm as a result of a strong desire to accommodate them. Regan (1996) has suggested that NNS tend to overgeneralise NS norms, so they end up producing more variants, thus raising the frequency of occurrence of a given NS variant.

Schleef (2013a) suggests that male Polish migrants to London have reinterpreted NS gender constraints by producing more velar variants than female NNS. Another way of handling this piece of information, by considering it an example of the acquisition of NS style-shifting constraints and not a reversed NS gender effects.

The mere suspicion that Polish migrants may have created a gender constraint suggests that gender differences are an invaluable aspect of Polish migrant's mindsets, it is, therefore, plausible to suggest that Polish migrant boys might be pressured to accommodate, including their English proficiency levels, so they end up overproducing velar variants as a result of over-exaggerating NS norms in their efforts to assimilate. The assumption is that it is not only middle-aged men that may be the most pressured to produce variables with certain social indices but also boys in certain communities. Benor (2004), for example, revealed similar tendencies among Orthodox Jewish boys to assimilate to the variable choices that are associated with education and masculinity in the Orthodox Jewish communities (t-release).

It is crucial here that I stress the point that I am only speculating since I am not aware of any perception-related details for the above-mentioned studies, I have also not conducted any matched guise tests and have not included any attitude-related data for this project, except for some few complimentary comments from my respondents.

Moreover, the seemingly reordering of NS style-shifting is also widely illustrated in many studies of acquisition of variation in French that has discrete formal and informal variation patterns. For example, advanced Irish learners of French have increased their rate of

"ne" deletion after arriving in France (Regan, 1996). Another example amongst advanced learners of French illustrates that they produced more "ne" deletion even in most formal situations (Dewaele, 2004), probably as a way of expressing their localness. Other advanced learners of French produced more /l/ deletion even in very formal scenarios (Howard et al., 2006).

The second example, I draw on, to illustrate that the patterns I found among old and middle-aged Arab NNS might be a persistent pattern across many similar communities is from Schleef et al. (2011) who studied the acquisition of variation in (ING) amongst NNS Polish migrants to London and Edinburgh.

Native speakers of English in Edinburgh illustrate the following rank order of constraints: style-shifting, the grammatical category, persistence, number of syllables and overt attitudes to Edinburgh dialect. NNS Polish migrants to Edinburgh seem to be sensitive to dialect-specific rates of the velar nasal (application value in the Schleef et al. 2011), although style-shifting was not found to be significant for them.

Although (ING) is a complex morphophonemic variable that is not primarily conditioned by articulatory constraints, Schleef et al.'s sample of Polish NNS in Edinburgh replicated grammatical constraints but with an internal hierarchy that diverges from that of NS; conforming more with Labovian nominal-verbal continuum, as a result of supra-local constraints on (ING) (Schleef et al., 2011). They replicated priming with the same internal hierarchy as NS, probably because this constraint is cognitively predictable, and replicate the number of syllables constraint but reorder the internal hierarchy.

What is interesting about Schleef et al.'s findings is that, like old and middle-aged AMs who created a phonological constraint on (ING) that is non-existent among NZSED speakers, Polish migrants have created a phonological constraint on (ING) as well; the following segment which has an internal hierarchy that is compatible with regressive

homorganic dissimilation (following velars favour /n/ and following apicals favour /ŋ/). They have also created a network constraint that is meaningful to them.

NS of English in London, display a different group of significant constraints on (ING); the following segment, style-shifting, preceding segment and priming. Polish migrants to London only replicated one significant factor group, the preceding segment with the same internal hierarchy found amongst NS in London as well as managing to have a rate of the velar nasal that is similar to NS rates. They, however, created two NNS-specific constraints; grammatical category and gender.

The creation of a grammatical category effect is interesting because Polish migrants to London seem to have applied grammatical conditioning to (ING) exactly as expected by the Labovian nominal-verbal continuum. This is probably due to supra-local constraints that NNS have picked from previous exposure to English as suggested by the authors. It could also be the fact that Polish migrants to the school investigated in the Schleef et al. study (2011) have been the recipients of extra attention by their English language teachers, within the framework of governmental and educational policies, to ease their integration into the British community, as the authors explained. This may have provided Polish migrants with more instructed language learning which would usually introduce grammatical explanations and categorical distinctions that their peer NS Londoners do not achieve being NS of English.

To conclude, it seems that NNS with developing L2 grammars are sensitive to dialect-specific constraints on variation, including rank ordering because old and middle-aged AMs have replicated the grammatical constraints on (ING) as found among NZSED. A point to mention here is that NNS-specific constraints have an influence on the acquisition of variation like age at arrival in the host country, proficiency in the L2, network involvement and attitude towards the L2 and its speakers, a matter that makes NNS cross-group comparisons harder to achieve for this variable.

5.14 Discussion of the results of (ING) among young AMs

This part of chapter 5 discusses the linguistic and the social constraints that condition variation on (ING) among young AMs in Wellington, more specifically:

1. Do young AMs in Wellington illustrate the same constraints on (ING) that are found significant in the NZSED? If the answer is yes, do young AMs in Wellington display rank ordering of constraints like those found in NZSED?

2. Do the constraints found significant for young AMs in Wellington have similar internal hierarchies as those found in NZSED for the variable (ING)?

3. What are the constraints that are not significant for young AMs in Wellington? And are there any young-AM-specific constraints?

4. Are there any implications of "transformation under transfer" (Meyerhoff, 2009a; Meyerhoff & Schleef, 2012)? If the answer is yes, are there any signs of weak transfer, strong transfer and calquing?

5. Are the patterns of variation for (ING) among young AMs found among other 1.5 generations migrants who live in similar contact settings?

As mentioned earlier, I have started with a model that included all age groups for AMs, this collective model, however, had many convergence problems and as outlined before, many reasons compelled me to create two separate models for old and middle AMs and young AMs. A persistence effect was detected for the collective model; when a previous token has an alveolar realisation, the current token favours alveolar realisations as well, and if a previous token has a velar realisation the current one is more likely to have a velar realisation. This constraint became not significant, however, when I created two AM groups.

5.14.1 The three lines of evidence outlined for (ING) among young AMs

What is interesting about the results for young AMs for (ING) is that they seem to illustrate patterns of variation that are a mixture of NS of NZSED and old and middle-aged AMs. They, however, sometimes diverge from both groups as in the rate of the alveolar nasal which is 71 %.

Young AM share the same significant and non-significant constraints on (ING) with NS of NZSED with a different strength of effects and sometimes even slightly different internal hierarchies for these constraints.The results indicate that the grammatical category and the following segment are the only two significant constraints on (ING), among young AM, with a gap in terms of the strength of effect for each constraint with range values of 60 and 37, respectively.

Progressives favour the alveolar nasal the most with a factor weight of 0.81, followed by gerunds with a factor weight of 0.61. Nouns and pronouns slightly favour the velar nasal while adjectives favour the velar nasal the most with a factor weight of 0.28 (disfavour the alveolar nasal). A similar ordering of internal levels was found among old and middle-aged AMs, where nominals slightly favour the velar nasal and adjectives favour the velar nasal the most. This created a wider continuum of grammatical categories than that found among NS of NZSED.

Young AMs seem to be taking the grammatical category further than NS of NZSED by introducing further levels. This might have been an application of supra-local constraints on variation; young AMs introduce more grammatical categories than the NS peers. A point to mention here is that the parents of this subgroup of AMs insisted on getting extra English classes for their children as well as enrolling them at schools that provided (EFL) English classes. Consequently, these children might have been introduced to grammatical distinctions that their peer NS have never taken into consideration.

What is interesting about the internal hierarchy for this constraint is that young AMs in Wellington seem to treat progressives and nominal gerunds alike, both favouring the alveolar nasal. This clustering of the two levels replicates the internal ordering of these two levels found among NS of NZSED; if a functional semantics approach is adopted progressives and gerunds provide very similar information. Nevertheless, this subgroup of AMs is more likely to favour velar nasals for nominal categories, thus diverging from NS of NZSED who slightly favour the alveolar nasal with a factor weight of 58 for nouns. Adjectives are the most favouring of velar realisations amongst young AMs.

The constraint with the second strongest effect is the following segment. Unlike old and middle-aged AMs, young AMs seem to have established that apicals present the context most favouring of alveolar nasals with a factor weight much higher than other following sounds. Like NS of NZSED, apicals favour the alveolar nasals the most with a factor weight of 0.67. Velars also seem to comply with NS norms as they favour the alveolar nasal more than other consonants.

A striking difference from the internal hierarchy among NS of NZSED is that young AM treat following pauses as slightly favouring the alveolar nasal. However, a look at the cross-tabulation of grammatical constraints with the following segment constraints in table 5.10 reveals that 87% of tokens followed by pauses are nominal, this might explain why the factor weight of the following pause for young AMs is hovering around the neutral point, whereas NS have the following pause as the context most favouring of the velar nasal. old and middle-aged AMs seem to find that the following vowels present a complex challenge for them, but younger AMs seem to follow a similar pattern to that of NS with a weaker factor weight of 30.

```
Table 5.10: Cross-tabulation of the grammatical category with the
following segment among young AMs. Application value is the alveolar
nasal
```

Response = /n/					
	Grammatical-category				
Following segment	adj	gerd	nominal	prog	total
Apical	2	5	3	50	60
Other-consonants	4	17	17	88	126
PAUSE	**11**	*7*	**12**	*8*	*38*
Velar	1	4	2	3	10
Vowel	3	18	18	51	90
total	21	51	52	200	324

5.15 Conclusions for (ING) among young AMs

Young AMs in Wellington illustrate weak transfer for the significant constraints on (ING). Specifically, young AMs seem to share the significant constraint on (ING) with NZSED; the following segment and grammatical category. They, however, do not allocate the same strength of effect found among NZSED speakers. young AMs have partially replicated the internal hierarchy for this constraint compared with NS of NZSED.

This means that AMs in Wellington are sensitive to dialect-specific constraints on (ING) variation; since the internal hierarchy, as argued earlier, is dialect-specific. Young AMs acquired the knowledge that progressives and gerunds behave alike and favour the alveolar nasal, while nominal categories favour the alveolar nasal less. The data did not present enough *-thing* compounds, and hence I cannot comment on it.

Young migrants seem to have also acquired the following segment constraint and almost replicate the internal hierarchy, with a decreased strength effect allocated to the constraint, compared with NS of NZSED.

In other words, young AMs have acquired the linguistic competence of NZE and their knowledge allows them to target NS-like variation patterns, but they still seem to have a weak transfer from NZSED. This result is unexpected, especially since that young AMs have acquired English shortly after arrival to Wellington and none of them was over five years of age. So, one would expect them to replicate NS norms. One plausible reason might be that their divergence from NS-patterns is more likely stylistic than it is a result of developing L2 grammars. Young AM are using what was once an L1 transfer effect, strategically, to convey social meanings that are meaningful to them.

5.16 Placing young AMs' patterns of variation on (ING) amongst other 1.5 and Second-generation of NNS in contact settings

Placing the results for (ING) variation for young AMs amongst other NNS-targeting studies of the acquisition of (ING) variation proved to be a harder challenge than that for old and middle-aged AMs. Conceptually, old and middle-aged AMs can be grouped with first-generation migrants and/or NNS in study abroad contexts, based on shared social constraints. They may have high or low proficiency in English, no previous exposure to English or formal education, migrants-mainly networks, mixed or non-migrant mainly networks as well as attitudes towards the host country, its speakers and perceived discrimination.

The young AMs may align with second-generation migrants and/or second-generation migrants that have distinct ethnolects. Since I am only comparing studies with the same variable, (ING), the pool of choice is limited. One study with striking similarities to mine examined the constraints on (ING) across three second-generation ethnic communities in Toronto, Canada. Walker (2012) based on previous research (Hoffman & Walker, 2010) assumes the British descendants to represent the NS baseline for (ING) across Torontonian ethnic minorities. For whom Ethnicity, the preceding segment, grammatical category and gender are the significant constraints on (ING), in that order.

The results for Walker (2012) illustrate that three heritage language speakers (Italian, Chinese and Portuguese) share the significant constraints found in the NS variety but have a different rank ordering, the strength of effect and internal hierarchies. Interestingly, Hoffman and Walker (2010:57) reporting on the same ethnic minorities state that "differences between generation within ethnic groups suggest that language transfer does not persist".

This suggests that heritage language speakers are stylistically using variation patterns as the authors state:

"We interpret the clear and significant differences within ethnic groups as evidence for the weak interpretation of ethnolects. Rather than tracing ethnolects to the imperfect second-language acquisition, substrate transfer, or lack of exposure to mainstream Canadian English, we suggest that a more plausible explanation for the observed effects may be associated with ethnic identity" (ibid:58).

But what is even more interesting is that all second and third generations of NNS created a following segment constraint that is higher in the rank ordering than the preceding segment constraint. Maintaining the same logic for interpreting this result, I would argue that NNS created a following segment constraint on (ING) originally as a substrate effect among first-generation migrants, that is used by Second-generation s as a marker of their ethnic identities.

This pattern of stylistic usage of what was previously a substrate effect is also evident for the young AMs of this study. It is unlikely that they have developing grammars, they are manipulating their linguistic competence to convey indexical meanings that are important to their community.

I conclude that AMs of all age groups, in my study, have patterns of acquisition that can be summarised as a tendency for "transformation under transfer". The different age

groups do, however, have different levels of replication of the NS constraints on variation for the two variables.

     The next chapter will introduce the conclusions for my research, highlighting the contribution of the study to sociolinguistic and variationist research as well as implications for further research.

Chapter 6: Conclusion

This thesis investigated the acquisition of variation by AMs to Wellington, New Zealand. Specifically, it set out to investigate how far AMs replicate patterns of variation found among NS of NZSED, for the variables CSD and (ING).

To achieve this, I first analysed the NZSED to create a solid understanding of NZE that the AMs have most likely had as their target variety. I ended up with a clear description of NS patterns of variation for CSD and (ING) in NZSED, with the focus on "the three lines of evidence" (see page 3).

The results for the NZSED illustrate that NS of this variety share significant constraints on CSD and (ING) with other varieties of NS English. The results also imply that the rank ordering of significant constraints, the associated strength of the effect, the internal hierarchies and sometimes the frequency of occurrence of a variant are very likely to be dialect-specific.

The results also suggest a change in progress in the rank orderings, internal hierarchies of the significant constraints and the frequencies of occurrence for the two variables when I compared the constraints on CSD and (ING) as reported in Bell and Holmes (1992) and Holmes and Bell (1994) with the results for NZSED, collected 20 years later.

I was able to assess how far AMs replicate NS norms on variation based on the benchmark established for the NS norms for the variables. After conducting initial modelling utilising parametric and non-parametric regression tools, it was established that separate models for the AM's age-groups provided a better fit for the data.

The results indicated that AMs are prone to "transformation under transfer" illustrating different degrees of replication of NS norms including instances of weak transfer, strong transfer and/or calquing.

Old and middle-aged AMs, whose social features qualify them as first-generation migrants to Wellington, display strong transfer of NS constraints on CSD and weak transfer of NS constraints on (ING). Whereas young AMs, who qualify as 1.5 generation, display instances of weak transfer of NS norms on the variables.

The strong transfer of NS norms on CSD among old and middle-aged AMs is illustrated as a tendency to replicate NS significant constraints with the same rank orderings and very similar internal hierarchies, only diverging from NZSED in terms of the associated effect size and the frequency of occurrence of the deleted stops i.e., /t,d/.

Nevertheless, the fact that old and middle-aged AMs display strong transfer of NS norms on CSD is an unexpected result, as we would usually suspect that first-generation migrants do not provide NS-like constraints.

Old and middle-aged Arab NNS have replicated the phonological constraints on CSD, whereas they did not illustrate a grammatical category constraint. The question to ask here is, why did old and middle-aged AMs acquire NS phonological constraints but not grammatical ones? Is it because first-generation NNS follow universal constraints on variation including articulatory ones? Or is it because articulatory constraints are perceptually salient and are, therefore, more likely to replicate by NNS? Or is it that certain types of variables, like CSD, are more readily acquired by NNS than others? What does this pattern indicate about NS constraints? Is it possible that NZSED is mainly phonologically-driven for CSD?

The literature, to date, is unclear about whether certain kinds of certain types of variables (including different levels of linguistic structures) or types of constraints (articulatory, cognitive, stylistic) are more or less likely to be associated with NNS replication patterns 1,2,3 mentioned above (Adamson & Regan, 1991; Regan, 1996; Dewaele, 2004; Mougeon et al., 2004; Howard et al., 2006; Labov et al., 2006; Abramowicz,

2007; Kluge et al., 2007; Dinkin, 2008; Hoffman & Walker, 2010; Schleef et al., 2011; Meyerhoff & Schleef, 2014).

Another point to mention is that evidence from NZSED suggest that this variety of NZE may be mainly phonologically-driven for the variable CSD as was suggested by the cross-tabulation conducted for NS that revealed that all grammatical categories favour deletion when the following segment is a consonant, 64% of the time. Even past tense categories, that are usually found to be the least favouring of deletion, seem to favour deletion when the word is followed by a consonant. Consequently, it is possible that there is no grammatical category constraint on CSD among NS of NZSED; hence the lack of this constraint among old and middle-aged AMs.

A second point to tackle is whether CSD is more readily acquired as a variable than other variables like (ING). One way to answer this question is by considering how far old and middle-aged AMs replicate NS constraints on (ING) and compare results and their implications.

The results for (ING) indicate that old and middle-aged AMs replicate the grammatical category constraint on (ING) but not the articulatory constraint, following segment, which is also found to be significant among NS of NZSED for (ING). They, nevertheless, have created a preceding segment constraint on (ING). This suggests that the constraint type (articulatory, grammatical, stylistic) does not determine how far they replicate NS norms, at least for these two variables. It is rather the significant NS constraints and their relative ranking order and strength effects that are relevant.

In other words, old and middle-aged AMs are sensitive to the NS significant constraints and their associated effect size. Accordingly, articulatory constraints on CSD among NS of NZSED that have the highest effect size are acquired by old and middle-aged AMs, as are the grammatical constraints on (ING). Old and middle-aged AMs would

replicate NS norms on variation as per their relative importance vis-à-vis NNS competence in the L2 which dominates how far they can go in the replication of NS norms.

This suggests that first-generation AMs are sensitive to NS significant constraints as well as to the relative strength of effects of such constraints. These constraints include phonological as well as grammatical and stylistic constraints. Accordingly, we might suggest that old and middle-aged AMs acquire NS significant constraints regardless of their type because they are sensitive to dialect-specific norms of NS variation.

Moreover, old and middle-aged AMs seem to be aware of NS social indices as reflected by the close frequencies of CSD and (ING) to NS frequencies of occurrence. They are also able to apply supra-local constraints on variation by filtering their previous exposure to English and tailoring their knowledge to add further levels to NS-constraints, like adding further details to the internal hierarchy of grammatical constraints on (ING).

Old and middle-aged AMs have also created a preceding segment constraint on (ING) that is non-existent in NZSED. This innovative articulatory constraint is most likely the result of a re-syllabification process that is rooted in linguistic transfer from Arabic. This implies that first-generation NNS acquire NS constraints that are valuable to them due to (i) the high size effect for a constraint among NS, (ii) its ability to host L1 transfer and supra-local knowledge across other NS varieties and (iii) NNS proficiency in the L2.

Young AMs, on the other hand, have already gained the linguistic competence of the L2, and I would have expected them to illustrate strong transfer from NZSED, at least for a phonological variable like CSD. The results, nevertheless, illustrate that although young AMs share the same significant constraints found among NS of NZSED for the variable CSD, they have different rank orderings, internal hierarchies and frequencies of occurrence and the patterns of "transformation under transfer" that they display for CSD is weak transfer. These results, although surprising, suggest that young AMs seem to be stylistically manipulating

NS norms of variation to convey social indices. Eckert (2013:97) emphasises the conscious use of variation patterns stylistically to convey certain social meanings; she states:

> "The emphasis on stylistic practice in the third wave places speakers not as passive and stable carriers of dialect, but as stylistic agents, tailoring linguistic styles in ongoing and lifelong projects of self-construction and differentiation. It has become clear that patterns of variation do not simply unfold from the speaker's structural position in a system of production but are part of the active - stylistic - production of social differentiation."

Moreover, the results for CSD and (ING) among young AMs suggest that phonological variables, like CSD, may be more readily used by NNS as a host of substrate effects and ethnolectal marking illustrating the weak transfer of NS norms. Specifically, the preceding segment constraint on CSD may host L1 transfer, which might be used strategically by 1.5 and second-generation migrants to display a variety of identity marking or even the emergence of an ethnolect (cf. Hoffman and Walker (2010)).

Whereas the constraints on (ING) are strongly transferred by 1.5 generation migrants and I propose that the articulatory constraints on the variable (ING) has less room for L1 transfer and therefore are less likely to host the stylistic use of variation patterns as a reflection of identity marking (compared with articulatory constraints on CSD among NNS of English). Nevertheless, young AMs were able to reflect their mixed status between NS of NZSED and first-generation AM by applying supra-local grammatical constraints on (ING) and accordingly illustrating weak transfer of NS constraints on (ING).

These patterns are not specific to AMs; other NNS living in low-linguistically diverse contact settings share similar patterns and set themselves aside from patterns found among young mobile people living in multicultural metropolises.

One of the implications of my study is highlighting the importance of understanding NNS-specific social factors and how various combinations of these result in differential patterns of the acquisition of variation in an L2.For example, some contact settings, especially recently, might be more hostile towards newcomers than others. One may consider Syrian-refugee migration to Europe, especially to geographically close destinations to Syria such as Greece or Italy.

These countries and their people have strong feelings towards the refugees which would probably affect how these refugees handle L2 acquisition including variation. Some countries might even have strict regulations towards migrants like apartheid for example, or refugee camps away from the locals. Other countries like Sweden, New Zealand and Canada have policies that encourage immigration and celebrate diversity. This would most likely affect the migrants' attitudes and consequently the process of acquisition of L2 variation.

Overt positive attitudes towards the host country, the target variety and the target speakers motivate NNS to converge with NS norms. Conversely, overt negative attitudes would most likely motivate NNS to accentuate the differences.

A strong sense of identity including ethnicity and ethnic orientation may also affect the acquisition of variation. In other words, sometimes it is not the host country's reaction towards migrants that matters, but it is how these migrants view themselves in relation to the structure of the host communities and their perceived attitudes. For example, the contrast between Hoffman and Walker (2010)'s findings regarding Chinese NNS in Toronto and how they tend to diverge from the mainstream language variety, and observations drawn about Asians adopting mainstream American English in the USA (Lo & Reyes, 2004; Boberg, 2014) provide further evidence that ethnicity and social isolation cannot be discussed independently of one another and other social factors.

The settlement history of ethnic groups, their settlement patterns and accommodation and work-related practices may also influence how far they replicate NS norms of variation. For example, when migrants cluster in the so-called ethnic neighbourhoods, they are most likely to occupy certain jobs and to maintain closed social networks with other migrants. This is also likely to result in ethnic marking. For example, Jewish and Italian migrants brought up inside ethnic neighbourhoods in Montréal demonstrated ethnic marking in their variety of English (Boberg, 2014).

This project adds to the emerging literature on developments in the quantification of linguistic data (mixed effects modelling, random intercepts, decision trees) and the understanding of how language data calls for different statistical measures than other sciences. I have experimented with the utilisation of non-parametric regression tools to assess their added value to our understanding of variation patterns. I have found that these techniques are helpful particularly to identify instances of interaction across constraints, that may be well missed using typical parametric regression and cross-tabulation techniques.

Considering the findings discussed above, I am planning to carry on my investigation of patterns of "transformation under transfer" for more complex linguistic variables, like morpho-syntactic, semantic and pragmatic ones. This would inform our understanding of patterns of variation across a wider range of variables, and we would, accordingly, understand the true role of variable type versus the role of constraint type on the acquisition of variation.

It would also be interesting to replicate my study with other Arab communities in New Zealand; for example, the well-incorporated Lebanese communities versus the newly arrived Syrian refugee communities. What is interesting about these communities is that the same contact setting, New Zealand, to which the Lebanese willingly migrated, is the

destination of Syrian refugees' forced migration. How these two groups perceive the host country, its language and people is an empirical question.

Within the Lebanese communities in New Zealand, it would also be interesting to look for traces of ethnolects as some evidence for phonetic and prosodic features with an ethnolectal aspect (Clothier, 2014; Cox & Palethorpe, 2005, 2006) have been detected in ethnically and linguistically diverse cities in Australia such as Sydney and Melbourne.

I also plan to try and re-interview Arabs in Wellington, especially that I am now an integrated member of the community. Re-interviewing, my friends and acquaintances, would enable me to tackle some of the issues that I was unable to cover in my thesis. For example, most of my respondents were unwilling to do any of the reading-aloud tasks and I was unable to document style-related NNS constraints across different tasks that are associated with different levels of attention paid to speech (minimal pairs, sentences, and a reading passage). I, therefore, was unable to establish a clear understanding of the NNS style constraints on variation for the two variables and whether there were any indications for L1 transfer and/or reinterpretation of style-shifting, the latter being common among NNS as discussed in chapter 5. I was, accordingly, unable to investigate an interesting proposition: the possibility that AM might hold attitudes that would affect how they perceive the social meaning of variation as well as the non-linguistic constraints governing such variation in the L2.

Research in the acquisition of variation by NNS in contact setting is still lacking, but there is also a promising interest among variationist to walk this untrodden path. I hope that my thesis will contribute to this growing field of knowledge.

References

Abrahamsson, N. (2001). *Acquiring L2 syllable margins : Studies on the simplification of onsets and codas in interlanguage phonology.* (6 Doctoral thesis, comprehensive summary), Centre for Research on Bilingualism, Stockholm Unviersity, Stockholm. Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-33981 DiVA database.

Abrahamsson, N. (2012). Phonological acquisition *The Encyclopedia of Applied Linguistics*: Blackwell Publishing Ltd.

Abramowicz, Ł. (2007). Sociolinguistics meets exemplar theory: Frequency and recency effects in (ing). *University of Pennsylvania Working Papers in Linguistics, 13*(2), 3. Retrieved from http://repository.upenn.edu/pwpl/vol13/iss2/3/

Adamson, H. D., Fonseca-Greber, B., Kataoka, K., Scardino, V., & Takano, S. (1996). Tense marking in the English of Spanish-speaking adolescents. In R. Bayley & D. R. Preston (Eds.), *Second language acquisition and linguistic variation*: John Benjamins

Adamson, H. D., & Regan, V. M. (1991). The Acquisition of community speech norms by Asian immigrants learning English as a second language: A preliminary study. *Studies in Second Language Acquisition, 13*(1), 1-22. doi:10.1017/S0272263100009694

Al-Wer, E. (2007). The formation of the dialect of Amman: from chaos to order. In C. Miller, E. Al-Wer, D. Caubet, & J. C. Watson (Eds.), *Arabic in the city: Issues in dialect contact and language variation*: Routledge.

Al-Wer, E. (2009). *Arabic Dialectology: In Honour of Clive Holes on the Occasion of His Sixtieth Birthday (Studies in Semitic languages and linguistics, v. 53)*: Brill Academic Publishers.

Babel, M., McGuire, G., Walters, S., & Nicholls, A. (2014). Novelty and social preference in phonetic accommodation. *Laboratory Phonology* (Vol. 5, pp. 123).

Bayard, D., & Krishnayya, S. (2001). Gender, expletive use, and context: Male and female expletive use in structured and unstructured conversation among New Zealand university students. *Women and Language, 24*(1), 1-15. Retrieved from https://search.proquest.com/docview/198817432?accountid=27719.

Bayley, R. (1994). Consonant cluster reduction in Tejano English. *Language Variation and Change, 6*(3), 303-326. doi:10.1017/S0954394500001708

Bayley, R. (1996). Competing constraints on variation in the speech of adult Chinese learners of English. In R. Bayley & D. R. Preston (Eds.), *Second language acquisition and linguistic variation* (pp. 97-120) (Vol. 10). Amsterdam, The Netherlands: John Benjamins Publishing Company

Bayley, R. (2005). Second language acquisition and sociolinguistic variation. *Intercultural communication studies, 14*(2), 1-15.

Bayley, R., & Langman, J. (2004). Variation in the group and the individual: Evidence from second language acquisition . *International Review of Applied Linguistics in Language Teaching* (Vol. 42, pp. 303).

Bell, A. (1977). *The language of radio news in Auckland a sociolinguistic study of style, audience & subediting variation.* ResearchSpace@ Auckland.

Bell, A., & Holmes, J. (1992). H-droppin': Two sociolinguistic variables in New Zealand English. *Australian Journal of Linguistics, 12*(2), 223-248. doi:10.1080/07268609208599478

Benor, S. B. (2004). Talmid Chachams and Tsedeykeses : Language, learnedness, and masculinity among Orthodox Jews. *Jewish Social Studies, 11*(1), 147-170. doi:10.1353/jss.2005.0001

Blondeau, H., & Nagy, N. (2008). Subordinate clause marking in Montreal Anglophone French and English. *Social Lives in languages. Sociolinguistics and multilingual*

speech communities. Celebrating the work of Gillian Sankoff (pp. 273-313).

Amsterdam. The Netherlands: John Benjamins

Boberg, C. (2004). Ethnic patterns in the phonetics of Montreal English. *Journal of Sociolinguistics, 8*(4), 538-568. doi:10.1111/j.1467-9841.2004.00273.x

Boberg, C. (2014). Ethnic divergence in Montreal English. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique, 59*(1), 55-82. doi:10.1017/S0008413100000153

Boersma, P., & Weenink, D. (2013). Praat: doing phonetics by computer. Retrieved from http://www.fon.hum.uva.nl/praat/

Britain, D. (2010). Supralocal regional dialect levelling. In C. Llamas & D. Watt (Eds.), *Language and identities* (pp. 193-204). Edinburgh: Edinburgh University Press.

Brown, G., & Yule, G. (1983). *Teaching the spoken language* (Vol. 2): Cambridge University Press.

Brown, L. (2011). *Korean honorifics and politeness in second language learning*. Amsterdam and Philadelphia: John Benjamins.

Brown, R. (1973). *A first language: The early stages*: Harvard University Press.

Bucholtz, M. (1999). "Why be normal?": Language and identity practices in a community of nerd girls. *Language in Society, 28*(2), 203-223. doi:undefined

Campbell-Kibler, K. (2007). Accent, (ING), and the social logic of listener perceptions. *American Speech, 82*(1), 32-64. doi:10.1215/00031283-2007-002

Campbell-Kibler, K. (2012). The Implicit Association Test and sociolinguistic meaning. *Lingua, 122*(7), 753-763. doi:https://doi.org/10.1016/j.lingua.2012.01.002

Cedergren, H. J., & Sankoff, D. (1974). Variable Rules: Performance as a statistical reflection of competence. *Language, 50*(2), 333-355. doi:10.2307/412441

Clark, L., & Schleef, E. (2010). The acquisition of sociolinguistic evaluations among Polish-born adolescents learning English: evidence from perception. *Language Awareness, 19*(4), 299-322. doi:10.1080/09658416.2010.524301

Clothier, J. (2014). *Adapting the Ethnic Orientation Questionnaire for use in a sociophonetic investigation of ethnolectal variation in Melbourne.* Paper presented at the A paper presented at New Ways of Analysing Variation (NWAV) Asia-Pacific 3, Victoria University of Wellington.

Cohen, P., & Labov, W. (1967). *Systematic relations of standard and non-standard rules in the grammars of Negro Speakers.* Paper presented at the 7th project literacy conference

Cox, F., & Palethorpe, S. (2005). *A preliminary acoustic phonetic examination of Lebanese Australian English.* Paper presented at the Australian Language and Speech Conference, Sydney, Australia. Retrieved March.

Cristià, A. (2011). Fine-grained variation in caregivers' /s/ predicts their infants' /s/ category. *The Journal of the Acoustical Society of America, 129*(5), 3271-3280. doi:10.1121/1.3562562

Cristià, A. (2013). Input to language: The phonetics and perception of infant-directed speech. *Language and Linguistics Compass, 7*(3), 157-170. doi:10.1111/lnc3.12015

Cristià, A., & Seidl, A. (2013). The hyperarticulation hypothesis of infant-directed speech. *Journal of Child Language, 41*(4), 913-934. doi:10.1017/S0305000912000669

Davydova, J., Tytus Agnieszka, E., & Schleef, E. (2017). Acquisition of sociolinguistic awareness by German learners of English: A study in perceptions of quotative be like *Linguistics* (Vol. 55, pp. 783).

Dewaele, J.-M. (2004). Retention or omission of the ne in advanced French interlanguage: The variable effect of extralinguistic factors. *Journal of Sociolinguistics, 8*(3), 433-450. doi:10.1111/j.1467-9841.2004.00268.x

Dewaele, J.-M. (2005). The acquisition of sociolinguistic competence in French as a foreign language: an overview. *Journal of French Language Studies, 14*(3), 301-319. doi:10.1017/S0959269504001814

Dinkin, A. J. (2008). The real effect of word frequency on phonetic variation. *University of Pennsylvania Working Papers in Linguistics, 14*(1), 97-106. Retrieved from http://repository.upenn.edu/pwpl/vol14/iss1/8

Drummond, R. (2010). *Sociolinguistic variation in a second language: The influence of local accent on the pronunciation of non-native English speakers living in Manchester.* The University of Manchester. (Doctor of Philosophy), The University of Manchester. Retrieved from https://www.escholar.manchester.ac.uk/uk-ac-man-scw:95944

Drummond, R. (2011). Glottal variation in /t/ in non-native English speech: Patterns of acquisition. *English World-Wide, 32*(3), 280-308. doi:doi:10.1075/eww.32.3.02dru

Drummond, R. (2012). Aspects of identity in a second language: ING variation in the speech of Polish migrants living in Manchester, UK. *Language Variation and Change, 24*(1), 107-133. doi:10.1017/S0954394512000026

Eckert, P. (1989). Jocks and burnouts: Social categories and identity in the high school. New York: Teachers College Press.

Eckert, P. (1996). Vowels and nail polish: the emergence of linguistic style in the preadolescent heterosexual marketplace. In N. Warner, J. Ahlers, L. Bilmes, M. Oliver, S. Wertheim, & M. Chen (Eds.), *Gender and Belief Systems* (Vol. 183, pp. 183-190): Berkeley: Berkeley Women and Language Group.

Eckert, P. (2000). *Language variation as social practice*. Oxford: Wiley-Blackwell.

Eckert, P. (2008a). Variation and the indexical field1. *Journal of Sociolinguistics, 12*(4), 453-476. doi:10.1111/j.1467-9841.2008.00374.x

Eckert, P. (2008b). Where do ethnolects stop? *International Journal of Bilingualism, 12*(1-2), 25-42. doi:10.1177/13670069080120010301

Eckert, P. (2012). Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation. *Annual Review of Anthropology, 41*(1), 87-100. doi:10.1146/annurev-anthro-092611-145828

Eckert, P. (2013). Ethics in linguistic research. *Research methods in linguistics*, 11-26.

Eckert, P., & McConnell-Ginet, S. (1992). Think practically and look locally: Language and gender as community-based practice. *Annual Review of Anthropology, 21*(1), 461-488. doi:10.1146/annurev.an.21.100192.002333

Edwards, J. G.H. (2001). Linguistic constraints on the acquisition of English syllable codas by native speakers of Mandarin Chinese. *Applied Linguistics, 22*(3), 338-365. doi:10.1093/applin/22.3.338

Edwards, J. G. H. (2011). Deletion of /t, d/ and the acquisition of linguistic variation by second language learners of English. *Language Learning, 61*(4), 1256-1301. doi:10.1111/j.1467-9922.2011.00672.x

Edwards, J. G. H. (2016). The deletion of /t, d/ in Hong Kong English. *World Englishes, 37*(2), 138-167. doi:10.1111/weng.12166

Fasold, R. W. (1972). *Tense marking in Black English: A linguistic and social analysis. Urban Language Series, No. 8* (Vol. 8).

Coles, F. A. (1994). *Language maintenance institutions of the Isleño dialect of Spanish. Las jarchas mozárabes: Forma y significado.* Barcelona: Crítica. Retrieved from http://www.jstor.org/stable/4168498

Floccia, C., Delle Luche, C., Durrant, S., Butler, J., & Goslin, J. (2012). Parent or community: Where do 20-month-olds exposed to two accents acquire their representation of words? *Cognition, 124*(1), 95-100. doi:https://doi.org/10.1016/j.cognition.2012.03.011

Forrest, J. (2015). Community rules and speaker behaviour: Individual adherence to group constraints on (ING). *Language Variation and Change, 27*(3), 377-406. doi:10.1017/S0954394515000137

Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics, 34*(4), 409-438. doi:https://doi.org/10.1016/j.wocn.2005.08.002

Galili, T. (2017). R-bloggers. Retrieved from https://www.r-bloggers.com/

Gambino, S. A. (2001). Towards the development of exit criteria for a Newcomer Center: Investigating the relationship between affective variables and the acquisition of English as a second language. *ETD Collection for Pace University*. Retrieved from http://digitalcommons.pace.edu/dissertations/AAI3002143/

Gas, S. (1979). Language transfer and universal grammatical relations. *Language Learning, 29*(2), 327-344. doi:10.1111/j.1467-1770.1979.tb01073.x

Geeslin, K. L., & Long, A. Y. (2014). *Sociolinguistics and second language acquisition: Learning to use language in context*. New York: Routledge.

GhafarGafarSamar, R. (2000). *Aspects of second language speech: A variationist perspective of second language acquisition.* University of Ottawa. Retrieved from http://idochp2.irandoc.ac.ir/FulltextManager/fulltext15/diss/16/16288.pdf

Gordon, E. (1998). The origins of New Zealand speech: The limits of recovering historical information from written records. *English World-Wide, 19*(1), 61-85. doi:doi:10.1075/eww.19.1.05gor

Grunel, P. (1987). *Clinical phonology* (2nd ed.). London: Croom Helm.

Guy, G. R. (1980). Variation in the group and the individual: The case of Final Stop Deletion. In W. Labov (Ed.), *Locating language in time and space* (pp. 1-36): Academic Press.

Guy, G. R. (1991). Explanation in variable phonology: An exponential model of morphological constraints. *Language Variation and Change, 3*(1), 1-22. doi:10.1017/S0954394500000429

Guy, G. R., & Boberg, C. (1997). Inherent variability and the Obligatory Contour Principle. *Language Variation and Change, 9*(2), 149-164. doi:10.1017/S095439450000185X

Guy, G. R., & Boyd, S. (1990). The development of a morphological class. *Language Variation and Change, 2*(1), 1-18. doi:10.1017/S0954394500000235

Guy, G. R., Hay, J., & Walker, A. (2008). Phonological, lexical, and frequency factors in Coronal Stop Deletion in early New Zealand English. *Laboratory Phonology, 11*, 53-54.

Hazen, K. (2006). IN/ING variable. In K. Brown (Ed.), *Encyclopedia of Language and Linguistics* (pp. 5--581): Elsevier.

Hazen, K. (2008). (ING): A Vernacular baseline for English in Appalachia. *American Speech, 83*(2), 116-140. doi:10.1215/00031283-2008-008

Hazen, K. (2011). Flying high above the social radar: Coronal Stop Deletion in modern Appalachia. *Language Variation and Change, 23*(1), 105-137. doi:10.1017/S0954394510000220

Hoffman, M. F. (2004). *Sounding Salvadorean: Phonological variables in the Spanish of Salvadorean youth in Toronto*: University of Toronto.

Hoffman, M. F., & Walker, J. A. (2010). Ethnolects and the city: Ethnic orientation and linguistic variation in Toronto English. *Language Variation and Change, 22*(1), 37-67. doi:10.1017/S0954394509990238

Holes, C. (1993). *The uses of variation: A study of the political speeches of Gamal Abd Al-Nasir.* Paper presented at the Perspectives on Arabic Linguistics: Papers from the Fifth Annual Symposium on Arabic Linguistics.

Holmes, J., & Bell, A. (1994). *Consonant Cluster Reduction in New Zealand English.* Retrieved from

Holmes, J., Bell, A., & Boyce, M. (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*: Department of Linguistics, Victoria University.

Holmes, J., Roberts, M., Verivaki, M., & Aipolo, A. (1993). Language maintenance and shift in three New Zealand speech communities. *Applied Linguistics, 14*(1), 1-24. doi:10.1093/applin/14.1.1

Hornsby, D. (2007). Regional dialect levelling in urban France and Britain. *Nottingham French Studies, 46*(2), 64-81. doi:10.3366/nfs.2007-2.005

Houston, A. C. (1985). *Continuity and change in English morphology: The variable (ING).* (PhD dissertation), University of Pennsylvania.

Howard, M., LemÉE, I., & Regan, V. (2006). The L2 acquisition of a phonological variable: the case of /l/ deletion in French. *Journal of French Language Studies, 16*(1), 1-24. doi:10.1017/S0959269506002298

Huddleston, R., & Pullum, G. K. (2002). The Cambridge grammar of English. *Language. Cambridge: Cambridge University Press*.

Hymes, D. (1972). Models of the interaction of language and social Life. In J. Gumpertz & D. Hymes (Eds.), *Directions in sociolinguistics: The ethnography of communication* (pp. 35-71). New York: Holt, Rinehart and Winston.

Immigration New Zealand. (2017). Statistics. Retrieved from

    https://www.immigration.govt.nz/about-us/research-and-statistics/statistics

Jaspaert, K., & Kroon, S. (1993). Methodological issues in language shift research. In G.

    Extra & L. Verhoeven (Eds.), *Immigrant languages in Europe* (pp. 297-308).

    Clevedon: Multilingual Matters.

Jiang, M., Green, R. J., Henley, T. B., & Masten, W. G. (2009). Acculturation in relation to

    the acquisition of a second language. *Journal of Multilingual and Multicultural*

    Development, 30(6), 481-492. doi:10.1080/01434630903147898

Johnson, D. (2009). Getting off the GoldVarb standard: Introducing Rbrul for Mixed-Effects

    Variable Rule Analysis. *Language and Linguistics Compass, 3*(1), 359-383.

    doi:10.1111/j.1749-818X.2008.00108.x

Kachru, B. B. (1982). South Asian English. *English as a world language, 4*.

Kachru, B. B. (1992). *The other tongue: English across cultures*: University of Illinois Press.

Kendall, T. (2010). *Accommodating (ING): individual variation in mixed-ethnicity*

    interviews. Paper presented at the HESELWOOD, Barry; UPTON, Clive. Proceedings

    of the Thirteenth International Conference on Methods in Dialectology (Methods

    XIII). Frankfurt: Peter Lang.

Kern, F. (2011). From learner varieties to ethnic styles. In F. Kern & M. Selting (Eds.),

    Ethnic styles of speaking in European metropolitan areas (Vol. 8, pp. 1-18).

    Amsterdam and Philadelphia: John Benjamins Publishing.

Kerswill, P., & Williams, A. N. N. (2000). Creating a New Town koine: Children and

    language change in Milton Keynes. *Language in Society, 29*(1), 65-115.

    doi:undefined

Kluge, D. C., Rauber, A. S., Reis, M. S., & Bion, R. A. H. (2007). *The relationship between*

    the perception and production of English nasal codas by Brazilian learners of

English. Paper presented at the Eighth Annual Conference of the International Speech Communication Association.

Labov, W. (1963). The social motivation of a sound change. *<i>WORD</i>, 19*(3), 273-309. doi:10.1080/00437956.1963.11659799

Labov, W. (1966). The social stratification of English in New York City. *Center for Applied Linguistics, Washington DC*.

Labov, W. (1972). *Sociolinguistic patterns*: University of Pennsylvania Press.

Labov, W. (1975). The quantitative study of linguistic structure. In K. Dahlstedt (Ed.), *The Nordic languages and modern linguistics.* (pp. 188-244). Stockholm: Almqvist & Wiksell.

Labov, W. (1984). Field methods of the project on linguistic change and variation. In J. Baugh & J. Sherzer (Eds.), *Language in use : readings in sociolinguistics* (pp. 28-53). Englewood Cliffs, N.J.: Prentice-Hall.

Labov, W. (1989). The child as a linguistic historian. *Language Variation and Change, 1*(1), 85-97. doi:10.1017/S0954394500000120

Labov, W. (1997). Resyllabification. In F. Hinskens, R. Van Hout, & L. Wetzels (Eds.), *Variation, change and phonological theory* (Vol. 146, pp. 145–179). Amsterdam and Philadelphia: John Benjamins Publishing.

Labov, W. (2001). *Principles of linguistic change, social factors*. Malden, Oxford and Basil: Blackwell.

Labov, W. (2011). *Principles of linguistic change, cognitive and cultural factors* (Vol. 3): John Wiley & Sons.

Labov, W. (2014). The sociophonetic orientation of the language learner. In C. Celata & S. Calamai (Eds.), *Advances in sociophonetics* (pp. 17-29). Amsterdam and Philadelphia: John Benjamins.

Labov, W., Ash, S., Baranowski, M., Nagy, N., Ravindranath, M., & Weldon, T. (2006). Listeners' sensitivity to the frequency of sociolinguistic variables. *University of Pennsylvania Working Papers in Linguistics, 12*(2), 105-129.

Laturnus, R., Vilchez, N. P. d., & Guy, R. C. a. G. (2016). .Dialect, priming, and frequency effects on (-ING) variation in English. *New ways of analyzing variation 45*

Léglise, I., & Chamoreau, C. (2013). *The interplay of variation and change in contact settings*: John Benjamins

Levon, E., & Buchstaller, I. (2015). Perception, cognition, and linguistic structure: The effect of linguistic modularity and cognitive style on sociolinguistic processing. *Language Variation and Change, 27*(3), 319-348. doi:10.1017/S0954394515000149

Lim, L. T., & Guy, G. R. (2005). The limits of linguistic community: speech styles and variable constraint effects. *University of Pennsylvania Working Papers in Linguistics, 10*(2), 13.

Linguistic Data Consortium (LDC). (2017). How LDC Data Inspires Research. Retrieved from https://www.ldc.upenn.edu/

Lo, A., & Reyes, A. (2004). Language, identity and relationality in Asian Pacific America. *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA), 14*(2), 115-125. Retrieved from http://journals.linguisticsociety.org/elanguage/pragmatics/article/download/429/429-723-1-PB.pdf

Lybeck, K. (2002). Cultural Identification and second language pronunciation of Americans in Norway. *The Modern Language Journal, 86*(2), 174-191. doi:10.1111/1540-4781.00143

Maclagan, M., & Hay, J. (2010). Sociolinguistics in New Zealand In M. J. Ball (Ed.), *The Routledge Handbook of Sociolinguistics Around the World* (pp. 159-169). USA and Canada: Routledge.

Major, R. C. (2004). Gender and stylistic variation in second language phonology. *Language Variation and Change, 16*(3), 169-188. doi:10.1017/S0954394504163059

Mendoza-Denton, N. (2008). *Homegirls : Language and cultural practice among Latina youth gangs*.

Mendoza-Denton, N. (2014). *Homegirls: Language and cultural practice among Latina youth gangs*: John Wiley & Sons.

Mendoza-Denton, N. (1996). 'Muy Macha': Gender and ideology in gang-girls' discourse about makeup. *Ethnos, 61*(1-2), 47-63. doi:10.1080/00141844.1996.9981527

Meyerhoff, M. (2002). Communities of practice. In Chambers, J.K, & Trudgill, P & Schilling-Estes, N (eds) The Handbook of Language Variation and Change , Blackwell Publishing Ltd.

Meyerhoff, M. (2009a). Replication, transfer, and calquing: Using variation as a tool in the study of language contact. *Language Variation and Change, 21*(3), 297-317. doi:10.1017/S0954394509990196

Meyerhoff, M. (2009b). Sociolinguistic variation and change. *Linguistics-Encyclopedia of Life Support Systems* (pp. 202-224).

Meyerhoff, M., Schleef, E., & Clark, L. (2009c). *Becoming local: Exploring adolescents' sociolinguistic limits and potential.* Paper presented at the British Association of Applied Linguistics (BAAL).

Meyerhoff, M., & Niedzielski, N. (2003). The globalisation of vernacular variation. *Journal of Sociolinguistics, 7*(4), 534-555. doi:10.1111/j.1467-9841.2003.00241.x

Meyerhoff, M., & Schleef, E. (2012). Variation, contact and social indexicality in the acquisition of (ing) by teenage migrants1. *Journal of Sociolinguistics, 16*(3), 398-416. doi:10.1111/j.1467-9841.2012.00535.x

Meyerhoff, M., & Schleef, E. (2014). Hitting an Edinburgh Target: Immigrant Adolescents' Acquisition of Variation in Edinburgh English. In R. Lawson (Ed.), *Sociolinguistics in Scotland* (pp. 103-128). London: Palgrave Macmillan UK.

Meyerhoff, M., Schleef, E., & MacKenzie, L. (2015). *Doing sociolinguistics: A practical guide to data collection and analysis*: Routledge.

Miller, C., & Caubet, D. (2009). Arabic sociolinguistics in the Middle East and North Africa (MENA). In M. Ball (Ed.), *Handbook of Sociolinguistics Around the World* (pp. 238-256): The Routledge.

Milroy, L. (1980). *Language and social networks*: Wiley.

Milroy, L., & Gordon, M. (2003). *Sociolinguistics: Method and interpretation*. Malden and Oxford John Wiley & Sons.

Ministry of Culture and Heritage of the New Zealand Government. (2014). *Te Ara: The Encyclopedia of New Zealand*.

Ministry of Business, I. a. E. (2017, 25 November 2016). Trends and outlook 2015/16. Retrieved from http://www.mbie.govt.nz/info-services/immigration/migration-research-and-evaluation/trends-and-outlook/2015-16

Moore, E., & Podesva, R. (2009). Style, indexicality, and the social meaning of tag questions. *Language in Society, 38*(4), 447-485. doi:10.1017/S0047404509990224

Mougeon, R., & Nadasdi, T. (1998). Sociolinguistic discontinuity in minority language communities. *Language, 74*(1), 40-55. doi:10.2307/417564

Mougeon, R., Nadasdi, T., & Rehner, K. (2005). Contact-induced linguistic innovations on the continuum of language use: The case of French in Ontario. *Bilingualism: Language and Cognition, 8*(2), 99-115. doi:10.1017/S1366728905002142

Mougeon, R., Rehner, K., & Nadasdi, T. (2004). The learning of spoken French variation by immersion students from Toronto, Canada. *Journal of Sociolinguistics, 8*(3), 408-432. doi:10.1111/j.1467-9841.2004.00267.x

Nagy, N., Blondeau, H., & Auger, J. (2003). Second language acquisition and "real" French: An investigation of subject doubling in the French of Montreal Anglophones. *Language Variation and Change, 15*(1), 73-103. doi:10.1017/S0954394503151034

Nagy, N., & Meyerhoff, M. (2015). Extending ELAN into variationist sociolinguistics. *Linguistics Vanguard* (Vol. 1, pp. 271).

Neu, H. (1980). Ranking of constraints on/t, d/deletion in American English: A statistical analysis. *Locating language in time and space, 37*.

New Zealand Statistics. (2013, 14 July 2015). 2013 Census. Retrieved from http://www.stats.govt.nz/Census/2013-census.aspx?gclid=Cj0KCQjw9uHOBRDtARIsALtCa940PjiUQWkIr22WW_aK35-aO_ARzifp9P0M6BkaE8LxVwqpGbDmYK8aAloTEALw_wcB

Ochs, E. (1992). Indexing gender. In A. Duranti & C. Goodwin (Eds.), *Rethinking context: Language as an interactive phenomenon* (pp. 335). Cambridge, New York and Melbourne.

Okamoto, S. (2014). The use and interpretation of "regional" and "standard" variants in Japanese conversation. In K. Kabata & T. Ono (Eds.), *Usage-based approaches to Japanese grammar* (pp. 279-303): John Benjamins.

Patrick, P. L. (1991). Creoles at the intersection of variable processes: -t,d deletion and past-marking in the Jamaican mesolect. *Language Variation and Change, 3*(2), 171-189. doi:10.1017/S095439450000051X

Patrick, P. L. (1999). *Urban Jamaican Creole: variation in the mesolect*: John Benjamins Publishing.

Podesva, R. J. (2007). Phonation type as a stylistic variable: The use of falsetto in constructing a persona1. *Journal of Sociolinguistics, 11*(4), 478-504. doi:10.1111/j.1467-9841.2007.00334.x

Podesva, R. J. (2010). Constraints on the social meaning of variation: The case of Released /t/: Summer School of Sociolinguistics. The University of Edinburgh.

Podesva, R. J., Roberts, S. J., & Campbell-Kibler, K. (2002). Sharing resources and indexing meanings in the production of gay styles. *Language and sexuality: Contesting meaning in theory and practice*, 175-189. Retrieved from http://www.ling.ohio-state.edu/~campbell-kibler.1//Podesva_ea02.pdf

Poplack, S., & Tagliamonte, S. (2001). *African American English in the diaspora*: Wiley-Blackwell.

R Development Core Team. (2008). R: a language and environment for statistical computing. Retrieved from www.R-project.org

R Development Core Team. (2013). R: a language and environment for statistical computing. Retrieved from www.R-project.org

Regan, V. (1996). Variation in French interlanguage: A longitudinal study of sociolinguistic competence. In R. Bayley & D. R. Preston (Eds.), *Second language acquisition and linguistic variation* (pp. 177-201). Amsterdam and Philadelphia John Benjamins

Roberts, J. L. (1994). *Acquisition of variable rules:(-t, d) deletion and (ing) production in preschool children.* The university of Pennsylvania.

Roberts, J. L. (1997). Acquisition of variable rules: a study of (-t, d) deletion in preschool children. *Journal of Child Language, 24*(2), 351-372. doi:undefined

Romaine, S. (2003). Variation in language and gender. In J. Holmes & M. Meyerhoff (Eds.), *The handbook of language and gender* (pp. 98-118). Malden, Oxford and Melbourne Blackwell.

Sankoff, D. (1978). *Linguistic variation: Models and methods*: Academic Press.

Sankoff, D. (1988). Variable rules. In A. Ulrich, N. Dittmar, & K. J. Mattheier (Eds.), *Sociolinguistics* (pp. 984-997). Berlin: Walter de Gruyter.

Sankoff, D., Tagliamonte, S., & Smith, E. (2005). Goldvarb X. Computer program. Department of Linguistics, University of Toronto, Canada.

Sankoff, G., & Blondeau, H. (2007). Language change across the lifespan: /r/ in Montreal French. *Language: Journal of the Linguistic Society of America, 83*(3), 560-588. doi:10.1353/lan.2007.0106

Santa Ana A, O. (1996). Sonority and syllable structure in Chicano English. *Language Variation and Change, 8*(1), 63-89. doi:10.1017/S0954394500001071

Santa Ana A., O. (1991). *Phonetic simplification processes in the* English of the Barrio: A cross-generational sociolinguistic study of the Chicanos of Los Angeles.*

Sayahi, L. (2014). *Diglossia and language contact: Language variation and change in North Africa*: Cambridge University Press.

Schleef, E., Meyerhoff, M., & Clark, L. (2011). Teenagers' acquisition of variation: A comparison of locally-born and migrant teens' realisation of English (ing) in Edinburgh and London. *English World-Wide, 32*(2), 206-236. doi:doi:10.1075/eww.32.2.04sch

Schleef, E. (2013a). *Developmental sociolinguistics and the acquisition of T-glottalling by immigrant teenagers in London*. Paper presented at the International Conference on

Language Variation in Europe and the Variation in Language Acquisition Workshop in 2012. http://www.erikschleef.eu/wp-content/uploads/2012/08/Tdevelopm_Schleef.pdf

Schleef, E. (2013b). Migrant teenagers' acquisition of sociolinguistic variation: The variables (ing) and (t). In P. Auer, J. Caro, & G. Kaufmann (Eds.), *Language variation – European perspectives IV* (pp. 201-213). Amsterdam: John Benjamins.

Schleef, E., & Turton, D. (2016). Sociophonetic variation of like in British dialects: effects of function, context and predictability. *English Language and Linguistics*, 1-41. doi:10.1017/S136067431600023X

Schleef, E. (2017a). Developmental sociolinguistics and the acquisition of T-glottalling by immigrant teenagers in London. In G. d. Vogelaer & M. Katerbow (Eds.), *Variation in language acquisition.* (pp. 311-347). Amsterdam: John Benjamins.

Schleef, E. (2017b). Social Meanings across Listener Groups. *Journal of English Linguistics, 45*(1), 28-59. doi:10.1177/0075424216686149

Schleef, E., Flynn, N., & Barras, W. (2017c). Regional diversity in social perceptions of (ing). *Language Variation and Change, 29*(1), 29-56. doi:10.1017/S0954394517000047

Schreier, D. (2003). Convergence and language shift in New Zealand: Consonant cluster reduction in 19th Century Maori English. *Journal of Sociolinguistics, 7*(3), 378-391. doi:10.1111/1467-9481.00230

Schreier, D. (2005). *Consonant change in English worldwide: Synchrony meets diachrony*: Palgrave Macmillan.

Sclafani, J. (2009). Martha Stewart behaving Badly: Parody and the symbolic meaning of style1. *Journal of Sociolinguistics, 13*(5), 613-633. doi:10.1111/j.1467-9841.2009.00427.x

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics in Language Teaching* (Vol. 10, pp. 209).

Sharma, D., & Sankaran, L. (2011). Cognitive and social forces in dialect shift: Gradual change in London Asian speech. *Language Variation and Change, 23*(3), 399-428. doi:10.1017/S0954394511000159

Shopen, T. (1978). Research on the variable (ING) in Canberra, Australia. *Talanya, 5*, 42-52.

Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language & Communication, 23*(3), 193-229. doi:https://doi.org/10.1016/S0271-5309(03)00013-2

Smith, J., Durham, M., & Fortune, L. (2009). Universal and dialect-specific pathways of acquisition: Caregivers, children, and t/d deletion. *Language Variation and Change, 21*(1), 69-95. doi:10.1017/S0954394509000039

Suleiman, Y. (2011). *Arabic, self and identity: A study in conflict and displacement*: Oxford University Press.

Tagliamonte, S. (2002). Comparative sociolinguistics. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The Handbook of Language Variation and Change* (pp. 729-763): Blackwell

Tagliamonte, S. (2004). *Somethi[N]'s Goi[n] On! Variable (ing) at ground zero.Language Variation in Europe.* Paper presented at the Language variation in Europe: Papers from the Second International Conference on Language Variation in Europe, ICLaVE 2, Uppsala University, Sweden, June 12-14,2003, Uppsala University Department of Scandinavian Languages.

Tagliamonte, S. (2006). *Analysing sociolinguistic variation*: Cambridge University Press.

Tagliamonte, S. (2011). Variation as a window on universals. In P. Siemund (Ed.), *Linguistic Universals and Language Variation* (pp. 128-168): Walter de Gruyter.

Tagliamonte, S. (2012). *Variationist sociolinguistics: Change, observation, interpretation* (Vol. 40): John Wiley & Sons.

Tagliamonte, S., & Temple, R. (2005). New perspectives on an ol' variable: (t,d) in British English. *Language Variation and Change, 17*(3), 281-302. doi:10.1017/S0954394505050118

Tagliamonte, S. A., & Baayen, R. H. (2012). Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change, 24*(2), 135-178. doi:10.1017/S0954394512000129

Tamminga, M. J. (2014). *Persistence in the production of linguistic variation.* (Dissertation), University of Pennsylvania. Retrieved from http://repository.upenn.edu/edissertations/1466/

Te Ara: The encyclopaedia of New Zealand. (2014). https://teara.govt.nz/en/history

Temple, R. (2003). *What is '-t, d deletion'? A new look at an old sociolinguistic variable.* Paper presented at the Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS-15), Barcelona, Spain.

Torgersen Eivind, N., Gabrielatos, C., Hoffmann, S., & Fox, S. (2011). A corpus-based study of pragmatic markers in London English *Corpus Linguistics and Linguistic Theory* (Vol. 7, pp. 93).

Torgersen, E. N., & Szakay, A. (2012). An investigation of speech rhythm in London English. *Lingua, 122*(7), 822-840. doi:https://doi.org/10.1016/j.lingua.2012.01.004

Trudgill, P. (1974). *The social differentiation of English in Norwich*: Cambridge University Press.

Uritescu, D., Mougeon, R., & Handouleh, Y. (2002). Le comportement du schwa dans le français parlé par les élèves des programmes d'immersion française. In A. Baudot & C. Tatilon (Eds.), *La Linguistique Fonctionnelle Au Tournant Du Siècle: Actes Du*

Vingt-quatrième Colloque International de Linguistique Fonctionnelle, Collège Universitaire Glendon, Université York, Toronto (Ontario) Canada, 23-27 Juin 2000 (pp. 335-346): Editions du Gref.

Uritescu, D., Mougeon, R., Rehner, K., & Nadasdi, T. (2004). Acquisition of the internal and external constraints of variable schwa deletion by French immersion students *IRAL - International Review of Applied Linguistics in Language Teaching* (Vol. 42, pp. 349).

Wagner, S. E. (2008). *Linguistic change and stabilization in the transition from adolescence to adulthood.* The university of Pennsylvania.

Walker, J. (2012). Velar nasals in Toronto English. New Ways of Analyzing Variation 41. Indiana University, Bloomington.

Warren, P. (2002). NZSED: Building and using a speech database for New Zealand English. *New Zealand English Journal, 16*, 53-58. Retrieved from http://search.informit.com.au/documentSummary;dn=596453763969271;res=IELNZC

Wells, J. C. (1982). *Accents of English: Beyond the British Isles* (Vol. 3): Cambridge University Press.

Wheelerm, M. W. (2005). Cluster reduction: Deletion or coalescence? *Catalan Journal of linguistics, 4*, 57-82.

Wiese, H. (2009). Grammatical innovation in multiethnic urban Europe: New linguistic practices among adolescents. *Lingua, 119*(5), 782-806. doi:https://doi.org/10.1016/j.lingua.2008.11.002

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). *ELAN: a professional framework for multimodality research.* Paper presented at the Proceedings of LREC.

Wolfram, W. (1984). Unmarked Tense in American Indian English. *American Speech, 59*(1), 31-50. doi:10.2307/454992

Wolfram, W. (1985). Variability in tense marking: A case for the obvious. *Language Learning, 35*(2), 229-253. doi:10.1111/j.1467-1770.1985.tb01026.x

Wolfram, W., & Christian, D. (1979). *Dialogue on dialects*: Center for applied linguistics.

Wolfram, W. A. (1969). A Sociolinguistic description of Detroit Negro Speech. Urban Language Series, No. 5 (pp. 237): Center for Applied Linguistics.

Woods, H. (1978). *A socio-dialectal survey of the English spoken in Ottawa: A study of sociolinguistic and stylistic variation.* The university of Ottawa.

Young-Scholten, M., & Archibald, J. (2000). Second language syllable structure. In J. Archibald (Ed.), *Second language acquisition and linguistic theory* (pp. pp. 64-97): Oxford: Blackwell.

Young, R. (1991). *Variation in interlanguage morphology*. New York: Peter Lang

Zentella, A. C. (1997). *Growing up Bilingual: Puerto Rican children in New York*. Oxford: Wiley-Blackwell.

Zoom North America. (2017). The Zoom H1 Handy Recorder.   Retrieved from https://www.zoom-na.com/products/field-video-recording/field-recording/zoom-h1-handy-recorder

Appendix A

Transcription protocol

1. Orthography and spelling:

I used the standard orthographic and spelling patterns of English, and I have double-checked for words I was not sure of, e.g. slang words like "pissed" for very drunk in New Zealand English. There were instances, in the NNS's corpus, where the respondents used words that are phonetically similar to another word, but that is irrelevant in the context of the conversation, so I kept these as uttered by the speakers.

2. Punctuation:

I used standard punctuation as it normally appears in writing, with no additional spaces around the punctuation marks. The acceptable marks in transcription protocol are periods, exclamation marks and question marks at the end of a sentence, and commas within a sentence.

3. Abbreviations and Acronyms:

I only used abbreviations that were used by the respondents, and I wrote Acronyms in caps.

4. Contractions:

I used contractions only when produced by speakers and there were instances of nonstandard forms like "*donno*", "*gonna*", "*wanna*".

5. Numbers:

All numerals are written out as complete words. Hyphenation is used for numbers between twenty-one and ninety-nine only.

6. Hyphenated words and compounds

I used hyphenation in compounds when required, in cases where there is a choice between writing a compound word as one word, a hyphenated word, or as two words with spaces in between them, I used hyphenation.

7. Disfluent speech:

Repetitions and false beginnings: single hyphen

Partial words: double hyphen after the written part of the word.

Incomprehensible words: I used the abbreviation "inc."

Pauses: three dots for audible pauses "…"

Paralinguistic intervention: such as *huh, er, uh, mm*, etc. are also indicated in standardized orthography.

Appendix B

Consent form

SCHOOL OF LINGUISTICS AND APPLIED LANGUAGE STUDIES

Project title: The acquisition of variation by Arab migrants to Wellington.
Researcher: Rania Zarour

I have read the Participant Information sheet, and I understand the intent and nature of the research project. I understand that my participation is strictly voluntary. I have had the opportunity to ask and receive satisfactory questions about the research. I understand that I may choose to withdraw from the study at any time within one year of this date, and to change the conditions under which my recording and transcript are used for future projects at any time; otherwise, I understand that the data will be stored in a secure location indefinitely.

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

☐ I agree to participate in this interview, either individually or as part of a small group (please cross out whichever does not apply).
☐ I am aware that our conversation is being digitally recorded, and that I have the right to request erasure of any portion of the recording that I am uncomfortable with or to withdraw from the study at any time.
☐ I grant you permission to use the interview material for any academic purposes relating to your doctoral research, such as discussions, presentations, teaching and any published or unpublished works. I understand that my confidentiality will be maintained always.
☐ I understand that I have one (1) year from this date to withdraw my participation from the doctoral research project.
☐ I grant you permission to keep and use my recording and transcript as part of a larger collection of sociolinguistic data, with the understanding that my confidentiality will be maintained always.
☐ I understand that I may, at any time, withdraw my participation from any future project not already underway.
☐ I grant you permission to share anonymized versions of my recording and transcript with other co-investigators in future collaborative linguistic studies.

_____   _____
Signature of participant   Signature of researcher

_____   _____
Date   Date

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:

Name (for use
in
correspondence)   _____

Postal address   _____

_____

Appendix C

Participant information sheet

SCHOOL OF LINGUISTICS AND APPLIED LANGUAGE STUDIES

Participant information sheet

Project title: The acquisition of variation by Arab migrants to Wellington.
Researcher: Rania Zarour

Introduction

My name is Rania Zarour; I am a PhD candidate at the School of Linguistics and Applied Language Studies at Victoria University of Wellington. I am conducting a research about the acquisition of New Zealand English by Arabic speakers living in Wellington. The progress of my research is supervised by Professor Miriam Meyerhoff and Professor Janet Holmes whose contact information is provided at the end of this sheet.
I would like to ask you to help me in my research by allowing me to make an audio recording of you in two simple tasks: reading a short passage and having a short conversation up to an hour. I may combine your results and other people's results in an analysis of acquisition of New Zealand English.

Procedures
Participation is strictly voluntary. If you agree to participate, you will be committing to one digitally-recorded (audio-only) conversation lasting between one and two hours. We can do the recording wherever is most convenient for you – your home, my home, somewhere on campus, somewhere else – so long as it's quiet. The conversation will be in two parts: the first will be a casual chat about whatever you feel like talking about; the second (and generally shorter) part will be reading a short passage.

Your confidentiality is important to me, so your recording will be assigned a number and a pseudonym, and these are the only identifiers that will be used in referring to the recording. I will be transcribing the interviews, and then anonymizing both the recording and the transcript – that is, I'll take out anything that might identify you as the speaker. I'll keep one unaltered copy of the recording as a backup, but I will be the only person with access to it, and the research itself will be conducted using the anonymized versions.

If you agree to participate and then later change your mind, that's okay – you're free to withdraw from the project at any time up to one year from the date of recording. After that point, it may be difficult to remove your data from any analyses that have already been carried out, but you will always be welcome to indicate that you do you not want your data

used in <u>future</u> projects. Similarly, you're also free to stipulate that certain portions of the recording be either deleted or not used in subsequent analysis.

Use of the data
Your recording and transcript will be used in both will primarily be used for my doctoral thesis, and any excerpts that find their way into either the written thesis or any associated presentations will be anonymized and attributed only to your pseudonym. I am the primary analyst for this project, but my supervisor and co-supervisor will also have access to the recordings and the transcripts.

Because I'm building a collection of sociolinguistic data, your recording may be used in more than one project. This means that your recording may be shared with co-investigators in future research projects; if this is the case, these researchers will only have access to the anonymized versions of the recording and transcript, and they will only know your pseudonym and some basic demographic information about you (such as your age, your gender, and where you grew up). If you prefer, you also have the option of stipulating that I am the only researcher who can access your recording or transcript at all. What this means is, unless you explicitly state otherwise, your recording and transcript will be kept indefinitely as part of a sociolinguistic data set, with all confidentiality precautions maintained for the lifetime of the archive. It will be stored as digital files on an external hard drive, which will be stored securely in a locked cabinet always. Note that, as new storage media are developed, the files may be changed to a new format, but will still be kept secure.

If you would like to have a copy of your recording or your transcript, I am happy to provide them to you. Likewise, if you would like to be kept up to date with either my doctoral thesis or any other publications that arise from the use of your data, I am happy to share these with you. Any address or contact information that you provide me will be used for the sole purpose of transmitting information to you and will not be shared with anyone else.

Contact details
If you have any questions, comments, concerns or thoughts about the project, please don't hesitate to get in touch with me:

P.S the contact details have been must be redacted from the electronic version of the thesis so

that it can be accepted into the Research Archive for public access.

Appendix D: specific social factors for AM in wellington

age group AM	age at arrival to NZ	Time spent in NZ	Intention to stay in NZ	Self reported proficiency	Self-reported affiliation	Ties with Arabic culture	Network Involvement	Occupation	Experience in English
Young AM 16-30	Less than 5 years of age	Since arrival	yes	NS	kiwi	Yes Member of CoFP in Wellington Has visited home at least once Maintains ties overseas	Mainly non-Arab	Student	Acquired English integratively in New Zealand. Were enrolled at ESL classes at school. Went to Sunday school/English classes Had playdates with NS of NZE on regular bases
Middle-aged AM 31-49	adult	9 months-less than 3 years	yes	Very good	Arab-new Zealander	Yes Member of CoFP in Wellington Maintains ties overseas	Mainly non-Arab	Employed	Received direct instruction to English Were enrolled at ESL classes offered by the government through several NGOs.
Old AM Over 50	adult	More than 15 years	yes	Good to excellent	Arab-new Zealander	Yes Member of CoFP in Wellington Maintains ties overseas	Mainly-Arab	Stay at home and retired	Received direct instruction to English Were enrolled at ESL classes offered by the government through several NGOs.

Appendix E

Extracts from AM's conversations

1. <u>Young AM who arrived in NZ when he was almost six years old</u>

"I play on the defense and you know after that they usually all go out to drink some alcohol, I go and drink a coke of course yeah"

"sometimes yeah I don't get that many just because it quite the furthest and so we were just like you know its half way across the world it has nothing to do with us… are just Australia awesome Australia that's all we care about".

2. <u>Young AM who arrived in NZ when he was two years old</u>

"they say I look kind of Kiwi….coz ha yeah when I was overseas they were like…oh you look so foreign you know you brothers can pass as being here but the way you dress the hair you know just the way you look is foreign and it's just a giveaway….but now that I have seen New Zealand and like lived up in new Zealand…..I can't you know I just can't go back there.. just what if you are I will go back there for a holiday…but to live there I don't think so ha-ha".

3. <u>Young AM who arrived in NZ when she was four years old</u>

"you know we get the most traditional people bring in their culture and their morals here, than I think back there…I feel like it's too much honestly…i think it's just yeah, I think that's that , cause I feel like , their, in their own environment , they feel as much eased up, and here you have to inflict all these religion and morals and values and all that, that…that you have to be like that, and I understand that, I fully respect that.. cause if I have kids, I 'll be exactly like that…honestly, I'll tell my daughter t, I'll be like, one two three and four, accept it accept it, don't accept it, I can I will have to teach you, you know".

"you can't help it, and at the end of the day, you want someone from your culture".

"you know, religiously, I'm there, I feel like, I'm confident within myself"

4. <u>Middle-aged male AM who have lived in NZ for seventeen years.</u>

"about fourteen hours to reach uhm fourteen hours Dubai yeah…I mean with stops maybe eighteen hours to reach the f the eastern middle east…so, once you\re here, you usually start getting holidays once every three four, five years. Instead of every year…"

"uh it's, it's much nicer than living in a single homogenized sort of uh city or country…uh for my taste at least".

5. <u>Middle-aged female AM who has been in NZ for ten years on and off.</u>

"oh, it's nice I like it yeah, so I love New Zealand…i feel like this is my real home it's my home…i have always felt that I do belong here I love Jordan…i love Jordan but see I, I feel like ummm…ha ha-ha yeah I just love New Zealand and I love wellington".

6. <u>Middle-aged female AM who has been in NZ for twenty years.</u>

"so, I didn't really mix with kiwis because they were not very open…cause twenty years ago, it was, you know, because of the language difference and…we were different we did not really mix together….so all my friends were either Arabs or Indians…or Chinese so always second language people…"