WPLC

Working Papers of the Linguistics Circle
of the University of Victoria

Vol. 22

Issue 1 (October 2012)

Working Papers of the Linguistics Circle of the University of Victoria

– Vol. 22 –

Published by the graduate students
of the University of Victoria Linguistics Department

Department of Linguistics
University of Victoria
P.O. Box 3045
Victoria, B.C., Canada
V8W 3P4

ISSN 1200-3344 (print)
ISSN 1920-440X (digital)

http://web.uvic.ca/~wplc | wplc@uvic.ca

Table of Contents

Acknowledgments

First we would like to acknowledge the contributions of the authors for their research efforts that fill the pages of this issue. A particular thanks also goes out to the volunteer committee who helped shape this volume. Thank you to the continued support from the Department of Linguistics, and especially to our faculty supervisor, Dr. D'Arcy. The current digital form of the WPLC is supported by Inba Kehoe, Scholarly Communications Librarian, and UVic Libraries Office of Scholarly Communications. We would also like to thank the UVic Work Study program (Student Awards and Financial Aid) for providing funding to hire an undergraduate student to aid with the production of the journal.

Preface to our 22nd volume

This volume of WPLC is a continuation of the most recent history of the journal. WPLC has been tied to the North West Linguistics Conference in varying degrees as a venue for the proceedings of the conference. For the past three years we have been working to put WPLC out as an annual publication from our department.

Two years ago we published a volume of a collection of papers by both graduate and undergraduate students. This gives the graduate students an opportunity to get feedback and publish works in progress, and the undergraduate students the chance to participate in the academic publishing process.

This volume, although smaller, continues to provide that same opportunity. Last year we enjoyed reading and publishing papers from across the continent for the NWLC. For this volume we reached out once again to linguists across the continent and collected papers from undergraduate and graduate students.

Enjoy!

Editorial Committee,
WPLC 22

Editorial committee

Not vowel epenthesis: Mandarin and Japanese ESL learners' production of English consonant clusters

Akitsugu Nogita & Yanan Fan
University of Victoria
akitsugu@uvic.ca, echofyn@gmail.com

The present study is based on Funatsu *et al.*'s (2008) experimental study about Japanese ESL learners' perception and production of vowel insertion. To further investigate the process of second language vowel insertion, the present study employed a reading task and a repetition task. A syllabification task was also conducted after each task. Both Japanese and Mandarin ESL participants were involved to explore the effect of language experience. The results showed that both Mandarin and Japanese ESL learners with a relatively short length of residence in Canada only occasionally inserted a short vowel in English consonant clusters when they immediately repeated sound stimuli. The between-group difference was that Mandarin speakers with correct phonological representations often attempted to produce consonant clusters without vowel epenthesis but occasionally failed in gestural coordination which resulted in a schwa-like vowel, while Japanese speakers mostly had incorrectly stored English consonant clusters with extra vowel phonemes in their interlanguage mental lexicon but they phonetically deleted or weakened such vowels when they imitated native English speakers' production.
Keywords: vowel epenthesis; vowel intrusion; vowel weakening; vowel deletion; interlanguage phonological representation

1 Introduction

It is well attested that second language (L2) learners may insert an extra vowel into a consonant cluster that is illegal in their first language (L1). For example, native speakers of English tend to insert [ə] in a Polish consonant cluster which is illegal in English, like /zb/ produced as [zəb] (Davidson & Stone, 2003). This paper examines whether Mandarin Chinese and Standard Japanese speaking English-as-a-Second-Language (ESL) learners insert a vowel in English consonant clusters with different motivations. If there are different processes in vowel insertion in L2, the learning and teaching of L2 consonant clusters will be improved.

This present study consists of both production tasks and syllabification tasks with Mandarin and Japanese ESL learners. The production tasks were designed based on Funatsu, Imaisumi, Fujimoto, Hashizume, and Kurisu's (2008) reading task and repetition task to investigate L2 learners' performance in English consonant clusters that are illegal in their L1, which we will discuss in

§2.2. The syllabification tasks are to examine L2 learners' knowledge of English syllables and whether they add extra vowels in their syllabification process. We found that in the production tasks, Japanese speakers inserted a vowel in an English consonant cluster more often than Mandarin speakers did. Both Mandarin and Japanese speakers tend to insert a vowel more frequently when they read English words aloud than when they repeated native speakers' production without written cues. In the syllabification tasks, regardless of the language background, the participants with explicit knowledge of English syllabification notably performed better than those without explicit knowledge. Our interpretation is that there are three different processes for vowel insertion: 1) those who had known syllabification for a long time (all the Chinese participants) correctly understood where to or not to pronounce a vowel but occasionally failed to pronounce consonant clusters because of difficulty in articulation, 2) those who had recently learned syllabification (a few of the Japanese participants) were still interfered by L1 phonotactics and epenthesized vowels, 3) those who did not know syllabification (the rest of the Japanese participants) did not know where not to pronounce a vowel.

2 Background and Hypothesis

2.1 Studies of Dupoux et al. and Funatsu et al.

Dupoux, Hirose, Kakehi, Mehler, and Pallier (1999) found that native Japanese speakers hear "illusory" [u][1] in a consonant cluster when they hear the VCCV structure where the CC is an illegal sequence in Japanese, such as [ebzo] perceived as [ebuzo]. Dehaene-Lambertz, Dupoux, and Gout (2000) approached the same topic by using a brain-based measure. When native French speakers heard the stimulus change from "igumo" to "igmo," they showed strong responses, while native Japanese speakers showed significantly weaker responses when hearing the same change. Dupoux et al. and Dehaene-Lambertz et al. concluded that Japanese speakers did not pay much attention to the difference between CuC and CC if the CC is a phonotactically illegal consonant cluster in Japanese. However, Funatsu, Imaisumi, Fujimoto, Hashizume, and Kurisu (2008) criticized that Dupoux et al. synthetically deleted [u] of the original sound [ebuzo] to make [ebzo] but listeners might have heard a coarticulation effect from the deleted vowel to the neighboring consonants. Funatsu et al. conducted their study consisting of a reading task and a repetition task. In the reading task, novice Japanese English-as-a-foreign-language (EFL) learners read aloud written stimuli that were both real English words and nonsense words. In the repetition task, the participants immediately repeated sound stimuli that were the same words as the written stimuli pronounced by a native English speaker. Funatsu et al.'s study mostly focused on word-initial /Cɹ/ sequences. In the reading task, the speakers

[1] Dupoux et al. (1999) used the symbol [u] for the Japanese high back vowel instead of [ɯ] to be consistent with French. Since roundness is not a distinctive feature in Japanese, either [u] or [ɯ] can be used. We use [ɯ] in this paper according to the custom.

inserted a full vowel (most of the time [o]) after /t/ or /d/, and [ɯ] elsewhere; note that these patterns are the same as vowel epenthesis in loanword adaptation (see §2.1 for more detail). However, in the repetition task, Japanese speakers correctly pronounced the target consonant clusters or inserted only a very short vowel. Thus, Funatsu et al. argued that Japanese speakers can correctly perceive consonant clusters without "illusory" vowels, and occasionally inserted short vowels were due to their gestural mistiming of coordinating two consonants rather than lexical epenthetic vowels.

However, we take a different view. In the repetition task in Funatsu et al.'s study, the short inserted vowels were still [o] before [t] and [d] and [ɯ] elsewhere, which was the same as the reading task and the loanword adaptation patterns. If these were really gestural mistiming, these would have been more like transitional vowel occurring in a gap between two consonants. If these were transitional vowels, there is not phonetic explanation why those were [o] before [t] and [d] and [ɯ] elsewhere. We interpret that in the Japanese EFL learners' production, underlyingly /o/ and /ɯ/ were still present but phonetically shortened or weakened when the participants were influenced by the sound stimuli. In other words, we suspect that as Funatsu et al. mentioned, Japanese speakers can correctly hear English consonant clusters at the phonetic level, and as Dupoux et al. (1999) mentioned, in Japanese speakers' phonological representation, there is still [o] before [t] and [d] and [ɯ] elsewhere in illegal consonant clusters, regardless of whether or not [o] or [ɯ] is phonetically present. Besides, we are also interested in other L1 speakers' behavior in the same tasks. For example, Mandarin speakers used both deletion and insertion in the production of English consonant clusters (see §2.3). As well, L2 learners' modification strategies vary in different tasks.

Our research questions are whether vowel insertion by Japanese learners of English is really gestural mistiming as Funatsu et al. discussed, and whether Mandarin speakers behave differently from Japanese speakers.

2.2 Mandarin Chinese and Standard Japanese syllables

The maximal syllable structure in both Mandarin Chinese and Standard Japanese is generally CGVX (C=consonant, G=glide, V=vowel, X=nasal or vowel) (Lin, 2001). The only complex (or branching) onset is a consonant-glide sequence: e.g. /mjæn/ "Noodle" in Mandarin and /kjó:to/ "Kyôto" in Japanese. These consonant clusters have a great sonority distance; a glide is the most sonorous consonant, followed by liquids, nasals, and obstruents, and the greater the sonority distance, the less marked the cluster is (Eckman & Iverson, 1993). This suggests that Mandarin and Japanese allow complex onsets but only unmarked ones. Note that there are some other interpretations for their complex onsets; in Japanese, a minimal pair like /makɯ́/ "drop curtain" vs. /mjakɯ́/ "pulse" is sometimes interpreted as the plain-palatal distinction and /mj/ can be transcribed as /mʲ/ like a single consonant as in Akamatsu (2000). In Mandarin, the glide can be included in a nucleus rather than an onset: e.g. /m-jæn/ instead of /mj-æn/. As

for codas, for both languages, a word-final coda can only contain nasals. A nucleus can be either a single vowel or a diphthong in Mandarin. Japanese has an additional condition that a nucleus can be a phonologically distinctive long vowel (Lin, Y., 2007).

In loanword adaption in Mandarin and Japanese, when phonotactically illegal consonant clusters are in the borrowed words, vowel epenthesis can be seen. Miao (2005) summarized the process of loanword adaption of complex onsets and codas in Mandarin. In the corpus, vowel epenthesis is the most common strategy. In the onset position, the epenthesis only occurs between the two consonants. In the coda position, the epenthesis may happen between the two consonants like CVCVC, or after the second consonant like CVCCV, or after the first consonant as well as after the second one like CVCVCV. Loanword adaptation in Japanese also prefers epenthesis, as in the complex onset "play" adapted to /pɯ.ɾé:/, the complex coda "old" to /ó:.ɾɯ.do/, and the syllable boundary "badminton" to /ba.do.míN.toN/. In Japanese the default epenthetic vowel in loanword adaptation is /ɯ/, but after /t/ and /d/, /o/ is the most common (Funatsu et al., 2008). From these processes, we infer that both Mandarin and Japanese speakers would prefer a CV structure. Thus, epenthesis might happen in the production of consonant clusters.

However, phonotactics in Mandarin and Japanese is more complicated. In more detail, Japanese allows three-mora syllables, which do not fit the aforementioned CGVX template (Vance, 2008). For example, the second syllable [homp] in [ni.homp.pói] (/ni.hoNQ.pói [2] /) "Japanesy" has a complex (or branching) coda (Vance, 2008, p.131), where as Chinese does not allow complex codas. On the other hand, Mandarin allows more types of consonant clusters in syllable boundaries. When the coda of the first syllable is a nasal (/n/ or /ŋ/), the onset of the second syllable can be a plosive (e.g. *benbu* 34[3], /pʰɤn.pʰu/, "headquarters"), a fricative (e.g. *fenshou* 13, /fɤn.ʂo/, "break up"), an affricate (e.g. *fenqi* 12, /fɤn.tɕʰi/, "difference"), a liquid (e.g. *benlai* 32, /pʰɤn.lai/, "original"), and a nasal (e.g. *benneng* 32, /pʰɤn.nɤŋ/, "instinct"). In contrast, in Japanese consonant clusters across syllables are allowed as long as the two consonants are a nasal followed by a homorganic consonant, or two consonants that agree in place (Inozuka & Inozuka, 2009), as in [ba.do.mín.toN] (/ba.do.míN.toN/) "badminton," [dén.tɕi] (/déN.tɕi/) "battery," and [tem.móŋ.ga.kɯ] (/[teN.móN.ga.kɯ/) "astronomy." Therefore, Mandarin speakers may have more experience in consonant clusters across syllable boundaries than Japanese speakers do.

Then again, at a more phonetic level, Japanese speakers may have more experience in consonant clusters with obstruents due to vowel devoicing. As a common generalization, the Japanese short high vowels /i/ and /ɯ/ are regularly devoiced between voiceless consonants and in a pre-pausal position. So-called devoiced vowels are often not true vowels but entirely disappear (Vance, 2008).

[2] /N/ is a placeless nasal, and /Q/ is the first half of a geminate consonant (Vance, 2008). Both are moraic.

[3] These numbers indicates phonemic tones in each syllable.

One between voiceless stops is realized as a fricative, as in /kitai/ → [kɨ̥tai] → [kçtai] "expectation," and it may entirely disappear when involving a fricative, as in /sɯtáiɾɯ/ → [sɯ̥táiɾɯ] → [stáiɾɯ] "style" (Vance, 2008, pp. 207-208). Ôno (2007) mentioned that contrary to Ôsaka or Kyôto dialects, Eastern dialects, including Standard Japanese, prefer to use closed syllables both phonologically with /N/ and /Q/ (see footnote 2) and phonetically with devoiced vowels as in the sentence-final polite copula /désɯ/ → [désɯ̥] → [dés]. Ôno speculates that the preference of closed-syllables in Eastern dialects may be due to a substratum of Japanese, such as the Ainu language, although there is much debate regarding the origin of Japanese. Interestingly, the Japanese singers named /ásɯka/ and /gákɯto/ with devoiced [ɯ̥] spell their own names as "Aska" and "GACKT" respectively instead of the regular spellings "Asuka" and "Gakuto." This may imply that these singers not only adopt fancy foreign-looking spellings but interpret [sɯ̥] and [kɯ̥] as the coda of the first syllable. Although Japanese is generally treated as a language with single onsets and codas, we argue that it allows branching onsets and codas on the basis of the preceding evidence. It also allows (pseudo) obstruent clusters at the phonetic level. It would not be very surprising if Japanese ESL learners can pronounce English consonant clusters without much difficulty.

Another difference between Mandarin and Japanese is that based on the traditional classification, Mandarin is a syllable-timed language in which a syllable is the smallest unit (Lin, 2008), whereas Japanese is a mora-timed language in which a mora is the smallest unit (Vance, 2008). Ueyama (2003) found that Japanese ESL learners who had been in the United States for more than five years still incorrectly syllabified English words more than 50% of the time and tended to divide a word into morae. Even advanced Japanese ESL learners may not be able to naturally acquire English syllabification without formal instruction. Ueyama speculated that a possible reason is that the English writing system does not show syllabification. In contrast, in Chinese orthography, each morpheme-syllable is represented by a character (Sun, 2006). Thus, Mandarin speakers have enough experience to identify syllable boundaries. More specifically, in Modern Chinese, morphological compounding leads to the dominance of disyllabic words. Unlike Old Chinese, two, sometimes three free words are allowed to form a compound word (Sun, 2006, p50). The preference of disyllabic words is also reflected by the prefix and suffix in Modern Chinese. When the free morphemes like "*lao*," "*xiao*," and "*zi*" are used as a prefix or suffix, the new word still has the same meaning with the original monosyllabic word. For example, "*hu*" and "*laohu*" have the same meaning "tiger" and "*bei*" and "*beizi*" share the same meaning "cup". All this subconscious knowledge may help them acquire the syllabic system, especially disyllabic words in English.

2.3 Vowel insertion in second language acquisition

Previous research on the acquisition of consonant clusters in L2 has been well documented. Most of these studies investigated this issue in terms of error types

and accuracy rate (Yoo, 2004; Hansen, 2001; Carlisle, 1998; Eckman and Iverson, 1994). In all the studies above, they found that epenthesis is the major modification strategy used by L2 learners. As well, previous studies accounted for learners' acquisition of the consonant clusters in terms of L1 transfer (Kim & Jung, 1998; Bayley, 1996; Broselow, 1987), markedness (Yoo, 2004; Hansen, 2001), sonority (Carlisle, 2006; Davidson, Jusczyk and Smolensky, 2004), and frequency (Davidson, 2006; Pitt, 1998). In a paper about Mandarin speakers' production of consonant clusters, Hansen (2001) used the interaction of all these linguistic constraints above to explain her results. In her study, she found two patterns of epenthesis by Mandarin speakers, CVCVC & CVCCV, which she explained as the interaction between the linguistic environment and the universal preferences of syllable structure. Interestingly, phonological vowel epenthesis by L2 learners may not be always phonetically present as a voiced vowel. Goad and Kang (2003) reported that Japanese ESL learners with higher proficiency commonly devoice and aspirate the word-final voiced stop in English (e.g. [fɑnt^{h}] "fond"). Such aspiration is considered as a voiceless vowel, often [ɯ̥] (/ɯ/) as mentioned in §2.1. Urbanczyk (1996) reported a similar phenomenon in Lushootseed; there is syllabic aspiration, which can be interpreted as a voiceless schwa. In fact, Goad and Kang also reported that Japanese ESL learners delete the final vowel in an English vowel final word and aspirate the preceding stop (e.g. [fɑnt^{h}] "(Jane) Fonda"). In other words, [fɑnt^{h}] "fond" and [fɑnt^{h}] "Fonda" are actually [fɑndɯ̥] and [fɑndə̥] respectively. Phonological epenthesis may phonetically occur as aspiration.

Recently, a few of the studies about L2 consonant clusters began to focus on the gestural model. Davidson et al. (2003) used ultrasound to detect insertion caused by gestural mistiming in L2. Some Russian speakers in this study did not necessarily use phonological epenthesis to repair illegal consonant clusters, but rather failed to employ the appropriate gestural coordination for English initial consonant clusters, which is vowel intrusion. Davidson (2006) conducted a study comparing the acoustic values of a lexical schwa and a so called "transitional schwa." She found that English speakers' pronunciation of "a transitional schwa" was significantly different from a lexical schwa. The values of the first formant (F1), second formant (F2) and duration were much lower in "a transitional schwa" regardless of linguistic environment. Zsiga (2003) found that the articulatory timing in L1 could be transferred to L2. She found the presence of unexpected release burst in Russian speakers' English consonant clusters. Zsiga explained that the audible release burst is typical in Russian which is not common in English and the Russian speakers transferred the L1 articulatory timing pattern to English. Gafos (2002) claimed that the pattern of consonantal coordination is language specific. Generally, there are two ways to produce a transition between two consonants. One is "close transition" which means the release of the first gesture and the target of the second gesture occur at the same time. Thus, there is no release of the first consonant. The other is called "open transition" in which the onset of the second gesture aligns to the c-center of the first gesture as shown in Figure 1.

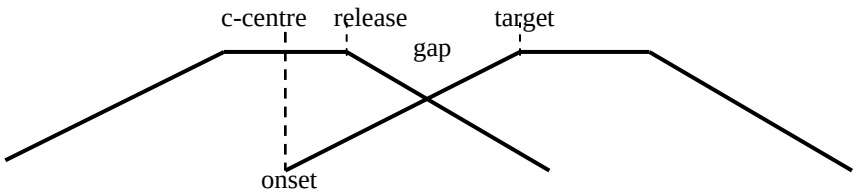

Figure 1. "Open transition" explained by Gafos (2002).

This coordination results in a gap or an open vocal tract between the two consonantal gestures which provides the possibility that the inserted vowel between two consonant is due to gesture mistiming. When L2 learners encounter English consonant clusters, which, according to Gafos, have the "close transition," they may fail to manage the articulatory timing and pull apart two gestures to make an open vocal tract between two consonants, resulting in a short schwa-like vowel or an intrusive vowel.

While epenthesis is phonological and intrusion is phonetic, the third approach is induced errors. We cannot overstate Japanese ESL learners' loanword interference as Goble (2001) found that there is an inordinate amount of loanword interference in Japanese English-as-a-Foreign (EFL) learners. Martin (2004) also states that there is a strong tendency to use *Katakana*-English in English classes in Japan. There is a wide range in the definition of *Katakana*-English, but in this paper, we define *Katakana*-English as almost standardized extension of Japanese loanword adaptation to all the English words: for example "This is a fridge" is almost unanimously pronounced by Japanese ESL/EFL learners as ディス イズ ア フリッジ /dísɯ izɯ a ɸɯɾíQdʒi/ where none of these words is the Japanese vocabulary. When Japanese ESL learners pronounce extra vowels, it may be due to *Katakana*-English stored in their interlanguage mental lexicon regardless of their ability to articulate.

As for the difference between Mandarin and Japanese speakers, Eckman (1981) found more insertion for advanced Mandarin-English speakers, while Weinberger (1987) found equal proportion of insertion and deletion for intermediate level Mandarin-English speakers. Weinberger (1994) explained the difference in terms of recoverability. The advanced learners have more knowledge about the target lexicon than intermediate learners. Thus they will more often realize the ambiguity. To avoid the ambiguity, they will employ more insertion. On the other hand, in Japanese, Smith (2006) mentioned that English words are usually imported to Japanese through written materials rather than spoken materials. There are also a few auditorily imported English loanwords, but there is a notable difference. In auditorily imported words, deletion of consonants is preferred, as in /ɾamɯne_/ "lemonade" where /d/ is deleted and /hé_boN/ "Hepburn" where /p/ is deleted (p. 68), while in orthographically imported words, vowel epenthesis is preferred, as in /aisɯkɯɾíi:mɯ/ "ice-cream." Since Japanese EFL learners commonly use aforementioned *Katakana*-English pronunciation, it is expected that the Japanese speakers with little exposure to

native English speakers' production may prefer vowel insertion especially with written material. As mentioned above, advanced ESL learners are expected to insert devoiced vowels.

2.4 Hypothesis of three different types of vowel insertion

We hypothesize that there are three insertion types: 1) an intrusive vowel, which is not a lexical vowel (or a vowel in the phoneme inventory) but occurs between a consonant cluster when the first consonant is released before the second consonant starts, 2) an epenthetic vowel, which is a lexical vowel that occurs to satisfy lexical syllabification (Hall, 2006; Davidson et al., 2004), and 3) a vowel in misinterpreted L2 phonological representation, which is a lexical vowel that exists in L2 learners' incorrectly memorized mental lexicon. For example, when a Japanese ESL learner pronounces /ɹʌgbi:/, in 1), s/he attempts to produce /gb/ as [gb] without being interfered by L1 phonological constraints but fails to coordinate the two consonants in articulation, which results in a short schwa-like vowel. Or s/he has acquired the L2 phonological structure but has not acquired its phonetic characteristic. In 2), s/he is aware of the L2 /bg/ sequence which is phonotactically illegal in L1, but he/she still has trouble performing it in practice and the L1 syllable structure comes out. In 3), s/he has incorrectly stored "rugby" as /ɹʌgɯbi:/ in his/her interlanguage mental lexicon and pronounced it without knowing that it is incorrect. This study examines whether Mandarin and Japanese ESL learners insert a vowel in English consonant clusters, and if they do, which of the three types is used.

3 Research design

3.1 Participants

We recruited six Mandarin and eight Japanese ESL learners in Canada reportedly at lower-intermediate to intermediate proficiency levels with relatively short length of residence (3 to 13 months) in an English speaking country. Three Mandarin speakers were male and the rest were female in their 20's. Japanese speakers were all from Kantô or Chûbu regions. None of the participants knew other languages besides English and their L1. No hearing problem was reported. Mandarin speakers were labelled as M1, M2, and so forth and Japanese speakers were labelled as J1, J2, and so forth.

3.2 Stimuli

Funatsu et al. mostly focused on /p, b, t, d, k, g/-/ɹ/ sequences. There is a possibility that their Japanese participants pronounced vowels, which might have been phonetically realized as devoiced vowels after voiceless consonants. Dupoux et al. (1999) also used stimuli where devoiced vowels can occur. Our focus in the present study is vowel insertion. Thus, the appearance of voiceless

vowels might affect our results. For this reason, we only chose voiced stops. As well, the target clusters in Funatsu et al.'s study were mostly in the word initial position, which is salient in speakers' perception (Toro, Sebastia´n-Galle´s, & Mattys, 2009). Also, a stop-/ɹ/ cluster is great in sonority, which is relatively unmarked; stops are the least sonorous and /ɹ/ is the most sonorous consonant in English except for glides. In other words, Funatsu et al. used relatively unmarked clusters, which may have enabled novice Japanese EFL learners to perform well, but they may have more difficulty in more marked clusters. Therefore, we use more marked clusters (smaller in sonority distance than stop-/ɹ/), /b, d, g/-/l, m, n, O/ (O=obstruent) (3 × 4 = 12 combinations) in the more marked word-medial position.

The stimuli were both real English words and non-sense words. There were 24 critical items (12 reals and 12 nonsenses) and 16 fillers (8 reals and 8 nonsenses). All the critical items were disyllabic (except for "badminton") with the primary stress on the first syllable and most of the fillers were not disyllabic. The real words were considered familiar to the participants. The nonsense words were made based on English phonotactics. In the nonsense words we avoided /u/ and /oʊ/, which are perceptually similar to Japanese common epenthetic vowels /ɯ/ and /o/ respectively (Nishi, Strange, Akahane-Yamada, Kubo, & Trent-Brown, 2008) in order to avoid potential influence by /oʊ/ or /u/ in the adjacent syllables. These 20 real words were randomized in order, and so were the nonsense words, and the written stimuli and audio stimuli were in different order. The written stimuli were printed on a sheet. The audio stimuli were produced by a phonetically trained female native speaker of Canadian English in an anechoicchamber in the UVic Linguistics Speech Research Lab. The speaker produced each word five times and only the third one, which we assumed to be the most stable and natural, was used as a stimulus. Following are the stimuli. Parenthesized words are fillers.

1. Real words: subject, webmail, webnet, tablet, foodbank, badminton, Sydney, badly, rugby, eggman, magnet, ugly, (avocado, banana, coconut, fruit, grape, ice-cream, strawberry, vegetable)

2. Nonsense words: ebdet, gabmee, gabno, cabla, idgay, cadma, pednay, edlee, agday, egmad, hegneb, agla, (ba, cantukpeg, gamboozee, jeejee, ma, muzz, smecks, sna)

3.3 Procedure

There were four tasks: two production tasks and two syllabification tasks. In the first task, the participants were recorded reading aloud the written real words and nonsense words. The recording was done with the software Audacity set at 44100 Hz and 32-bit float in the UVic Phonetics Lab. In the second task, the participants were asked whether they know what 'syllable' (or "*yinjie*" and "*onsetsu*" in Mandarin and Japanese respectively) was, and they were asked to

separate each written word into syllables. We demonstrated how to divide the Mandarin word "*yinjie*" (syllable) into "yin-jie" by making a pause between syllables to Mandarin participants. To the Japanese speakers, because most Japanese participants did not know syllable, we demonstrated the multi-CV-syllable Japanese word "*wasabi*" divided into "wa-sa-bi," and explained that the monosyllabic Japanese word "*ka*" (mosquito) could not be divided into a smaller unit. Because the number of syllables and moras in both "*wasabi*" and "*ka*" match, the participants would not figure out the difference between syllable and mora. The participants who did not know syllable were instructed to syllabify according to their impression. In the third task, they listened to each stimulus without looking at written cues and immediately mimicked the stimulus. To avoid practice effects, the participants heard each stimulus only once, except for a few cases when the participants could not repeat at all. In the fourth task, the participants listened to each word and divided it into syllables without looking at written cues.

For data analysis, error patterns and duration of inserted vowels were analyzed on the phonetic software PRAAT. Duration was measured from the left-most zero crossing of the first pitch pulse to the right-most zero crossing of the last pitch pulse on the waveform. Both authors individually analyzed the data and when there were disagreements, we discussed to make the final decisions. When we could not make a decision, we consulted with a third phonetician about how the token should be measured.

4 Results

4.1 Overall results

We categorized five levels of coordination of consonant clusters. 1) "Feature change"; two consonants are so close that phonological assimilation occurs: e.g. the nonsense word "cadma" pronounced as "canma" (nasal assimilation) by J5 and J6, "badly" pronounced as "banly" (sonorant assimilation) by J8 all in the repetition task. 2) "Unreleased"; a cluster shows close transition without release burst. 3) "Released"; the first consonant has release burst or aspiration. 4) "Coalescence"; if participants inserted a vowel, the first voiced stop in the cluster ended up intervocalic and spirantized to a voiced fricative, and then the inserted vowel was absorbed by the fricated stop, which turned into a long fricative or even approximant (e.g. "rugby" → [ɹʌgɯbi] → [ɹʌɣːbi]), or the inserted vowel was merged also with the following /l/ (e.g. "ugly" → [ʌgɯli] → [ʌɣɬːi]). 5) "Vowel insertion"; a vowel is inserted and the spectrogram shows formants and periodic pulses. As going from 1) to 5), the two consonants in the cluster becomes phonetically further apart. Besides these five vowels, we made the category "others" that includes misinterpretation of spellings, careless mistakes, and so on. For example, "g" in "rugby" pronounced as [dʒ] in the reading task was considered misinterpretation of the spelling-sound correspondence of <g>. Incidentally, Japanese speakers often pronounced /l/ as a flap [ɾ] or rather [ɹ]-like

sound because /ɾ/ is the only liquid phoneme in Japanese. It does not count as "feature change" or "others." Table 1 shows the overall results. Note that native English speakers typically unreleased the first consonant in a consonant cluster, but "unreleased" in Table 1 does not necessarily mean that the participants' productions were target-like, and other category does not necessarily mean strong foreign accents.

Table 1. The percentage and the number (bracketed) of each error type in each task. Nb. Each task contains 144 tokens (12 clusters × 2 word types (real, nonsense) × 6 participants) in the Mandarin group and 192 tokens (8 participants) in the Japanese group.

	Reading task		Repetition task		Total	
	Mand.	Japan.	Mand.	Japan.	Mand.	Japan.
1. Feature change	5.6% (8)	0% (0)	2.8% (4)	3.1% (6)	4.2% (12)	1.6% (6)
2. Unreleased	**35.4% (51)**	**13.5% (26)**	**41.0% (59)**	**35.9% (69)**	**38.2% (110)**	**24.7% (95)**
3. Released	36.1% (52)	12.0% (23)	48.6% (70)	30.2% (58)	42.4% (122)	21.1% (81)
4. Coalescence	0.7% (1)	4.2% (8)	0% (0)	2.6% (5)	0.3% (1)	3.4% (13)
5. Vowel insertion	20.8% (30)	66.1% (127)	6.9% (10)	28.1% (54)	13.9% (40)	47.1% (181)
Others	1.4% (2)	4.2% (8)	0.7% (1)	0% (0)	1.0% (3)	2.1% (8)

First, vowel insertion occurred much more frequently by Japanese speakers. Coalescence also occurred more often by Japanese speakers. As mentioned in §2.3, beginner Mandarin ESL learners prefer deletion while intermediate learners are more inclined to vowel insertion for the sake of recoverability. In the present study, all the participants were at lower intermediate level. They may not be aware of recoverability as much as advanced learners. Thus, less insertion might be used to repair the nonnative clusters. As well, it is understandable that the Japanese participants preferred vowel insertion especially with written material in the reading task due to *Katakana*-English pronunciation (extension of loanword adaptation to all the English words).

To compare reading and repetition tasks, in both groups, "Vowel insertion" (as well as "Coalescence," which is a type of vowel insertion in this experiment as mentioned above) dramatically decreased from the reading task to the repetition task, while unreleased stops, which are the closest to target-like productions, increased. These results agree with Funatsu et al.'s study, suggesting that both Mandarin and Japanese participants perceived consonant clusters and succeeded to imitate target-like clusters in many cases. Interestingly, a paired T-test shows that Mandarin speakers' unreleased stops were not significantly more frequent in the repetition task ($p>0.1$), while Japanese speakers showed

significant increase ($p<0.001$). This suggests that the sound stimuli may have been more helpful for Japanese speakers than for Mandarin speakers. In addition, released stops were more frequently pronounced in the repetition task in both groups. Again, this increase is significant in the Japanese group ($p<0.005$), while in the Mandarin group it is marginally significant ($p=0.056$). This indicates that both groups changed the vowel insertion strategy to the release or unreleased strategy in the repetition task, but this change is notable in the Japanese group. We will discuss this further in §5. Feature change decreased in the repetition task in the Mandarin group, while feature change occurred only in the repetition task in the Japanese group. In the Mandarin group, deletion-like feature change sometimes occurred: e.g. "idgay" realized as [ɪggeɪ] where place assimilation took place in /d/ resulting in the geminate [gg]. This may be the trace of the aforementioned non-advanced Mandarin ESL learners' preference of the deletion strategy. In the repetition task, the sound stimuli may have helped Mandarin speakers avoid the deletion-like strategy. In contrast, Japanese speakers prefer insertion in the reading task, so that feature assimilation could occur between the two consonants. In the repetition task, the sound stimuli helped Japanese speakers pronounce consonant clusters, which triggered feature assimilation between the consonants.

4.2 Vowel insertion patterns

We also did find a duration difference between the inserted vowels in the two tasks like Funatsu et al.'s study. Table 2 reports the overall frequency of insertion and the mean duration of the inserted vowels in two tasks according to the consonantal contexts. Table 3 to 6 show individual vowel insertion, release (including aspiration), and unreleased productions. We omitted feature change, coalescence, and others due to their small numbers of occurrences.

Table 2. The frequency of insertion and mean duration of inserted vowels. Nb. For each cluster, the Mandarin group has 12 occasions (6 participants × 2 word types) and the Japanese group has 16 occasions (8 participants).

clusters	Reading task Frequency of insertion (Mean duration)		Repetition task Frequency of insertion (Mean duration)	
	Mand.	Japan.	Mand.	Japan.
/b/-Obs	1 (34ms)	12 (39ms)	0	6 (24ms)
/bm/	1 (54ms)	2 (41ms)	0	0
/bn/	3 (26ms)	12 (40ms)	0	4 (12ms)
/bl/	5 (40ms)	11 (44ms)	2 (46ms)	7 (27ms)
/d/-Obs	1 (75ms)	11 (47ms)	0	1 (27ms)
/dm/	1 (18ms)	13 (42ms)	1 (36ms)	6 (29ms)
/dn/	4 (42ms)	12 (62ms)	2 (61ms)	4 (48ms)
/dl/	5 (39ms)	12 (53ms)	3 (37ms)	8 (47ms)
/g/-Obs	0	9 (31ms)	0	0
/gm/	0	6 (30ms)	0	2 (27ms)
/gn/	4 (97ms)	15 (39ms)	0	10 (23ms)
/gl/	5 (59ms)	12 (59ms)	2 (51ms)	6 (40ms)

Table 3. Consonant clusters in the real words by each Mandarin speaker. Nb: 'V Insertion' Indicates the number of vowel insertion; 'Mean dur.' indicates mean duration of the inserted vowels; 'Released' indicates the number of release or aspiration; 'Unreleased' indicates the number of unreleased stops.

Participants	M1		M2		M3		M4		M5		M6	
	Read	Rep	Read	Rep	Read	Rep	Read	Rep	Read	Rep	Read	Rep
V Insertion	5	2	2	2	1	0	4	1	3	1	0	0
Mean dur.	(31)	(45)	(25)	(21)	(24)		(71)	(80)	(40)	(56)		
Released	5	5	8	8	4	3	4	5	3	10	1	0
Unreleased	**2**	**4**	**2**	**2**	**7**	**9**	**3**	**6**	**6**	**1**	**11**	**12**

Table 4. Consonant clusters in the nonsense words by each Mandarin speaker.

Participants	M1		M2		M3		M4		M5		M6	
	Read	Rep	Read	Rep	Read	Rep	Read	Rep	Read	Rep	Read	Rep
V Insertion	3	1	3	1	0	0	4	1	5	1	0	0
Mean dur.	(51)	(43)	(27)	(33)			(88)	(48)	(69)	(65)		
Released	6	8	9	9	3	5	3	7	6	10	0	0
Unreleased	**0**	**2**	**0**	**2**	**7**	**6**	**2**	**3**	**0**	**0**	**11**	**12**

Table 5. Consonant clusters in the real words by each Japanese speaker.

Participants	J1		J2		J3		J4		J5		J6		J7		J8	
	Rea	Rep	Rea	Rep	Rea	Rep	Rea	Rep	Rea	Rep	Rea	Rep	Rea	Rep	Rea	Rep
V Insertion	7	6	7	4	5	0	10	7	9	4	8	6	12	6	2	2
Mean dur.	(42)	(28)	(22)	(20)	(47)		(28)	(32)	(61)	(57)	(36)	(29)	(60)	(40)	(17)	(14)
Released	2	1	4	5	3	4	0	4	2	3	3	2	0	5	3	1
Unreleased	**0**	**2**	**1**	**3**	**4**	**8**	**2**	**1**	**1**	**4**	**0**	**3**	**0**	**1**	**7**	**8**

Table 6. Consonant clusters in the nonsense words by each Japanese speaker.

Participants	J1		J2		J3		J4		J5		J6		J7		J8	
	Rea	Rep	Rea	Rep	Rea	Rep	Rea	Rep	Rea	Rep	Rea	Rep	Rea	Rep	Rea	Rep
V Insertion	6	1	5	1	8	0	11	2	8	3	8	4	12	7	9	1
Mean dur.	(41)	(20)	(36)	(27)	(71)		(33)	(33)	(73)	(31)	(49)	(35)	(49)	(29)	(26)	(35)
Released	3	4	0	4	0	3	0	7	1	4	1	3	0	3	1	5
Unreleased	**2**	**5**	**3**	**7**	**2**	**9**	**0**	**3**	**2**	**4**	**0**	**4**	**0**	**2**	**2**	**5**

In the reading task, insertion happened in almost every type of cluster (except /b/-Obs and /gm/ by Mandarin speakers). Within tasks, insertion by Mandarin speakers occurred less often before obstruents and /m/ in both tasks. After /d/ and before /l/, there seems a bigger chance of insertion. By Japanese speakers, a similar tendency can be seen although it is not as clear as that of Mandarin speakers. Even in the reading task, there were only two occasions of vowel insertion in /bm/ by Japanese speakers. The /bm/ cluster may be the easiest for Japanese speakers. As for vowel duration, vowels by Japanese speakers in the reading task are clearly longer than those in the repetition task. Each participant showed almost the same tendency as displayed in Table 5 and 6. In contrast, Mandarin speakers' vowels in the reading task were not necessarily longer, probably because there were too few occasions to observe the tendencies. In more detail, the inserted vowels by Mandarin speakers in the repetition task were all schwas. This indicates that in the repetition task insertion was likely to be intrusion, while in the reading task, some of the cases may have happened at a phonological level. On the other hand, in the Japanese group, a short schwa-like vowel occurred only twice in J8's reading task, but most of the others were the common Japanese epenthetic vowels [o] after [d] and [ɯ] elsewhere in both reading and repetition tasks, as in Funatsu et al.'s study. J8's schwas may be intrusion but we considered the others to be epenthetic or vowels originating from misinterpreted L2 representation. The syllabification tasks will tell more details.

4.3 Syllabification Tasks

As insertion happened, how speakers syllabify these words with target consonant clusters may reflect the process of insertion. Thus, we asked the speakers to syllabify each word (both target words and fillers) and produce each syllable separately in both reading task with written stimuli and repetition task with sound stimuli.

First, we asked the participants' knowledge of English syllabification. Reportedly, all the Mandarin speakers knew it, while only J1 and J8 in the Japanese group had received formal instruction in Canada, but not in Japan, about a month and half a year prior to the experiment respectively. J2 reported that she might have been taught it in an English phonetics class at her university in Japan but she forgot it. The numbers of errors were summarized in Table 7.

Table 7. The numbers of syllabification errors in each task by each participant (reading task–repetition task). NB. Those with * had never been taught English syllabification.

	M1	M2	M3	M4	M5	M6		
Real words	1–1	5–5	0–3	6–4	3–5	1–3		
Nonsense words	0–0	2–0	2–2	1–0	1–0	1–0		
Overall	1–1	7–5	2–5	7–4	4–5	2–3		
	J1	(*)J2	*J3	*J4	*J5	*J6	*J7	J8
Real words	4–4	17–8	4–7	7–8	8–7	16–16	6–9	0–0
Nonsense words	1–1	17–11	4–3	2–4	6–2	17–17	3–5	0–0
Overall	5–5	34–19	8–10	9–12	14–9	33–33	9–14	0–0

Overall, only one Japanese participant (J8) who had been taught syllable half a year ago did the entire syllabification correctly. Although all the Chinese participants reported that they all knew the concept of syllable, none of them got full marks. The overall correctness of the syllabification task by the Japanese speakers is 65% in the reading task and 68.1% in the repetition task including the extreme case of J8, while the results of the Mandarin speakers were 90.4% in both tasks. This difference may partially come from the fact that Mandarin is a syllable-timed language while Japanese is a mora-timed language, but come more from the fact that all the Mandarin participants had taken formal instructions of English syllabification in China while none of the Japanese participants had in Japan. In fact, the participants with explicit knowledge of syllabification (without * marks) notably made fewer errors than those without it (with *). J8 performed better than J1 probably because J8 had known English syllabification for several months, while J1 had known it only for a month. This suggests that formal instruction can help ESL learners become aware of the concept of syllabification in English. Although it is unwise to generalize based on such a small sample size, recall that Japanese ESL learners with more than five years of residence in the U.S. cannot acquire English syllabification without formal instructions, as mentioned in §2.2. Interestingly, J2 who reportedly learned syllable in an English phonetics course at university in Japan did not correctly syllabify the stimuli. Makino (2008) stated that an English phonetics course at university in Japan is a mere drop in the bucket for Japanese EFL learners who did not take formal instruction of pronunciation at the beginning. The difference among EFL in Japan and China, and ESL can be a future topic. There was no significant difference between the reading and repetition tasks calculated by paired t-test ($p=1$ Mandarin, $p>0.1$ Japanese). However, both speakers made significantly more errors with real words ($p<0.005$ Mandarin, $p<0.05$ Japanese). One of the possible reasons is they were biased by loanword pronunciation.

In the Mandarin group, among the incorrect tokens, the errors can be categorized in Table 8, which shows that all the errors but one occurred for monosyllabic and trisyllabic words. Mandarin speakers seemed to be good at disyllabic words as mentioned in §2.2.

Table 8. Syllabification errors in the reading task and the repetition task by the Mandarin speakers (Totally 240 tokens). Nb. For "vegetable," either "vege-ta-ble" or "ve-ge-ta-ble" was counted correct.

Error types	Examples and frequency
Referring to foot	straw-berry (9); avo-cado (8); vege-table(5); coco-nut (4); bad-minton (3); ugly (1)
Exclude coda	frui-t (1); gra-pe (1); co-co-nu-t (3); mu-zz (3); sme-cks (2)
Separate clusters	f-ruit (2); sme-k-s (2); smeck-s(1); s-na (1)

From the syllabification, we can see that the Mandarin participants did not have problem in syllabifying the target consonant clusters, except M4 misjudged "ugly" as one syllable in the reading task (and J1 made exactly the same error as well). Errors were caused by other positions or other cluster. For the target clusters, phonologically Mandarin speakers accept the consonant clusters and assign them to the correct syllables, which rules out a vowel in misinterpreted L2 phonological representation. The participants had already acquired the phonological structure of the target clusters, but their phonetic abilities did not reach a ready state. About error patterns, participants syllabified the trisyllable words into feet, which are higher units in the prosodic hierarchy. In the monosyllable words, they exclude the codas or pull apart the consonant clusters, suggestion the participants preferred universally unmarked CV syllables. However, such errors also have the possibility to the results of native language transfer, because CV and CVC syllables are the basic syllable structures in Mandarin Chinese. In any case, the participants were on the way to acquire the English syllabification and prosody.

As for Japanese speakers, contrary to our expectation that they may syllabify English words based on morae, the actual error patterns were quite inconsistent. For example, J2 divided "webnet" into "we-b(ɯ)-ne-t(ɯ̥)" referring to a mora while dividing "icecream" into "i-ce-cream." Individual differences also varied: e.g. the nonsense word "gamboozee" syllabified as "ga-m-boo-zee" in reading but correctly in repetition by J2, as "gam-boozee" in repetition but correctly in reading by J4, as "gamboo-zee" in repetition but correctly in reading by J5, and as "ga-m-boo-oo-zee-ee" in both tasks by J6. Even the CV sequences "banana" was incorrect six times, like "ban-nana" in repetition by J2 and "banana" without division by the others. This implies that the Japanese participants did not know what to do in the syllabification task. We interpret that their errors were random or pre-systematic errors. Their correct answers were also likely due to chance, except for J1 and J8. This time, we simply counted the numbers of errors in each word due to their random error patterns, shown in Table 9. The only words correctly syllabified by Japanese speakers were the one-syllable CV nonsense words "ba" and "ma." Unlike Chinese speakers, Japanese speakers made errors with the target clusters.

Table 9. Syllabification errors in the reading task and the repetition task by the Japanese speakers (Totally 320 tokens). Nb. For "vegetable," either "vege-ta-ble" or "ve-ge-ta-ble" was correct.

Word types	Words and frequency
3-or-more-syllable wrd.	badminton (7); avocado (9); banana (6); coconut (10) ; strawberry (13); vegetable (10); cantukpeg (5); gamboozee (5)
2-syllable wrd.	subject (7); webmail (4); webnet (4); tablet (4); foodbank (4); Sydney (6); badly (4); rugby (3); eggman (4); magnet (4); ugly (8); ebdet (5); gabmee (4); gabno (9); cabla (7); idgay (5); cadma (5); pednay (4); edlee (8); agday (4); egmad (3); hegneb (5); agla (5); jeejee (1)
1-syllable wrd.	fruit (4); grape (5); ice-cream (5); muzz (4); smeks (7); s-na (7)

What we are focusing on is not how Japanese speakers syllabified, but how they pronounced each word in the syllabification task. For example, J6 correctly pronounced the consonant cluster in "eggman" without releasing /g/ in the repetition task, while when she syllabified in the repetition task, she pronounced [ɛ-gɯ-mæn] with clear [ɯ] after [g]. Other participants also tended to add clear [ɯ] or [o]. Even if they heard a native English speaker's production without vowel insertion, they still heard "illusory" [ɯ] or [o] as Dupoux et al. (1999) mentioned. This indicates that even though the participants could hear and pronounce target-like consonant clusters at the phonetic level as Funatsu et al. found, they still heard vowels at the phonological level. In other words, [g] and [gɯ], for example, are allophonic variations of /gɯ/ in their interlanguage. Interestingly, when Japanese speakers correctly syllabified (by chance), Japanese speakers often did not insert a vowel, suggesting that not inserting a vowel at the surface representation (SR) is not the biggest problem. The more serious problem is that they misinterpreted English consonant clusters, assuming that there is a vowel in underlying representation (UR) in their interlanguages. Their vowel insertions are neither epenthesis nor intrusion, but their misinterpreted L2 phonological representation.

More interestingly, when syllabifying, J1 and J8 looked careful not to add extra vowels. When they happened to add a vowel, they self-corrected it, which gave us a glimpse of their, especially J8's, correct phonological representations. This rules out misinterpreted L2 representation. However, J1 and J8 sometimes inserted [ɯ] or [o] in the production tasks. To figure out these inserted vowels' true identity, we did a follow-up interview to J1 and J8, who gave two answers. First, their old bad habit, namely *Katakana*-English, sometimes came out when they got careless. Second, J8 sometimes focused too much on syllable division. The first case is considered as epenthesis, and the second case may be intrusion.

An even more interesting case is J2 who had taken a phonetics course in Japan. She frequently divided the stimuli into morae instead of syllables and added a vowel. However, the vowels added in wrong places were often quite short [ɯ̆] or [ŏ] and were often aspiration or devoiced vowels [ɯ̥] or [o̥] after

voiceless consonants: e.g. [sʌ-bɯ̆-dʒɛ-kɯ̥-to̥] "subject." She might have mistaken morae for syllables, but possibly she had some explicit knowledge of where to and not to pronounce a vowel. She might have attempted to pronounce consonants alone and have ended up with a short vowel or a devoiced vowel, as in /sʌ-b-dʒɛ-k-t/ → [sʌ-bɯ̆-dʒɛ-kɯ̥-to̥]. In her case, vowel insertion may be epenthesis, rather than misinterpreted L2 phonological representation. Still, further study will be needed.

Referring to Table 5 and 6, J8 most correctly pronounced consonant clusters. Interestingly, J8's overall accent was not necessarily the best, such as the issue of [l]. J1 and J2 were better in /l/. According to the limited number of samples, the relationship between explicit knowledge of syllabification and production of consonant clusters was observed. As for J1, who just started to learn syllable, she still seems under the development from *Katakana*-English to target-like English syllabification. J3 showed a dramatic difference between the reading task and the repetition task in production as shown in Table 5 and 6. It is probably because J3 was good at imitating sounds although she did not have explicit knowledge of syllable.

As for Mandarin speakers, they did not add a vowel in syllabification unlike Japanese speakers, except M5 occasionally did. M5 also frequently inserted vowels in the production tasks compared to others. Mandarin speakers, especially M2 and M5, frequently devoiced voiced stops in coda and aspirate them in syllabification. M2 and M5 also frequently aspirated voiced stops in the target clusters in the production tasks. In other words, Mandarin speakers correctly understand phonological representations, but either did not know the English phonetic characteristic or do know it but sometimes fail to produce it.

5 Discussion

5.1 Japanese speakers' L2 phonological representation

The most commonly inserted vowels by Japanese speakers were [o] before /d/ and [ɯ] elsewhere, the common epenthetic vowels in Japanese loanwords, which agrees with the results of Funatsu et al. (2008). In the real words, as Funatsu et al. interpreted, the participants likely associated the English words with the corresponding Japanese loanwords. In fact, their stress pattern also occasionally reflects their loanword interference. For example, P4, P5, P6, P7, and P8 put stress on "min" in "badminton," which corresponds to the Japanese loanword /badomíNtoN/ with the phonemic accent nucleus in /mí/. However, even in nonsense words the participants had never seen, they clearly inserted vowels. Nogita (2010) reported that many Japanese learners of English are never taught the English basic symbol-sound correspondence rules in the six years of English curriculum in Japan. Therefore, it is likely that Japanese learners of English have built their own English symbol-sound correspondence rules. Considering the fact that they almost always added an extra vowel, Japanese learners' interpretation of the English orthography is *abugida* or alphasyllabary, rather than alphabet. Each

consonant letter, probably except for <n> that corresponds to the moraic placeless nasal /N/, has a default following vowel which is pronounced every time a following vowel is required according to Japanese phonotactics; for example, the italicized consonant letters in "su*b*je*c*t" are pronounced as /bɯ/, /kɯ/, and /to/ with the default vowels although there are no vowels in the spelling. The followings are some examples of Japanese speakers' interpretation of the English symbol-sound correspondence rules: <b> - /bɯ/, <c> - /kɯ/, <t> - /to/ (or /tɯ/).

As for the repetition task, Japanese speakers less frequently inserted vowels, but they also released the first consonants in clusters, or devoiced the first consonants and aspirated them, as shown in Table 1, 5, and 6. Japanese speakers' aspiration in a consonant cluster is considered a voiceless vowel, often [ɯ] as mentioned in §4.2. This interpretation is consistent with the interpretation that there is a vowel after a consonant in their UR. When the participants released the first consonant, such as [b] in 'subject', the release was actually /ɯ/ in their mind, which was phonetically minimized. This interpretation is also consistent with their UR, rather than interpreting as gestural mistiming. Even if Japanese speakers pronounced target-like clusters, there was still /ɯ/ in their UR that was phonetically minimized to ø. In Japanese ESL learners' interpretation of English words where a vowel exists in the first place, the vowel is allophonically weakened or deleted. A learner misinterprets 'gb' as /gɯb/ but this /gɯ/ is altered to [gɯ], [k^h], [g] (with release), [g̚] and so forth in SR. We consider these vowel alternations as free allophonic variations. Figure 2 to 5 show examples of vowel alternations by the Japanese participants. Note that only F4 with clear vowel is from the reading task and the others are all from the repetition task.

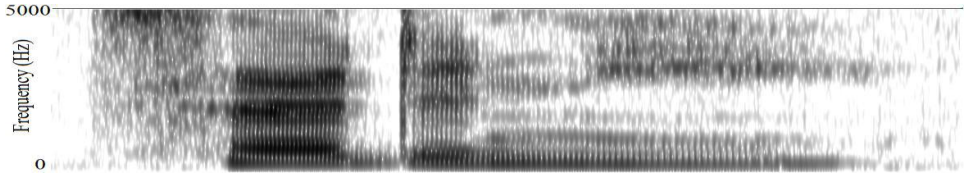

Figure 2. Full vowel: "Sydney" with [ɯ] insertion produced by J3 in the reading task

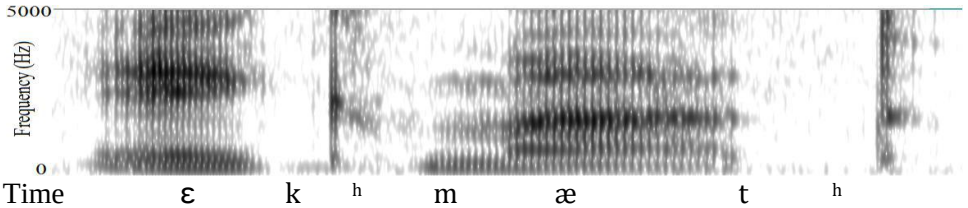

Figure 3. Weakening: "egmad" with aspirated /g/ produced by J4 in the repetition task

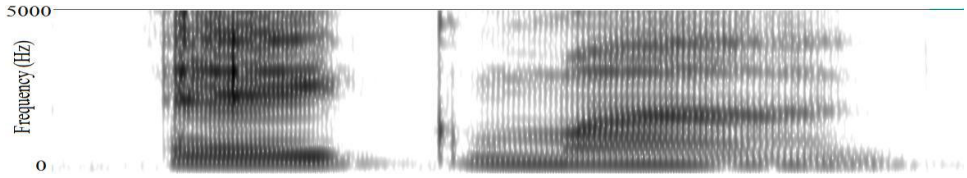

Figure 4. Weakening: "eggman" with release produced by J5 in the repetition task

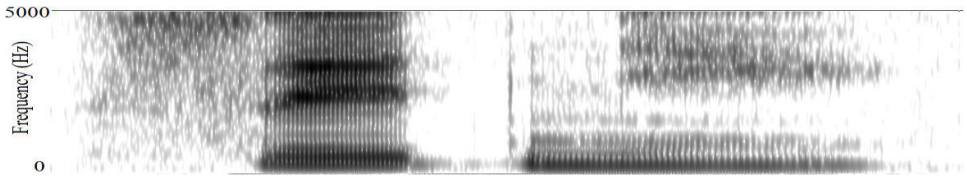

Figure 5. Deletion: "Sydney" with no release produced by J3 in the repetition task

J1 and J8 who had explicit knowledge of English syllabification and possibly J2 may have more similar processes to Mandarin speakers to some extent.

5.2 Mandarin speakers' process of vowel insertion

The Mandarin ESL learners did insert a vowel between two consonants but only in a few cases in the reading task and fewer in the repetition task. Like Japanese speakers, the vowels inserted in the repetition task are shorter than those in the reading task. Meanwhile, the syllabification tasks indicated that Mandarin speakers assigned these clusters correctly. In their processes of target consonant clusters, there was no inserted vowel underlyingly. Moreover, their inserted vowels were mostly schwas rather than lexical vowels, especially in the repetition task. Thus, we inferred that the inserted vowels were due to factors other than native phonological transfer. The Mandarin speakers still may have not fully acquired the articulatory timing patterns of English consonant clusters and they could not correctly manage the coordination of the clusters and break the close transition into open transition, which resulted in an audible sound. If articulatory mistiming or is true in the present study, participants' random insertion may be explained. Incidentally, M6's production of consonant clusters was the best among all the participants. He might have almost acquired target-like clusters in L2. His syllabification task was quite good as well. Figure 6 to 9 show Mandarin speakers short vowel insertion, aspiration, unreleased, and a nasal assimilation, all in the reading tasks.

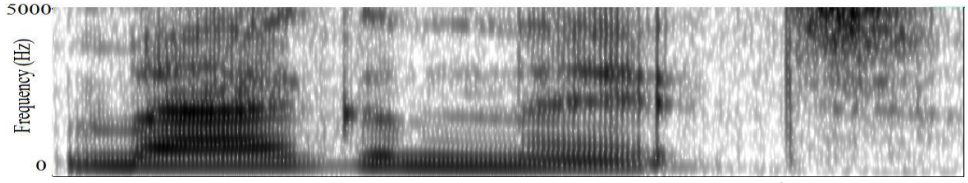

Figure 6. "magnet" with short [ə] insertion produced by M1 in the reading task

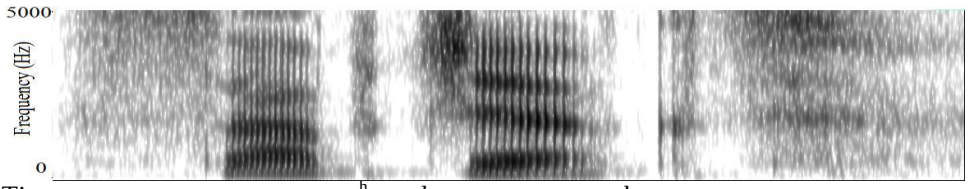

Figure 7. "subject" with devoiced and aspirated /b/ produced by M2 in the reading task

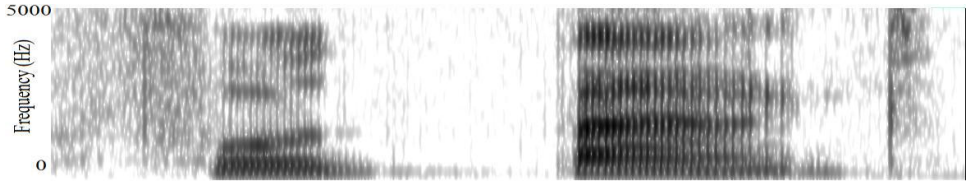

Figure 8. "foodbank" with no release produced by M6 in the reading task

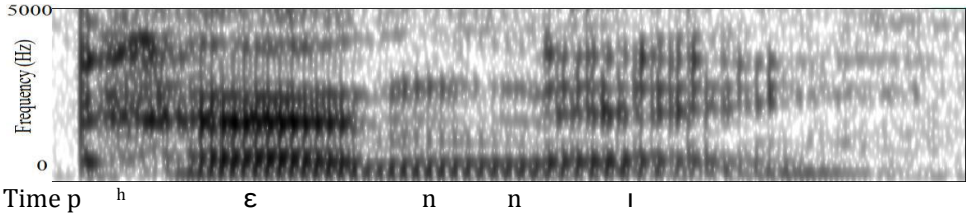

Figure 9. "pednay" with /d/ realised as [n] pronounced by M3 in the reading task

To sum up, Figure 10 and 11 show the three types of vowel insertion. Figure 10 shows (1) intrusion and (2) epenthesis, and Figure 11 shows (3) misinterpreted L2 phonological representation and the variations of SRs. The example is "rugby" that contains an illegal cluster /gb/ in Mandarin and Japanese.

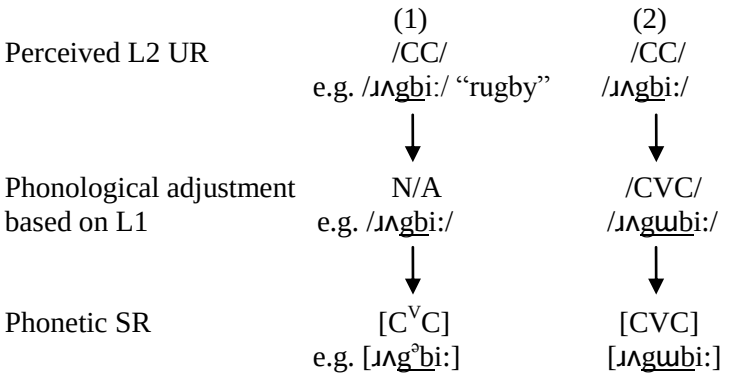

Figure 10. The difference among (1) intrusion, (2) epenthesis.

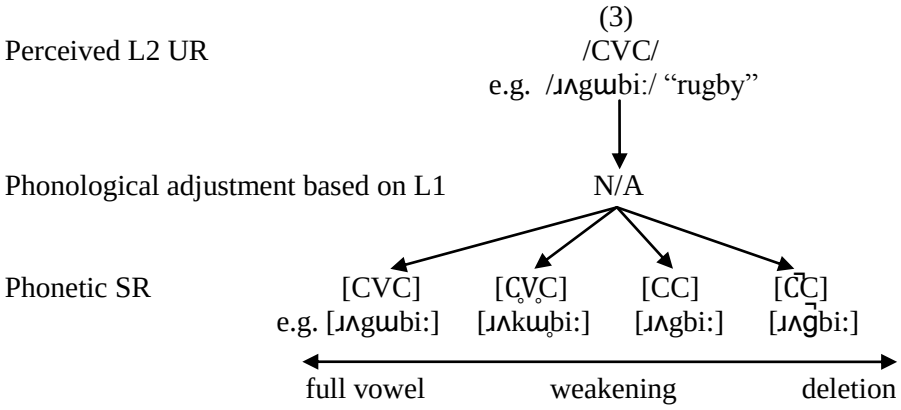

Figure 11. (3) Misinterpreted L2 phonological representation and phonetic manipulation.

5.3 Teaching Implication

There are three processes, intrusion, epenthesis, and misperceived L2 representation. Since interlanguages are in progress, a learner may not exclusively use only one process. Six of the Japanese participants' vowel insertion was mostly misperceived L2 representation. That of the mandarin participants was mostly intrusion, especially in the repetition task. That of J1 and J8 were somewhere between the other Japanese speakers and Mandarin speakers.

For six of the Japanese speakers, English syllabification should be explicitly taught. Their problem is that they do not know when to pronounce consonant clusters, but not necessarily how to pronounce clusters. They should first know the simple rule that if there is a vowel in the spelling, they are supposed to pronounce a vowel (except for silent letters as in "e" in "cake"), and otherwise, they are not supposed to. For the rest of the participants, since they pronounced consonant clusters better in the repetition task, auditory input can

help them acquire the gestural timing of coordination in consonant clusters in the target language.

5.4 Limitation and future study

The major limitation is that it is still not clear whether the participants' inserted vowels with shorter durations were vowel intrusion or epenthesis. In the future research, detailed acoustic measurement should be conducted to detect the gestural mistiming in the production of both Japanese and Mandarin speakers. Also, an ultrasound study is needed to investigate this phenomenon from an articulatory perspective. As well, if Japanese ESL learners' vowel insertion is a knowledge issue, there needs to be an experiment consisting of pre-test, lesson, and post-test in order to examine whether knowledge of syllabification will improve their production of consonant clusters. We also attributed Japanese speakers' more frequent vowel insertion to the difference in language education between Japan and China, but we have to more carefully investigate the educations in these two countries to back up our interpretation. In the methodology, as the English /l/ is notoriously problematic for Japanese ESL learners, all but J1 and J2 almost always pronounced more flap-like or more rhotic liquids. Therefore, we could not observe true C-[l] clusters by these participants. As well, in the present study, the target clusters were across the syllable boundaries, so that tautosyllabic clusters were not examined. We do not know whether we can generalize our interpretations in tautosyllabic clusters.

6 Conclusion

Assuming that vowel insertion in English consonant clusters is always vowel epenthesis due to transfer of L1 phonology is too simplistic. In English consonant clusters that are not allowed in either Mandarin Chinese or Standard Japanese, the Mandarin and Japanese ESL learners in our experiment inserted a vowel with different processes, mainly because of their explicit knowledge of English pronunciation. The Mandarin speakers' vowel insertion was much less frequent than that of the Japanese speakers. Two syllabification tasks confirmed that the Mandarin participants phonologically assigned the two consonants correctly to different syllables, which might indicate that the inserted vowels may not be a problem in UR. Their inserted vowels were often short schwa-like vowels rather than lexical vowels. Therefore, we interpret their productions to be a case of intrusion caused by gestural mistiming. Still, there is detailed measurement needed to judge whether the results indicate epenthesis or intrusion. In contrast, in the case of six of the Japanese participants, a vowel in consonant clusters existed in the first place in their interlanguage phonological interpretation because most Japanese learners of English are exposed to almost standardized *Katakana*-English whose orthography is *abugida* or alphasyllabary; each consonant-letter has a default following vowel which is pronounced every time a following vowel is phonotactically required. Also, none of our subjects had taken

formal instruction of English syllabification during the six years of English curriculum in Japan. However, all the Japanese participants successfully deleted those incorrectly existing vowels at least several times in the repetition task. Such vowel deletion or weakening are free allophonic variations and there is still a vowel in their UR. Therefore, even when they phonetically heard the stimulus [sʌb̚dʒɛk̚t], for example, they syllabified it as /sʌ.bɯ.dʒɛ.kɯ.to/ or the like. This can support Dupoux et al.'s (1999) conclusion that Japanese speakers hear an illusory vowel between an illegal CC sequence. Two of the Japanese participants, who recently learned English syllabification in Canada, behaved more like Mandarin speakers, except their old habit, *Katakana*-English, still came out, which is considered epenthesis. Because Mandarin speakers had known syllables much longer than these two Japanese speakers, Mandarin speakers may have been almost internalized L2 phonological structures and could pronounce consonant clusters without vowel epenthesis or phonological adjustment, while these two Japanese speakers who were new to syllable still had phonological L1 interference. Meanwhile, both Mandarin and Japanese speakers inserted a shorter vowel and less often in the repetition task than in the reading task. This indicates that both speakers can detect and produce the phonetic (not necessarily phonological) difference between CC and CVC, which supports Funatsu et al.'s (2008) conclusion that Japanese speakers can perceive and produce consonant clusters.

Acknowledgements

We are grateful to Dr. Ewa Czaykowska-Higgins for her guidance in this study and expertise in phonology. We are also grateful to all the participants.

References

Akamatsu, T. (2000). *Japanese phonology*. LINCOM Studies in Asian Linguistics 38.

Carlisle, R. (1998). The acquisition of onsets in a markedness relationship: A longitudinal study. *Studies in Second Language Acquisition*, 20, 245-260.

Coulmas, S, (2003). *Writing System*. New York: Cambridge University Press.

Dehaene-Lambertz, G., Dupoux, E. & Gout, A. (2000). Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience, 12*(4), 635.

Davidson, L., & Stone, M. (2003). Epenthetsis Versus Gestural Mistimimg in Consonant Cluster Production: An Ultrasound Study. In G. Garding & M. Tsujimura (eds.), *WCCFL 22 Proceedings*. Cambridge, MA: Cascadilla Press, 165-178.

Davidson, L., Jusczyk, P., & Smolensky, P. (2004). The initial and final states: Theoretical implications and experimental explorations of richness of the base. In R. Kager, W. Zonneveld, & J. Pater (Ed), *Fixing Priorities:*

Constraints in Phonological Acquisition. Cambridge: Cambridge University Press.

Davidson, L. (2006). Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics*, 34, 104-137.

Dupoux, E., Hirose, Y., Kakehi, K., Mehler, J., & Pallier, C. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology, 25*(6), 1568-1578.

Eckman, F. (1991). On the naturalness of interlanguage phonological rules. *Language Learning, 31*, 195–216.

Eckman, F., & Iverson, G. (1993). Sonority and markedness among onset clusters in the interlanguage of ESL learners. *Second Language Research*, 9, 234–252.

Eckman, F., & Iverson, G. (1994). Pronumciation difficulties in ESL: Coda consonants in English interlanguage. In Yavas, M. (Ed.), First and Second Language Phonology, 251-266. San Diego, CA: Singular.

Funatsu, S., Imaisumi, S., Fujimoto, A., Hashizume, M., & Kurisu, K. (2008). Do Japanese speakers perceive nonexistent vowels in non-native consonant clusters?. *The Journal of the Acoustical Society of America, 123*(5), 3072.

Gafos, A. I. (2002). A grammar of getural coordination. *Natural Language & Linguistic Theory*, 20, 269-337.

Goad, H. & Kang, H. (2003). Word-final syllabification in L2 acquisition with emphasis on Korean learners of English. In Juana M. Liceras, Helmut Zobl & Helen Goodluck (eds.) *Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference*. Somerville, Mass.: Cascadilla Press. 122–129.

Goble, D. (2001). Loanword-induced interference in Japanese students' foreign language acquisition: Developing student awareness through experiential learning . *Journal of Chikushi Jyogakuen Junior College, 36*, 55-77.

Hall, N. (2006). Cross-linguistic patterns of vowel intrusion. *Phonology 23*. The United Kingdom: Cambridge University Press. 387-429.

Hansen, J. G. (2001). Linguistic constraints on the acquisition of English syllable codas by native speakers of Mandarin Chinese. *Applied Linguistics*, 22, 338-365.

Inozuka, E., Inozuka, H., & Machida, K. (Ed.). (2009). *Nihongo onsêgakuno shikumi*. Tôkyô, Japan: Kenkyûsha.

Kim, D. Y., & Jung, Y. (1998). Adult ESL Korean speakers' interlanguage acquisition of consonant clusters. *SECOL Review*, 22, 150-169.

Lin, H. (2008). Interlanguage rhythm in the English production of Mandarin speakers. *Proceedings of the 8th Phonetic Conference of China and the International Symposium on Phonetic Frontiers*. Beijing.

Lin, Y. (2001). *The Sounds of Chinese*. Cambridge University Press: United Kingdom. 115-117.

Makino, T. (2008, September). *English pronunciation teaching and phonetic study*, Presented a Phonetics Mini Symposium (1), Ôsaka, Japan.

Mao, R. P. (2005). Loanword adaptation in Mandarin Chinese: Perceptual, phonological and sociolinguistic factors. Ph D. Dissertation in Stony Brook University. 105-110.

Martin, A., (2004). The 'katakana effect' and teaching English in Japan. *English Today, 20*(1), 50-56.

Nishi, K., Strange, W., Akahane-Yamada, R., Kubo, R., & Trent-Brown. S. A. (2008). Acoustic and perceptual similarity of Japanese and American English vowels. *The Journal of the Acoustical Society of America, 124*(1), 576-588.

Nogita, A. (2010). Do Japanese ESL learners' pronunciation errors come from inability to articulate or misconceptions about target sounds?.*Working Papers of the Linguistics Circle of the University of Victoria, 20*, 82-116.

Ôno, S. (2007). *Nihongono genryûo motomete*. Tôkyô, Japan: Iwanami-shoten.

Pitt, M. A. (1998). Phonological process and perception of phonotactically illegal consonant clusters. *Perception & Psychophysics*, 60 (6), 941-951.

Smith, J. (2006). Loan phonology is not all perception: Evidence from Japanese loan doublets. in Timothy J. Vance & Kimberly A. Jones (eds.). *Japanese/Korean Linguistics 14. Palo Alto: CLSI.*, 63-74.

Sun, Ch. F. 2006. *Chinese: A Linguistic Introduction*. UK: Cambridge University Press.

Urbanczyk, S. (1996). *Patterns of Reduplication in Lushootseed*. Ph.D. dissertation, University of Massachusetts.

Ueyama, M. (2003) Awareness of L2 syllable structure: The case of L2 Japanese and L2 English. *Journal of the Phonetic Society of Japan*, *7*(2), 84-100

Vance, T. J. (2008). *The sound of Japanese*. New York: Cambridge University Press.

Weinberger, S. H. (1987). The influence of linguistic context on syllable structure simplification. In G. Ioup & S. H. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system*, 401-417. Rowley, MA: Newbury House.

Weinberger, S. H. (1994). Functional and phonetic constraints on second language phonology. In M. Yavas (Ed.), *First and second language phonology*, 283-302. San Diego, CA: Singular Publishing Group.

Yoo, H. (2004). A longitudinal study of consonant cluster acquisition. *Studies in Phonetics, Phonology and Morphology* 10.3. 481-503.

Rhotic lenition as a marker of a dominant character type in northern Mandarin Chinese

Matthew M. Pankhurst
University of Western Ontario
mpankhur@uwo.ca

This paper looks at social identity with respect to two types of rhotacization: vowel rhotacization, in which a vowel is r-coloured, and consonant rhotacization (i.e. rhotic lenition). Recent studies (Zhang 2005, Zhang 2008) have investigated character type as a sociolinguistic variable affecting rhotacization in Mandarin Chinese speech. Rhotacization, in turn, has sociocultural associations that differ by both geographic region and regional character types (Lee 2007). I argue that in Northern China, there is a correlation between rhotic lenition and dominant, particularly masculine, social identities. In this study, I interview thirteen Mandarin speakers from Henan Province, a distinctly northern—but not northeastern—prefecture. Participants are interviewed and possible lenited tokens are counted. I hypothesize that a positive correlation between identity and lenition will be seen in speakers who perceive themselves as having dominant personalities; that people who identify with a dominant character type will exhibit more tokens of consonant rhotacization in casual speech. To explain this phenomenon, I take the view that there is a prevalent linguistic ideology linking vowel rhotacization with rurality, low social class, and Northeastern identity. I will show that among speakers of Henan Mandarin, vowel rhotacization is an overt marker of this identity, whereas consonant rhotacization (i.e. rhotic lenition) is less overt. Rhotic lenition is a unique and critical variable which functions as a marker of a dominant character type without establishing a Northeast identity.
Keywords: Lenition; Rhotacization; Rhoti; Chinese; Sociolinguistics

1.1 Introduction

Rhotacization among speakers of Mandarin Chinese has been shown to correlate with personal and group identity. Recent studies (Zhang 2005, Zhang 2008) have investigated character type as a sociolinguistic variable affecting rhotacization in Mandarin Chinese speech. Rhotacization, in turn, has socio-cultural associations that differ by both geographic region and regional character types (Lee 2007). I argue that in Northern China, there is a correlation between rhotic lenition and dominant, unrestrained social identities. This paper looks at social identity with respect to consonant rhotacization (i.e. rhotic lenition).

1.2 Rhotacization and Rhotic Lenition

Rhotacization is a word with two meanings. It can refer either to the r-colouring of a vowel (which occurs in rhotic accents) or the process in which a consonant becomes rhotic (a change in the manner of a consonant to become an approximant). The widespread variety of Mandarin Chinese spoken in the North of China exhibits both types: vowel rhotacization and consonant rhotacization (i.e. rhotic lenition). (1) and (2) are examples of the respective phonological processes:

(1)	/pan+ɻ/ →	[paɻ]	“dish”
	/pai+ɻ/ →		“card”
	/ɻ/ →	[Øɻ:]	“son”
	/san.ʂʐ/ →	[san.ɻʐ]	“thirty”
(2)	/ɕiaʊ̯.ʂʐ.hoʊ̯/→	[ɕiaʊ̯.ɻʐ.hoʊ̯]	“as a child”
	/ɕia.tʂʰaŋ/ →	[ɕia.ɻaŋ]	“homestyle”

Vowel rhotacization in Mandarin Chinese is commonly known as *Erhua* (Simplified Chinese: 儿化). Erhua refers to a tendency among Northeastern dialects—especially that of Beijing—whereby speakers add the suffix *er* (SC: 儿) to the last syllable in a word, usually to indicate a diminutive or endearing meaning (Zhang 2000). The addition of the diminutive suffix is peculiar in Standard Mandarin because it is the only non-syllabic morpheme accounted for by the writing system. According to Lin (1992), there are two suffixes in Chinese: *er* and *zi*, and they both occur as root morphemes, with the respective meanings of “child” and “infant” (Lin, 1992). However, in Standard Mandarin, *er* undergoes phonological change when attaching to a root morpheme; hence, the root-morphemic meaning of *er* is not present in affixing. Rather, the suffix may express an endearing meaning (*bao.benr*), change the literal meaning of the word altogether (*xi.fu* vs. *xi.fur*), or as shown in Zhang (2000), be socially (but not grammatically) obligatory (*huar*) or illicit (*panr*). It is noteworthy that whether this process expresses semantic diminutiveness, a change in literal meaning, or is considered obligatory or illicit is entirely based on regional and social variables. Rhotacization is a highly salient phonological feature which is stigmatized everywhere in China outside the Northeast provinces Heilongjiang, Jilin, and Liaoning. It carries strong connotations of rurality, poverty, and Northeasternness. Erhua is known by Chinese speakers and linguists alike to be the “heavy r-sounding” feature which contributes to the Beijing accent Zhang (2000).

Consonant rhotacization in Mandarin Chinese is a much less salient feature and may also be described as *erhua* by some speakers. Rhotic lenition is simply the change of manner from a consonant to the retroflex approximant [ɻ].

It facilitates rapid speech and may be related to stress-timing and prosody, for along with Erhua, rhotic lenition is the other of the two Beijinger "smooth operator" social variables investigated in Zhang (2000). Chinese speakers often consider rhotic lenition to be *erhua*, and thus the same social stigmas that follow vowel rhotacization are associated with rhotic lenition. However, since it is a much less salient feature socially, it is unlikely to carry such strong connotations as *erhua*. Zhang (2008) shows a correlation between the two in Beijing speakers.

While Chinese phonologists generally agree that these features are different in nature, this paper proposes that there are some common characteristics between Mandarin Chinese vowel and consonant rhotacization.

1.3 Xenophobia and Nationalist Linguistic Ideologies of Rhotacization

The choice of using rhotacization in Mandarin Chinese is a site of linguistic and political struggle. While Beijingers may consider it a "gift of gab" (Zhang 2008), there is a strong resentment for this sound among Mandarin speakers across Southern China. This xenophobic (Li 2003) resentment of Erhua is deep-seated in the history of foreign occupation of China. Li (2003) outlines Mandarin Chinese as a "Tartar-Chinese dialect," highlighting that the [ɻ] sound in China is not a native sound. This in turn could help explain the stigmatization of rhotacization: since it is both foreign and non-obligatory, its use may be considered the result of an active choice to use a foreign element rather than Chinese element, or could be perceived as being intrinsically non-Chinese.

Standard Mandarin, meant to be exemplary of one of the varieties spoken in Beijing, is a written, artificial language ideal published in 1955, which accounts for the way people in Beijing speak and the way people across the country ought to speak. It has remained unchanged since its inception and thus does not account for language change. But the ideologies of the creators of the standard are minor issues when considering the number of speakers; more relevant are the popular ideologies of the masses who have been educated in Standard Mandarin and have bought into its imbedded connotations. With regard to rhotacization, this language ideology functions to help improve the social status of people who use vowel rhotacization within social constraints and to decrease the status of those who use it outside of the constraints—including rhotacization of consonants.

The constraints of when to use vowel rhotacization are very explicit in Standard Mandarin: some words either always have the diminutive suffix, while the rest never do. For example, *huar* ("flower"), *yi dianr* ("a little bit"), and *liaotianr* ("to chat") are all meant to be taught as having an obligatory retroflex suffix, while other words such as *panr*, *xifur* are taught as being necessarily bare: *pan* and *xifu*, respectively. In fact, there are no linguistic constraints governing which syllables speakers can and cannot add the diminutive suffix to in daily speech (Zhang 2005); if there are constraints, they are created socially, and since "language creates identity and discontinuity" (Blommaert 1998) socially-imposed constraints will be able to create identity through uniting and dividing those who

conform to and ignore the constraints.

However, since consonant rhotacization is less overt, the government officials in charge of creating Standard Mandarin in 1955 do not seem to have noticed this linguistic feature. (Alternatively, the feature could be a relatively new sociolinguistic phenomenon.) Standard Mandarin has no rules for rhotic lenition. In this light, rhotic lenition could have the effect of portraying a local accent without being associated with the negative qualities that come with rhotacizing outside of the limits of standardness. If this is true, then consonant rhotacization in Northern China may not be considered taboo at all. Although as shown by Zhang (2005) the two types of rhotacization seem co-related, the use of rhotic lenition could be interpreted as a rebellious, unrestrained response to a national standard. For the reasons discussed, such a response could not have such a strong connotation in the South, the Northeast, or Beijing; it would be indigenous to a certain character type in urban Northern China.

1.4 Ubiquity and Stigma of Rhoatacization in Henan Province

Henan Province is a Northern, but distinctly not Northeastern, province in China, situated to the south of Beijing. Collins (1998) states "ideology is generated in particular social sites, often sites of conflictual exchange." The idea that mild rhotacization is patriotic does not become part of an ideological belief until it is encountered displaying potential for conflict. A farm is hardly the place for social power struggles, especially among social classes, and thus it is unlikely to perpetuate an ideology associating a linguistic feature with a low social class. Wealth, culture, education, and fashion are often not tension-raising topics in daily farm life. However, a high school speech contest for scholarship admittance, a group interview for a reporter job, or a party with government leaders are three examples where tensions over wealth, culture, education, and fashion are easily created, and thus these are potential sites for the propagation of linguistic ideologies about those ideas. In a group interview for a job as a reporter in Zhengzhou, proximity of speech to the standard will be closely examined, and a careful thought to place rhotic vowels only on certain words could show not only a high level of compliance, but an air of fashion and education. A scholarship speech contest—in which language, education, and prestige are very explicitly linked—could instill the idea that a light (but standard) use of rhotacization is highly prestigious and cultured.

Henan Province is among the most populous and densest provinces in China, home to over 100 million people. Everyday life in the Henan's capital, Zhengzhou, is one where wages range from ten cents per hour to countless thousands of dollars per day; class consciousness is made glaringly obvious by the juxtaposition of peasants and government officials. Extending this notion, where there are more varieties of language spoken, the greater the potential for cultural differences between people. Cosmopolitan Zhengzhou, like Cosmpolitan Beijing, therefore, is an example of a site where ideologies of rhotacization are prevalent.

On one hand, Zhengzhou is a Northern province in which rhotacization is widespread. On the other hand, if the standard use of rhotacization creates a perceived association with an identity of an urban, educated, cultured, prestigious, and wealthy Beijinger, the complete disregard of the social constraints set by Standard Mandarin will associate a speaker with an opposite identity: an impoverished, non-prestigious, uncultured, uneducated, and rural Northeasterner. The jarring contrast between the associated character types attributed to the speakers of Henan's most and least rhotic accents could be expected to create a sociolinguistic divide between two groups of people in Zhengzhou.

2.1 Hypothesis

The purpose of the study is to see whether there is a correlation between character type and rhotic lenition in Mandarin speakers from Northern China. If a distinct correlation between language and personality is seen in two different groups, there is likely to be a sociolinguistic reason. I propose that there will be a divide between the percentage of rhotacized tokens between extroverted and introverted Mandarin Chinese speakers from Zhengzhou, China.

3.1 Methodology

For this study, 13 native Mandarin speakers from Zhengzhou, Henan Province attending the University of Western Ontario were asked to participate in individual interviews conducted in Mandarin Chinese. All speakers were aged 18-22 and had spent at least 15 years living in Zhengzhou. There were nine men and four women. Each interview lasted for 10-15 minutes.

The participants were each asked a series of five questions. Firstly, each participant was to mention whether he or she was introverted or extroverted, and whether he or she smoked. The participants were then asked to describe a particularly dangerous or life-threatening event from their past. Fourthly, they were asked to tell a story of memorable shopping experience from China, especially one involving conflict with a shopkeeper. Finally, they were asked to construct eleven sentences, each including an integer from 25-35.

The first question was phrased as “Are you introverted or extroverted? For example, how talkative are you with new acquaintances?” The purpose of this question was to gain a self-assessment of the participants' character type. Since rhotacization is stigmatized as being a marker of rurality, low social class, and lack of education, the men who choose use it may be considered to be crude and unrestrained. Many Chinese people remark that erhua is a particularly masculine feature, and mention that the men in the North are bigger and gruffer. While there is no way to empirically test for a variable as abstract as gradable masculinity, speakers who are more open may be less concerned with receiving negative judgment from others. Smoking is also strongly associated with masculinity in China, with the majority of men being smokers. Extrovertedness

was ranked on a scale of 1 to 5, with 1 being very introverted and 5 being very extroverted.

The third question was included to elicit for natural speech, especially the discourse marker jiushi. Each participant told a fairly involved story involving a personal near-death experience. Each participant produced several utterances of *jiushi* among many other tokens.

The fourth question was designed to elicit for *duoshao*, "how much?" and dangshi ("at the time"). Participants described a particularly memorable conflict with a shopkeeper in China. All participants described asking for prices using *duoshao* and several expressed some regret with *dangshi*.

The purpose of the fifth question was to elicit for *ershi* "twenty" and sanshi "thirty." While constructing each sentence, participants generally said the number once before the sentence and once mid-sententially. Non-sentential utterances were not counted, but any utterance that could be interpreted as part of a phrase was counted. Since *ershi* was the only environment in which rhotic lenition would take effect immediately following another rhotic sound, it was important to capture and contrast this token with sanshi.

At the end of the interview, if there were common tokens unsaid, the participants were prompted to say a sentence with a particular word. For instance, eliciting for duoshao, an effective question was "How do you ask for a price?" Speakers generally answered with *zhege dongxi duoshao qian* ("How much does this stuff cost?").

Following the interviews, I listened to each recording, scanning for all lenited and deleted tokens—not only instances of rhotic lenition. For each token, each speaker had two categories: lenited and non-lenited. All utterances of potentially lenitable tokens were recorded, whether they were said using a standard pronunciation or using a lenited form.

Upon counting the tokens, two token categories became apparent based on phonological environment: syllables with the null medial [ɻ̩] and those with the medial vowel [a]. Since the high-frequency tokens *jiushi*, *ershi*, and *sanshi* were the ones most elicited for, a new category was made specifically to look at these tokens. A category was also made to consider all tokens. Instances of non-rhotic lenition or deletion were not included in any of the four token categories. The percentage of lenited utterances was calculated for all individuals in all four categories.

3.2 Methodological Considerations: words

There is continuous debate among Chinese linguists regarding the existence of the word in Chinese. It has been argued (Duanmu 2000) that Chinese has neither derivational morphology nor words in the traditional sense. In fact, the word *ci* ("word") did not exist in the Chinese language until the 19th century. While many of the tokens examined in this study (e.g. *bushi*, *buzhidao*) do not necessarily conform to the Western idea of a word, for the sake of simplicity this paper will nevertheless refer to them as words. Although it may appear to be logically

circular to state that lenition only happens within a word and that the tokens that this study considers to be words all include lenitable consonants, there is in fact reason for making this distinction. Words such as *shi* ("ten") or *shitou* ("rock") are never lenited unless incorporated into a larger word. In turn, words such as *zhidao* ("to know") following random words are not candidates for lenition; for example, *wo zhidao* ("I know") can never lenite. Thus, there needs to be some lexical boundary in which consonants can undergo lenition. This paper will refer to utterances within these boundaries as words. I will not attempt to answer this problem, but only acknowledge that there is some lexical constraint on lenition in Mandarin Chinese.

3.3 Methodological considerations: lenition and syllable deletion in *ershi*

Like all lenition, rhotic lenition is a phonological process of manner in which a consonant is pronounced as a more sonorous sound. It often occurs in and facilitates rapid speech, in much the same way as deletion.

In the recordings from this study, the majority of utterances of the word ershi, "twenty," are pronounced ambiguously. In these utterances, the distinction between er and ershi is inaudible; the entire syllable seems deleted. This is due to the the first and second syllables both beginning with the retroflex approximant [ɻ]. This ambiguity is particularly evident in phrases such as ershi liu. By examining the phonetics of seemingly deleted syllables alone, one cannot make a strong case for lenition.

Traditional Chinese phonological inventories interpret the vowel in the word shi as an close-central unrounded vowel (Cheng 1973). This vowel is seen following these initials: [s], [ʂ], [tʂ], [tʂh], and [ɻ].A traditional phonological approach to syllable deletion would use the rules in (4) and (5) to account for the process in (6), and (3) to explain (7):

(4)	*[ʂ] deletion rule	[ʂ] → ∅ / φ__
(5)	*Nonfreestanding [ɨ] rule	[ɨ] → ∅ / #__
(6)	*Deletion	1. /əɻ.ʂɨ / → /əɻ.ɨ/
		2. /əɻ.ɨ / → [əɻ]
(7)	Rhotic lenition	/əɻ.ʂɨ / → [əɻ.ɻɨ]

While this explanation seems to explain the missing syllable, it does provide any underlying reason. The rules are created simply to explain the data and give no motivation for the deletion, especially in the face of variation. Although variation is seen in the ershi token, this topical description provides no common motivation for deletion and rhotic lenition. It describes everything but explains nothing. Furthermore, the vowel [ɨ] but this vowel would be the only close vowel in Mandarin Chinese that can exist with but not without an initial. Again, there is no explanation; Mandarin Chinese vowels are syllabic and, with

exception to ɨ, are all able to exist on their own. This analysis has more than one flaw.

Rather, an alternative explanation is to consider syllable structure, the unspecified initial [Ø] and the null medial [ʐ̩] (Duanmu 1990). [ʐ̩] describes a continuation of the sound produced by any of [s], [ʂ], [tʂ], [tʂʰ], or [ɻ̩]. It cannot exist without an corresponding initial.

In his account of vowel rhotacization, Duanmu proposes that [ɻ̩] can occupy both medial and final positions simultaneously: "[r] has a syllabic template and becomes [Ør:]. When used as a suffix, [r] has no independent syllabic template, so it replaces the last segment of the root" (Duanmu 1990). According to Duanmu, there are two types of [ɻ̩]: one is syllabic and one is not. He explains that the motivation for affixing is that an [ɻ̩] is not syllabic and must move to attach to another syllable; i.e. in affixation an isolated [ɻ̩] is motivated by a lack of syllabic medial. Thus, it is possible for an [ɻ̩] to occupy the initial, medial, or final position in a syllable. Using Duanmu's notation, the word er could be written as either [ɻ̩] or [ɻ̩̩] (i.e. [ɻ̩ː]), rather than [əɻ̩]. Consider reinterpreting (6) and (7) as a single rule in (8):

(8) Rhotic lenition (deleted syllable)

1. /ɻ̩.ʂʐ̩/ → /ɻ̩.̩ɻʐ̩/
2. /ɻ̩.̩ɻʐ̩/ → /ɻ̩̩.ʐ̩/
3. /ɻ̩.̩ɻʐ̩/ → [Øɻ̩ː]

To justify the rhotacization in (8), we must make some assumptions. Firstly, Duanmu (1990) states that in when rhotacizing, [ɻ̩] is nonsyllabic. If [ɻ̩] is nonsyllabic, and [ʐ̩] is a continuation of the preceding initial, [ʐ̩] would be the continuation of a nonsyllabic syllable. Therefore, in the nonsyllabic syllable [ɻʐ̩], [ɻ̩] would have motivation to attach to the preceding syllable. In the case of ershi, the preceding syllable has a syllabic [ɻ̩] and can accommodate another [ɻ̩] to become realized as [Øɻ̩ː]. The stranded [ʐ̩] will delete, being a continuation of a non-present initial.

It is due to this explanation that it is important to make the distinction between lenition and rhotic lenition, for the rules of rhotacization still apply to rhotic lenition, while they do not apply to regular lenition. However, to account for variation, it could be argued that they do not apply obligatorily. While the majority of utterances of ershi liu are realized as er liu (indicating a deleted syllable), this pronunciation is not completely ubiquitous. There exists variation in the token utterances. In cases where ershi liu seems pronounced erri liu, the following explanation would be sufficient. (9) is the result of the lenition process in (8) in which the [ɻ̩] is not obligatorily nonsyllabic.

(9) Rhotic lenition (deleted syllable)

1. /ɻ.ʂʐ/ → /ɻ.ɻʐ/
2. /ɻ.ɻʐ/ → /ɻ.ɻʐ/
3. /ɻ.ɻʐ/ → [Øɻː.ɻʐ]

The ambiguity between the end products of (8) and (9) is a result of being unable to distinguish [Øɻː] from [Øɻː.ɻʐ] (i.e. [Øɻɻ] from [Øɻɻ.ɻɻ]). The only audible distinction is the length of the utterance. Since the lenited tokens are usually found in rapid speech, it is not always possible to make the distinction between the end products of (8) and (9). However, since both results can be explained as processes of rhotic lenition, they should produce accurate results for the study.

The biggest criticism to this approach is that one could ask why rhotic lenition only causes affixing following a syllable ending in [ɻ]. Following obligatory affixing rules, this process should cause r-colouring of all syllables. This can be explained by the fact that the syllable [Øɻː] is already r-coloured, and pressure from prosody is a motivation for the initial [ɻ]to assimilate and attach to the preceding syllable. The finding of more r-coloured syllables besides [Øɻː] that cause rhotic lenition could prove this hypothesis.

4.1 Results: Environment

The gathered data show that rhotic utterances occur in the following base-form words: *ershi* (IPA: [ɻ.ʂʐ], translation: “twenty”), *sanshi* ([san.ʂʐ], “thirty”), *jiushi* ([tsiʊ̯.ʂʐ]“precisely”), *haishi* ([hai.ʂʐ], “still is”), kaishi (kai.ʂʐ, “to start”), dangshi (daŋ.ʂʐ, “at the time”), doushi ([doʊ̯.ʂʐ], all are), *bushi* ([pu.ʂʐ], “is not”), *xiaoshihou* ([ɕiaʊ̯.ʂʐ.hoʊ̯], “as a child”), *buzhidao* ([pu.tʂʐ.taʊ̯], “to not know”), jibenshang ([tɕi.pən.ʂaŋ], “basically”), wanshang ([wan.ʂaŋ], “night”), jiachang ([ɕia.tʂhaŋ), *tianshang* (in the sky), and *duoshao* ([tuo.ʂaʊ̯] how many). Because all of these tokens can be produced with or without lenition, they occur in variation. Lexically, these words do not have any common features except that they tend not to be nominal; linguistically, the immediate environment of this phenomenon can largely be attributed to phonological in addition to lexical constraints.

As shown in the data above, all of the token utterances are multisyllabic and begin their second or third syllables with a retroflex sibilant [ʂ], [tʂ] or [tʂh]. These sounds are always initials; the medial vowel of the corresponding syllable is always bare [ʐ̩], or [a]. If the syllable is bare, there syllable has is no final; if the medial is [a], the final is either [ʊ̯] or [ŋ].

Thus, the data show that rhotic lenition only occurs in an initial retroflex syllabant following a ; i.e. the following:

3)[1] [+retroflex] [+retoflex]
[+sibilant] → [–sibilant] / σ__
[–rhotic] [+rhotic]

The only possible consonants that undergo lenition occur in the initials of the following syllables: , [ʂʐ], [tʂʐ], [ʂaŋ], [tʂhaŋ], and [ʂaʊ̯], and they must be preceded by a stress. Tone is not a significant factor in determining which sounds can potentially be lenited by Henan Mandarin speakers, but future research could investigate the role of tone, stress and intensity on the variation of these sounds.

4.2 Results: Variation

A total of 332 token utterances were counted over 17 different words. 16 of the tokens involved retroflex rhotic lenition; the remaining token (*dongxi*) saw alveolo-palatal deletion. This token was not included in numeric results, but offers interesting insight into the nature of deletion. Lenition is a change of manner in a consonant to a more sonorant sound, but it does not necessarily account for a change of place. Rhotic lenition appears to only occur in retroflex tokens because the rhotic approximant in Chinese happens to be retroflex as well. In Mandarin Chinese, the sibilant [ɕ] is the most sonorant alveolo-palatal consonant. Therefore, the sibilant in *dongxi* ([toŋ.ɕi]), in an attempt to lenite, has no option but to change to to Ø and is deleted. Therefore, all realized lenition in Mandarin Chinese seen by this study is rhotic in nature.

Among men, there was sharp divide between their percentages of lenited token utterances. Among instances of high-frequency tokens, the extroverted men tended to lenite 25% more than the introverted men. The men who were smokers tended to lenite 22.5% more than the non-smokers. And the extroverted men tended to be smokers (83%). In general, there was a divide between two groups of male speakers, which roughly corresponded both with both smoking and extroversion. There were no men who rhotacized between 18% and 35%. All male participants either had a heavy r-sounding accent with 35-43% rhotacized, or 0-18%.

[1] Since this analysis is grounded in syllable structure (Duanmu 2000), and lenition always at the beginning of a second or syllable in word, word boundaries are inadequate to describe this environment. I have used σ rather than # to explain this process more accurately.

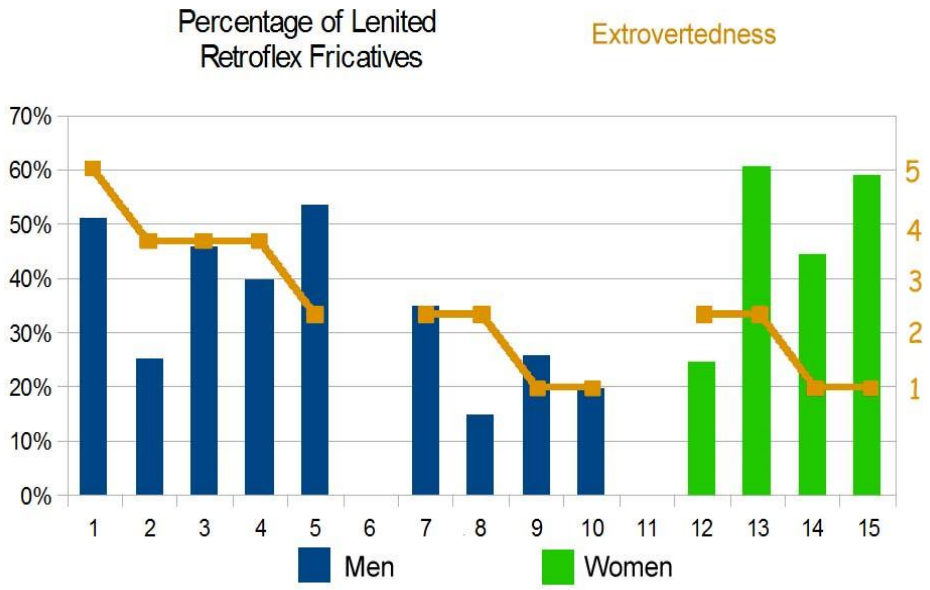

Figure 1.

This chart (Figure 1) compares the extrovertedness of the interviewed men and women with their total percentage of lenited token utterances. An apparent correlation is seen between extroversion and rhotacization in men.

Surprisingly, from this data, women from Zhengzhou tend to use rhotic lenition more often than men.

The following chart (Figure 2) displays the total number of token words uttered by each speaker involving the null medial [ʐ] "buzz" and the medial vowel [a].

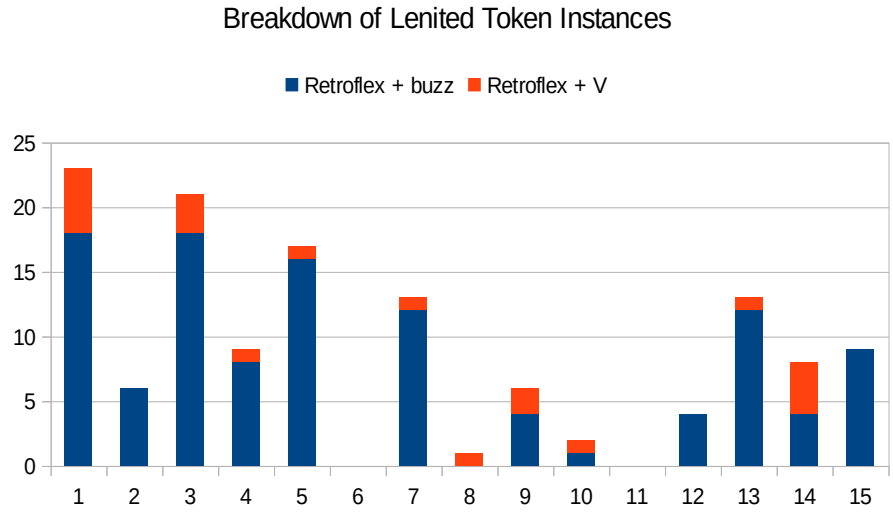

Figure 2.

In general, the more extroverted speakers spoke more than the introverted speakers. This may have had some effect on the data. The vast majority of tokens were [ʐ] party because the methodology elicited for them specifically, partly because they are common, and partly because there are more words that can be rhotacized with medial [ʐ] than with medial [a]. It is interesting that some speakers only lenite initials in words with certain types of medials. Further research could be done on a larger scale to see if this could be a significant factor.

4.3 Analysis

The data show that there is a strong distinction between the rhotacizing tendencies of two groups of Zhengzhou men. In general, the group that tends to rhotacize more also tends to smoke and be more extroverted. In Northern China, extroversion, smoking, and rhotacization are all signs of an unrestrained attitude affiliated with masculinity and social power.

Let us compare the social identity created by this sociolinguistic behaviour with that of using vowel rhotacization in excess. While an over-rhotacizing speaker from the North of China living in Zhengzhou necessarily violates the limit imposed by Standard Mandarin of which words should be r-affixed, a highly-leniting speaker will not violate any sociolinguistic norms. Using a high amount of rhotacization is likely to be interpreted as excessive (i.e. over-rhotacization), but there is no limit or gauge of excess for rhotic lenition. Language, after all, is a unifying force which can link several arbitrary identity traits together:

"That language is seen as a unifying force should be clear. Language assumes the character of a clear identity marker […] Yet, language is only one identity marker among others […] If feathers are predictive of beaks, eggs, and an ability to fly, so is a specific language predictive of a distinct history and culture." (Foucault, 1972)

In this light, it is natural that the production of variables with sociolinguistic weight will work to homogenize a speaker as either an included or excluded member of a group identity. While not as overt as vowel rhotacization, consonant rhotacization functions to establish the speaker as a member of a similar—but distinct—character type: one that includes a sense of local, urban unrestraint and a particularly masculine social dominance. Those who choose not to use rhotic lenition are instead associated with non-local attributes particular to the South of China and international communities, in line with Zhang (2008), which may be associated with being prestigious and refined.

While the sample of women from this study was too small to make any definite conclusions about women and rhotacization in Zhengzhou, their speech and personality were drastically different from the men; there was also much variation among so few speakers. It may be said that the unrestrained character type is not associated with women, who have different social expectations. Thus,

the sociolinguistic role of rhotic lenition is different between women and men.

4.4 Conclusion

By taking into consideration the roles of two sociolinguistic variables, and by contrasting their respective ideologies, it is possible to understand the way in which they are associated with character type. The capital of Henan Province, Zhengzhou is home to a particularly interesting variety of Mandarin Chinese: one that displays widespread rhotic lenition in the face of the stigmatization of vowel rhotacization. The result, among men, is a significant discrepancy between those who choose to lenite retroflex sibilants and those who choose not to. They are not just choosing between the sound of a tongue; they are choosing how they present themselves to the world.

References

Blommaert, J. and Verschuren, J. 1998. "The role of language in European nationalist Ideologies" in *Language Ideologies*. Schieffelin, Woolard and Kroskrity (eds). New York: Oxford University Press 189-210.

Duanmu, S. 1990. *A Formal Study of Syllable, Tone, Stress and Domain in Chinese Languages*. Massacheusetts Institute of Technology. 45-50.

Duanmu, S. 2000. Review of *The Morphology of Chinese: A linguistic and cognitive approach*. By Jerome L. Packard. Cambridge: Cambridge University Press, 2000. Pp. Xvi, 335.

Foucault, M. 1972. *The Discourse on Language* in The Archaeology of Knowledge. New York: Pantheon Books. 215-237.

Lee, E-X. 2007. *Chinese Nationals among "Overseas Chinese" in Singapore: The Sociolinguistic Authentication of Mainland Chinese Identities*. Ph.D Dissertation: The University of Texas at Austin. York: Pantheon Books. 215-237.

Li, C. W. 2003. "Conflicting notions of language purity: the interplay of achaising, ethnographic, reformist, elitist and xenophobic purism in the perception of Standard Chinese." *Elsevier*, 1-38.

Lin, Y. 1992. "Vocalic Underspecification in Two Mandarin Dialects: A Case against Radical Underspecification." *Journal of East Asian Languages*, Springer. 1, 3, 219-253.Wardhaugh, R. 2010. "Gender." *An Introduction to Sociolinguistics*. Blackwell Publishing Ltd. 333-354.

Philips, S. 2000. *Constructing Tongan Nation-State through Language Ideology in the Courtroom* in Regimes of Language. Kroskrity (ed). Santa Fe: School of American Research Press. 229-257.

Sangster, C. 2001. "Lenition of alveolar stops in Liverpool English." Journal of Sociolinguistics. 5/3, 2001: 401-412

Zhang, Q. 2005. Cosmopolitan Mandarin: The social meaning of a linguistic variable.” *Journal of Asian Pacific Communication*, John Benjamins Publishing Company.

Zhang, Q. 2008. Rhoticization and the 'Beijing Smooth Operator': The social meaning of a linguistic variable.” *Journal of Sociolinguistics*, Oxford: Blackwell Publishing Ltd. 12/2, 2008: 201-222.

Vowels in contact: A case study in Romanian

Sorina Dragusanu
University of Western Ontario
sdragusa@uwo.ca

From a desire to better understand the effect of both external and internal factors on vowel production in immigrant groups, this paper explores the change in vowel quality of two speakers of Romanian in Canada. The focus is on the behaviour of the central high vowel, which lacks a correlate in English, and the implication of age at the time of immigration. The Lobanov vowel extrinsic method for normalization of speaker data was used to analyze two speakers' vowel production. The raw and normalized means from each participant are compared to a control. Language internal factors were better at predicting change than language external ones. Both participants maintained their production of the high central vowel due to its phonetic distance. Further, both participants produced a more raised schwa than the control. This points to a divergence rather than convergence towards English.
Keywords: Vowels; Romanian; Vowel change

1 Introduction

Language contact situations have never been more common in the word's history. Immigrant groups negotiate between two or more languages on a daily basis. This interaction creates changes in at least one language, but the changes in minority languages do not always result in the attrition of that language. As Bullock (2004) points out, the processes lead to innovation.

The factors that lead some parts of the phonological system to be more susceptible to change than others are largely unknown. Previous accounts phonological distance may account for the preservation of some elements (Godson, 2004), while pressures to fit in to particular groups makes teenagers more susceptible to influence on their phonological system (Fought, 2004). From a desire to better understand the effect of both external and internal factors on vowel production in immigrant groups, this paper explores the change in vowel quality of two speakers of Romanian in Canada. Would a young adult speaker be more susceptible to convergence than someone in their 30s? What role does phonological distance play in the preservation of contrast?

In order to investigate these questions, the vowel productions of two speakers of Romanian, Dio and Tami, are analyzed. Dio was 30 years old at the time of immigration and Tami was 16. The six monophthongs of Romanian, /i, u, e, ə, o, a/ , are expected to show inter-speaker variation under English influence,

while the seventh vowel, /ɨ/, is expected to be preserved in both speakers because it does not have a counterpart in English.

2 Background

Studies on phonetic and phonological variation in contact situation try to work out which parts of the phonological systems of languages are most susceptible to change and which factors determine this. These categories help researchers interpret linguistic data in the context of society and its daily use. Attention has been given to both the ways in which a minority language can influence a dominant one (e.g. Hoffman and Walker, 2010) as well as how dominant languages result in the attrition of a minority language in its speakers (e.g. Chumak-Horbatsch, 1999; Montrul, 2002). Although most studies focus on the changes taking place the syntactic system, there are also those that tackle phonological changes. However, the factors that lead some parts of the speaker's phonology to be more susceptible to change than others remain largely unknown.

Convergence of phonological features is a common outcome (Bullock, 2004). One language, usually the minority, tends to become more like the other in the features that they share. Conversely, it has also been found that phonetic distance can, in some cases, work to preserve the features that are distinct in both languages. For example, in Godson's (2004) study, the quality of back rounded vowels of Armenian speakers in the U.S. was preserved rather than influenced by the English back rounded vowels.

An important addition to research conducted on these issues in recent years has been the consideration of external, social factors, and their role in change. Such as Chinen and Tucker (2005) and Kuunas (2009) focus on issues of identity in minority language speakers. Not only are community involvement and positive attitudes about one's langauge are linked with overall better langauge performance, but these studies show that by focusing on the speaker researchers can uncover more about the motivation and drives that shape and are shaped by the language system.

Age, another critical factor in language change, is not only relevant in the initial developmental stage of language acquisition but also in the creation of varieties at later ages as well. Fought (2004) shows how phonology can go through a variable stage during teenage years when young adults are trying to define identities and integrate in different group. This critical time period in a person's life can make the phonological system more susceptible to influence as outside pressures to conform act upon it. It is expected then that even though a speaker has a fully formed native phonology, if they were a teenager at the time of immigration they will show more variation overall and be more different from a native speaker than someone more matured at the time of immigration.

3 Questions and Methodology

The study is part of a larger effort to understand the effect of different linguistic and social factors on the phonology of heritage speakers. Data from two native speakers immigrated to Canada is presented to answer the questions:

a) What is the overall change in the vowel quality produced by speakers of differing ages?
b) What role does phonological distance play in the preservation of contrast?

The two speakers, Dio and Tami, that are analyzed were recruited through the larger investigation on Romanian heritage speakers. They were selected for this paper based on their entry time in the study and the different ages at which they entered Canada. Dio is a middle age male who immigrated to Canada over ten years ago. Tami is a female in her late twenties who immigrated at about the same time but was a teenager at the time and finished high school in Canada. Socio-economic class of the two speakers is harder to establish because Dio lived in Romania during the communist era, when personal wealth was not encouraged. Instead the participants' education level, orientation toward the linguistic market, and attitudes about the Romanian language are considered as more relevant indicators of their adherence to the standard language form.

A one hour long interview was conducted with each participant. As congruent with the larger research project four instruments were used to gather as many speech styles as possible adapted from Durand & Pukli (2004). For the purpose of vowel quality analysis a word list of 99 words was read by each participant. From this, vowel tokens were extracted using Praat voice analysis software. The list was composed using the help of the Swadesh list of 100 words (Swadesh, 1971), with a few substitutions made to reflect the stress pattern of the Romanian language.

A web-based interface for the statistical program R, NORM, was then used to create F1/F2 plots for each speaker (Thomas and Kendall, 2007). For baseline comparison, vowel samples were obtained from an online corpus of Romanian (Teodorescu et al., 2010). These were available as audio recordings and were extracted in the same way as the participants. The speakers were predominantly from the eastern part of Romania but reflect a standard accent learned through higher education and exposure to other parts of the country (Teodorescu et al., 2010)[1].

A sociolinguistic interview, about half an hour long, was also conducted in order to gather data about the background, language knowledge, language use, dialectal influences and language attitudes of the participants. The summary for each participant is presented in the next two sections.

4 Dio: Sociolinguistic assessment

[1] The term 'average speaker' is sometimes used in this paper to refer to the control. Since all speakers in this study are native speakers the 'average' represents the current pronunciation conventions of residents in Romania.

Dio is a native speaker of Romanian, 50 years of age, male, born in north-eastern Romania. He has completed his schooling including higher education in Romanian. He was born and has lived most of his life in the north eastern province. He can be considered a speaker of the standard language based on his education level and linguistic market.

Linguistic market considers the social networks that speakers are part of. Also taken into consideration is the multiplicity of contacts and the importance of standard language for occupation (Ash, 2004). His professional training is in geo-chemistry, requiring him to spend a considerable amount of time travelling the country. As an academic, he is required to speak in the most universally understood way. These activities require him to use the standard from of the language more often than the dialectal. In addition, he self-reflects that compared to his spouse, who is born in the same region but is less integrated in the linguistic market, and other relatives he notices that they have more of a regional accent[2].

Dio moved to Toronto about ten years ago and is well integrated in the Canadian society. His friends and colleagues reflect the diverse society of Toronto. Dio uses English as his primary language with which he mediates day-to-day events. Romanian is reserved only for family, and even then he sometimes slips into English with his adult daughter and his spouse. On the other hand, he has very strong connections with his remaining family and friends from Romania, holding almost daily telephone calls and visiting for extended periods of time at least once a year. Since daily events take place in English, it is expected that there will be some departure in his pronunciation compared to the control

Of the other language Dio knows, Japanese stood out as one he could once speak with fluency. However, it has gone unused for more than a decade now and he considers himself no more knowledgeable than a beginner.

Finally, Dio sees languages as important in transmitting cultural values and for that reason he is interested in keeping Romanian alive within himself and his family so that those values are not lost.

5 Tami: Socio-linguistic assessment

Tami is a female in her late twenties, who immigrated to Canada almost thirteen years ago. She is from a town in south western Romania, and has a mixed Serbo-Croatian background with both parents born in Romania. Although most of her early schooling has been in Romanian, she finished high school in Canada and is currently in graduate studies here too.

Tami does not meet the classification requirement for "dialectal voice" as described in the Teodorescu (2010) corpus. However, her linguistic market is

[2] The accent of the Moldova region of Romania can be characterized as a series of predictable, rule-governed changes, such as raising of the mind front vowel. The few words in the list that are prone to these changes did not show variation in the recordings.

much smaller compared to Dio. She speaks Romanian only to some of her relatives in Romania. To Romanian friends and family in Canada she often speaks in English. With her spouse she speaks mainly Spanish and with her children mostly in English. In her professional life as well, English is the norm.
To the researcher, her speech did not sound regionally affected. However, from an anecdote from a recent trip to the country's capital, Bucharest, it became obvious that her speech is altered enough to be considered "foreign" by an average speaker of Romanian. Her production of Romanian vowels will reflect this.

Tami is interested in bettering her knowledge of Romanian for herself. However, from a practical perspective she does not think she will teach her children Romanian, because she cannot at this time see that it would help them. The other languages that are available to her are also used in communications with friends and family so one can see that there is no immediate communicative need for maintaining or transmitting Romanian.

6 Vowel Plots

To compare the data obtained from the speakers to each other, a normalization technique that minimizes speaker difference and preserves dialectal differences had to be selected. The Lobanov vowel extrinsic method for normalization of speaker data was selected because it uses a single anchoring point to align speaker vowel spaces and it defines the edges of the space as well (Clopper, 2009). Since all the vowels of the speakers' system are sampled, there were no disadvantages or problems encountered with this method.

Figures 1 to 3 show the vowel spaces for Dio, Tami, and the average speaker (M) of Romanian created with the non-normalized means obtained from the raw data.

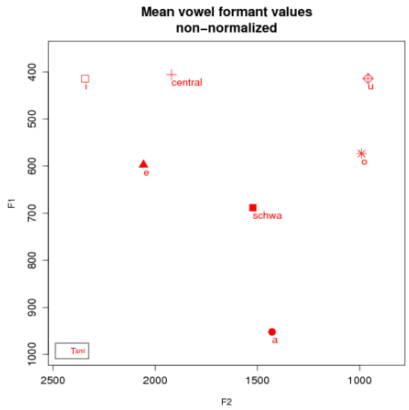

Figure 1. Mean value formant values for Dio

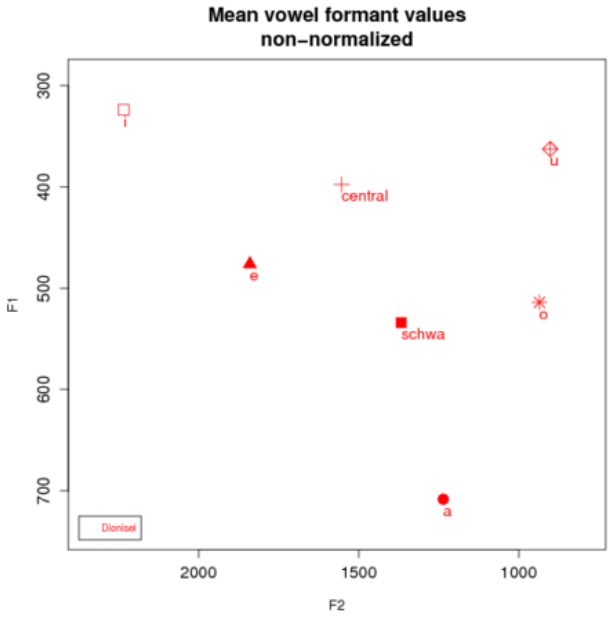

Figure 2. Mean formant values for Tami

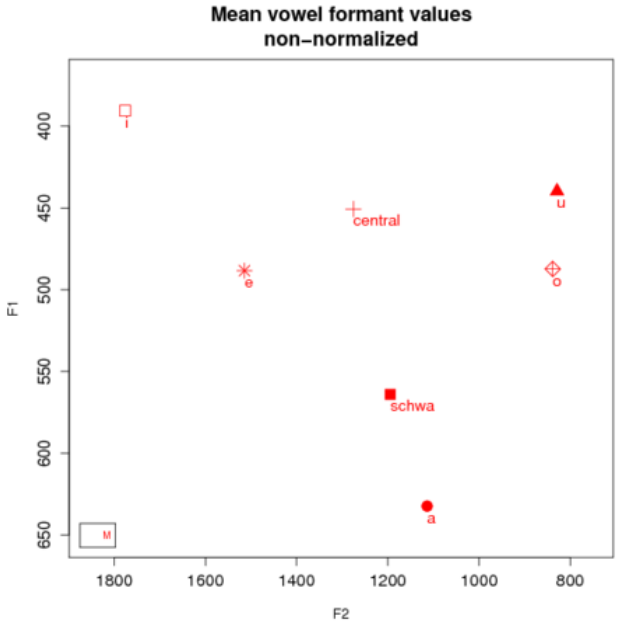

Figure 3. Mean vowel formant values for M.

The average speaker, M, sampled for this comparison is a male between the ages of 30-40 from eastern Romania. According to the classification used in the corpus by Teodorescu et. al. (2010), the speaker is a well educated, non-dialect speaker of Romanian.

From a visual comparison of the means, there is a noticeable difference in the pronunciation of the high vowels by Tami. Her front, center, and back rounded vowels are all in line, while Dio and M show a dip in the height of the central vowel. This is, in fact, because Tami's mean F1 values for /i/ and /u/ are higher than the other participants and match those of the central vowel. The numerical means for the high vowels are shown in Table 1.

Table 1. Formant means for high vowels by participant.

Speaker	Vowel	F1	F2	F3
Tami	i	415	2343	3311
Tami	u	414	959	2526
Tami	central	405	1920	2842
Dio	i	324	2233	2980
Dio	u	363	902	2526
Dio	central	398	1553	2854
M	i	391	1775	2262
M	u	440	829	2074
M	central	451	1275	2197

Already we can see that there are considerable differences between the participants and the control, however, since speakers' vocal tracts are shaped differently, it is not clear from the raw mean alone if these differences are significant or not. Normalization methods were then applied to these means to better understand the whether the differences observed in Tami are significantly different from the average speaker. The next set of F1/F2 plots, *Fig. 4* and *5*, show the mean vowel values for Tami compared to the average speaker and to Dio using the Lobanov normalization technique. As expected from the initial observations of the raw means, Tami's high front and high central vowels differ greatly from Dio and the average speaker. While the vowels /i/ and /ɨ/ are preserving distinction in the speech of Dio as those in the average speaker, Tami's pronunciation of the two vowels appears much closer together in the vowel space. The /i/ is dropping in height while the central vowel is raising. The two vowels are matching in F1 values in Tami's inventory. There is also a difference in the position of the vowels in the mouth with respect to frontness/ backness. The front vowel has a higher F2 while the central vowel has a lower F2. It seems the distinction between the vowel spaces of /i/ and /ɨ/ is lessening and the two are converging towards each other.

Figure 5 also reveals a departure in the pronunciation of the schwa. Dio seems to match Tami for this vowel, a fact confirmed by *Figure 6* which

compares all three speakers. Both participants produce the Romanian schwa much higher in the mouth than does the control. Further, there is a lowering and fronting of the mid-back rounded vowel in Dio's inventory.

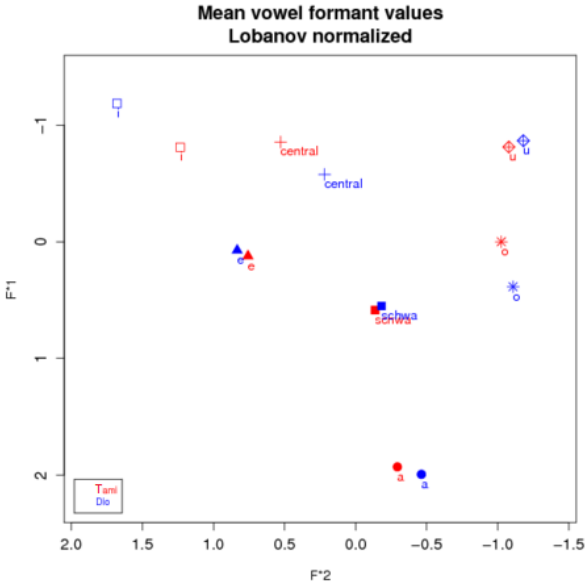

Figure 4. Comparison of normalized vowel formant means for Dio (blue) and Tami (red).

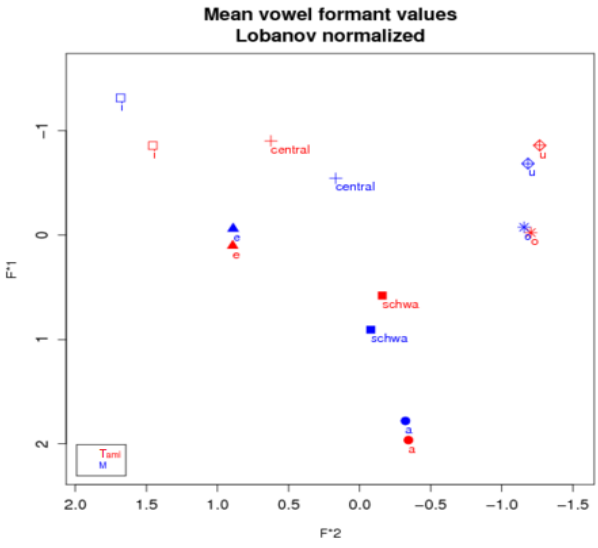

Figure 5. Comparison of normalized vowel formant means for M (blue) and Tami (red).

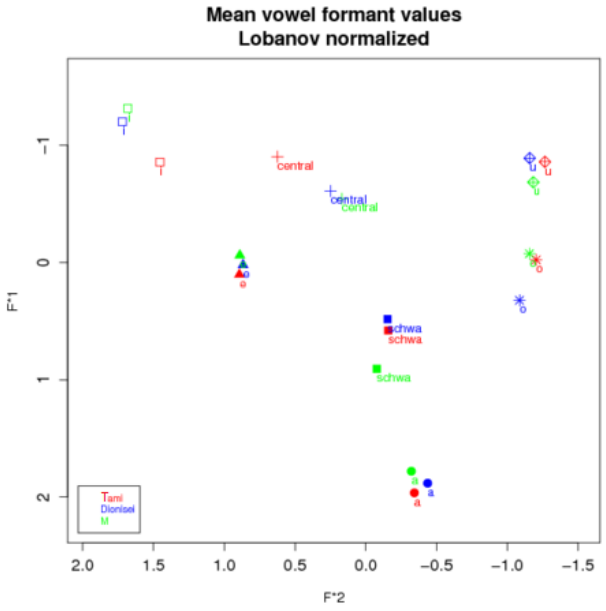

Figure 6. Normalized vowel means of all three participants.

The mean vowel plots show that some changes in the vowel production of the speakers in Canada are similar while other changes follow differing courses. In order to better understand the behaviour of each vowel, it is important to also look at the individual occurrences and deal with any outliers. The next set of plots, *Fig. 4a* and *4b*, show the raw values for every extracted occurrence of each vowel for Dio and Tami. These unpacked means show the distribution of the vowels within their respective vowel spaces.

With respect to the high vowels, Tami shows a close clustering of the central vowel, indicative of a consistent pronunciation even if it is more fronted than the average speaker. Her pronunciation of /i/ is however is spread along a continuum for height, showing a variation, but is not spreading to the vowel space of another vowel. The same conclusion cannot be made for the F2 values, which are fairly stable except for the three instances in the words /limba/, /mic/, and /cinci/[3], where the vowel is realized very back in the mouth, more like the high-back unrounded /ɯ/[4]. It is these three values that have affected the mean to show an /i/ with a lower F2 than in actuality. There appears to be a height ceiling above which Tami cannot produce the front vowels, corresponding with that of

[3] There is no phonological reason why these particular words should be pronounced that much farther in the mouth than the rest. The immediately preceding and following environments for /i/ do not belong to any natural class(es). While there are no minimal or near minimal pairs, the same vowel in the same immediate environment in the word /cine/ is realized at 2800Hz, radically different from the 1400Hz in /cinci/.

[4] /ɯ/ is not a phoneme of Romanian or Serbo-Croatian.

the high back rounded vowel. The problem is most obvious in the production of the central high vowel and the additional attention it receives in production. Tami also spends more time in the production of this vowel, around 40 to 50 ms longer than Dio, which could point to a possible pathology.

The unpacked means also show variation in the vowel space for the mid unrounded vowels. The variation in height seen in schwa is a reflection of the means seen in *Fig. 2b* and *3*. The values for /e/, however, which show a comparable mean to that of the average speaker, range both in height and frontness instead of clustering closer together. This is problematic for maintaining the distinction between /e/ and schwa as the vowel spaces of these are now encroaching on one another.

The low vowel is harder to judge because it has a larger vowel space in which it can be produced. However, the circular diffusion that can be seen in Dio's chart is not overlapping with any other vowel space while Tami's vertical dispersion is overlapping with the mid central vowel. It is either that, in order to keep the contrastive distinction between these two vowels, /ə/ moved into a higher vowel space to make room for the ascending /a/. The other possibility is that /a/ is being pulled to fill the gap left by the movement of the mid central vowel. The second scenario is more likely because Dio's /ə/ production is also higher than the control but without the subsequent movement of /a/.

Dio's unpacked means show much more variation in his pronunciation than the normalized means lead us to believe. The central vowel, for one, is produced in at least two distinct clusters. His production of this vowel is not as stable as Tami's. The cluster with the highest F1 value is very close to the vowel spaces of the mid vowels /e/ and schwa, but there is no actual overlap.

The high front vowels are hard to judge because the variation they display is well within its own space and does not produce any crossover.

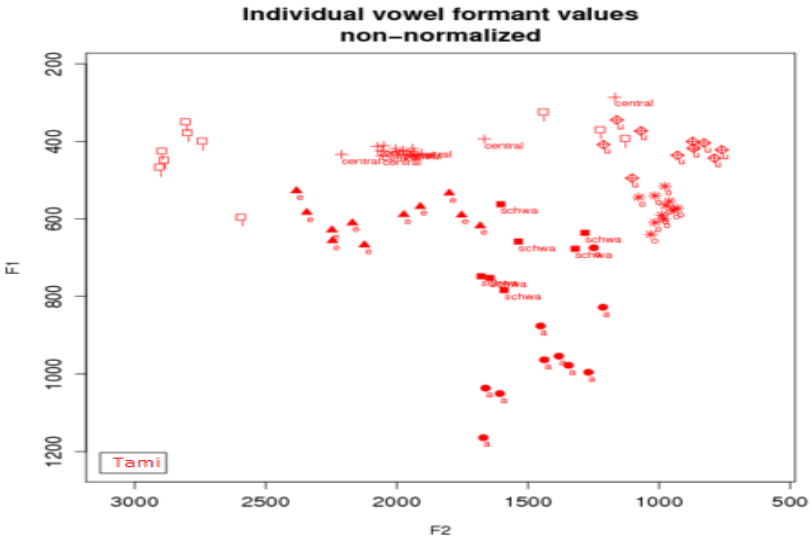

Figure 7. Individual vowel formant values for Tami

Figure 8. Individual formant values for Dio

7 Conclusion and Discussion

There is evidence of both the preservation of distinct language features in immigrants to Canada and of the role that external factors play in creating different patterns of change.

The central high vowel is preserved for both participants. Even Tami's atypical production of /ɨ/ can be attributed to this vowel's singular position. The movement in height of /ə/ can also be attributed to efforts of preservation. While both Romanian and English have a schwa in their phonetic inventories, the two are perceptually different. The English schwa is shorter and more variable (Silverman, 2011; Crosswhite, 1999), while the Romanian is more stable. It is a full vowel that can occur in stressed positions in a word, and has a slightly higher F1 value. Crosswhite (1999) values the English schwa no longer than 40 ms, the Romanian schwa in the data averaged over 100 ms. These features were distant enough to prompt a further distancing of the vowels in order to preserve the distinction between the Romanian mid central vowel and the English one.

Differences in pronunciation could not be attributed directly to age but rather physiological differences. Further, community involvement affected the participants overall language views but it also did not translate into an observable difference in pronunciation. The younger participant has more connections

outside of Romania while the older had already established a stable social life in Romania and only partly extended it in Canada. Their different attitudes about the transmission of the language are also in part fuelled by the arrangement of their social (and professional) contacts.

There is a great difference in the language experience of the participants but their actual language, as far as phonetics goes, does not reflect this. While social factors and language use are very intertwined in contact situations, linguistic internal factor were better at predicting/explaining changes in vowel pronunciation that were due to contact.

8 Future Directions

It is clear that changes are happening in both speakers after such prolonged contact with English. Further confirmation of the observed factors can be made by expanding the analyses to include more participants. As a springboard to the larger study on heritage language change, the study has great relevance. It remains to be seen how second and third generation Romanian-Canadians interpret the schwa and high central vowels. Further, the importance or prominence of language external factors may increase as younger generations build their identity away from their parents' native land.

Finally I would like to bring to attention the need for a better corpus of Romanian to be built with more native samples and especially those of monolingual dialect speakers.

References

Ash, S. (2004). Social class. In J. K. Chambers, P. Trudgill, and N. Schilling-Estes (Eds.), *The handbook of language variation and change* (402-422). Cambridge: Blackwell.

Bullock, B. E. & Green, C. (2004). Phonological convergence in a contracting language variety. *Bilingualism: Language and Cognition, 7*(2), 95-104.

Cheshire, J. (2004). Sex and gender in variationist research. In J. K. Chambers, P. Trudgill, and N. Schilling-Estes (Eds.), *The handbook of language variation and change* (423-443). Cambridge: Blackwell.

Chinen, K. & Tucker, G. R. (2005). Heritage language development: Understanding the roles of ethnic identity and Saturday school participation. *Heritage Language, 3* (1), 27-59.

Chumak-Horbatsch, R. (1999). Language change in the Ukrainian home: From transmission to maintenance to the beginnings of loss: 1. *Canadian Ethnic Studies, 31*(2), 61.

Clopper, C. G. (2009). Computational methods for normalizing acoustic vowel data for talker differences. *Language and Linguistics Compass, 3* (6), 1430–1442.

Crosswhite, K. M. (2000). Vowel reduction in optimality theory. *Dissertation Abstracts International.*

Durand, J. & Pukli, M. (2004). *PAC project*: *Methods and tools*. Toulouse: Université de Toulouse.

Fought, C. (2004). Ethnicity. In J. K. Chambers, P. Trudgill, and N. Schilling-Estes (Eds.), *The handbook of language variation and change* (444-470). Cambridge: Blackwell.

Godson, L. (2004). Vowel production in the speech of western Armenian heritage speakers. *Heritage Language Journal, 2* (1).

Hoffman, M. F. & Walker, J. A. (2010). Ethnolects and the city: Ethnic orientation and linguistic variation in Toronto English. *Language Variation and Change, 22,* 37–67.

Kunnas, N. (2009). Ethnic loyalty as an explanatory factor behind individual differences in variation. *SKY Journal of Linguistics*, *22*, 175-219.

Silverman, D. (2011). Schwa. In M. v. Oostendorp (Eds.), *Companion to Phonology* (628-642). Chichester, U. K: Wiley-Blackwell.

Phonological merging of front vowel monophthongs and diphthongs in New Zealand English televised media: The case of *Shortland Street*

Jordan Dyment
University of Western Ontario
jdyment2@uwo.ca

Phonological merging of the front-vowel monophthongs [ɛ]/[ɪ] and diphthongs [eə]/[iə] is a relatively recent development in the history of New Zealand English (NZE), reaching the level of speakers' awareness only in the past forty to fifty years (Bayard 1995, Gordon et al. 2004, Gordon and Maclagan 2001). Maclagan and Hay (2007) argued that contemporary NZE arose out of general raising in the phonology, such that words like *dress* now sound like [drɪs] and *square*, like [skwiə]. It has been demonstrated that phonological merging in NZE follows Labov's Principles of Change (c.f. Cheshire 2004), with women on the forefront, using raised variants (merging) more frequently than men (Gordon and Maclagan 2001, Maclagan and Hay 2007, Woods 1997). Since speakers became aware of this aspect of NZE, performers on radio, stage, and television began reducing their use of Received Pronunciation and opted for a more New Zealand vernacular (Bayard 1995). Indeed, as Tagliamonte and Roberts (2005) showed, the use of language on television can accurately reflect concurrent changes in the language of the speech community. Phonological raising in mid-to-high front vowel monophthongs and diphthongs was investigated in the long-running New Zealand medical drama, *Shortland Street*. Approximately 800 tokens (704 monophthongs and 80 diphthongs) from actors Olivia Tennet (b. 1991) and Amanda Billing (b. 1976) were collected from early 2008. The rate of monophthong raising was 6% – a rate significantly less than conversational speech (50-80%) – and all examples of diphthongs were categorically mid-vowels. Although raising is resisted on *Shortland Street*, when compared to Amanda, the younger Olivia raises more frequently. It is argued that Olivia's and Amanda's speech is a product of performance, and hence, the actors must be aware of NZE forms in order to resist them.

Keywords: New Zealand English; Phonological Merging; Phonological Raising; Front-vowels; Televised Media; Shortland Street; Goldvarb.

1 Introduction

Gordon and Maclagan (2001: 215) explain that New Zealand undergraduate linguistics students often confuse the transcriptions for word pairs like *fear* and *fair* and *sheer* and *share*. This anecdote highlights the situation of front vowel raising in New Zealand English (NZE), the variety of English spoken in New Zealand. Raising in NZE has been in development since the early 1900's (Bayard 1995, Gordon, Campbell, Hay, Maclagan, Sudbury, and Trudgill 2004 [henceforth, Gordon et al. 2004], Gordon and Maclagan 2001, Maclagan and Hay 2007, Woods 1997). Today, raising embodies the majority of speakers' speech, accounting for 50-80% of their utterances (ibid.). Despite the long history of development, speakers and listeners were not aware of front vowel raising until quite recently, approximately 30-40 years ago (Gordon and Maclagan 2001: 216). After the process was recognized, New Zealanders began accepting this unique aspect of their phonology with pride (Bayard 1995, Glass 2007). Most notably, television and radio personalities (i.e. performers and reporters) are increasing their use of NZE, whereas previously, a standardized Received Pronunciation was the norm (cf. Glass 2007).

In this paper, I present the results of the investigation of a sample of front vowel raising in the long-running New Zealand television series, *Shortland Street*. *Shortland Street* is produced and broadcast in New Zealand. Because of its home-grown background and the surge of front vowel raising in NZE, it is expected that speakers on *Shortland Street* will use NZE-characterized speech. On the other hand, *Shortland Street* is a performance, and the actors on the program augment their phonological choices to accomplish various goals. The resulting tension between "real speech" and "performance speech" will be explored. In the end, it appears that the latter wins out.

The development of raising in NZE begins with its history of development (sec. 1) followed by an overview of its phonological features (sec. 2). Next, previous research (sec. 3) and the current study (sec. 4) will be outlined in detail. The methodology (sec. 5) will then be presented. The results section (sec. 6) reviews the distribution of front vowel raising on *Shortland Street* as well as possible conditioning factors. A detailed discussion on "performance speech" (sec. 7) and possible directions of further study (sec. 8) conclude the paper.

The following research questions will be addressed:
(i) How accurately does New Zealand televised media – specifically, the drama *Shortland Street* – reflect NZE vernacular?
(ii) What is the rate of front vowel raising, and the status of merging, on *Shortland Street*?

1.1 History of NZE

Many theories have been put forward with regard to the origins of NZE. One of the most enduring of these proposals is that NZE developed from Australian and British English. The historical and linguistic connections that NZE shares with

these varieties are well recognized in both the academic and general communities (cf. Bayard 1995, Gordon et al. 2004: 74). However, the socio-political boundaries between New Zealand and Australia are deeply entrenched, and it is unlikely that New Zealanders would aspire toward Australian favour by adopting their linguistic norms (Bayard 1995; but cf. Gordon et al. 2004: 74). Second, linguists publishing research in the early nineteen-hundreds had suggested that NZE developed from Cockney English (Bayard 1995, Gordon et al. 2004: 71-72). Such assertions may stem from a stigmatized attitude toward, or from ignorance about, New Zealand and NZE. Gordon (Bayard 1995: 43-44) pointed out several embarrassingly obvious observations which contradicted this theory. For example, in order for NZE to have developed from Cockney English, it would need to be the case that the majority of immigrants to New Zealand emigrated from Cockney neighbourhoods in England. On the contrary, immigration to New Zealand was widespread across England and other UK countries, like Ireland and Scotland (Gordon et al. 2004: 40-41). Finally, contemporary research has focused on the origins of NZE as a result of dialect mixing or as a result of formation of a new dialect (Gordon et al. 2004). Dialect mixing differs from the above account in that, in dialect mixing, elements from several dialects converge into a new "mixed" dialect. Evidence for dialect mixing is found in tracing the heritage of linguistic features in NZE to the international communities from which they originated. According to new dialect formation, on the other hand, NZE developed along its own trajectory, subject to its own linguistic and social trends. Identifying these factors is a main occupation of sociolinguistics, an area to which we turn our attention in the following section.

1.2 The linguistic variable

NZE is easily recognizable by its raised front vowels (Bayard 1995, Gordon et al. 2004, Woods 1997). Over time, front vowels have raised into and invaded the phonological space of higher vowels, following a track of [æ] → [ɛ] → [ɪ] → [ɪ̄] (fig. 1). Examples of front vowel raising are illustrated in (1)-(2):

(1) a. *Olivia Tennet* on *Shortland Street*, episode 3908:
I wanna stay in one place, to go to school, to have two parents, to have friends.
[fɹɪndz]

b. *Olivia Tennet* on *Shortland Street*, episode 3903:
She'd be in bed all day getting better.
[bɛd] [gɛt^{h}ɪŋ] [bɛt^{h}ə]

(2) a. *Amanda Billing* on *Shortland Street*, episode 3946:
Not drive and endanger yourself and everyone else on the road.
[ɪvɹiwʌn]

b. *Amanda Billing* on *Shortland Street*, episode 3903:
I have said yes.
[sɛd] [jɛs]

The sociolinguistic motivation behind raising, its locality of origin within the phonological space of NZE, and whether movement is prompted by a push- or pull-chain (Maclagan and Hay 2007: 2) are issues that are beyond the scope of this article. However, evidence for chain shift appears to point in the direction of a push (ibid.) Woods (1997: 107), for example, argues from diachronic evidence (collected from two generations of a New Zealand family) that [æ] raising was in advance of [ɛ] raising; this finding suggests a push-chain. Maclagan and Hay (2007: 20), who show that for many young speakers raising of [e] is in advance of [i], seem to point in the same direction.

The encroaching of lower front vowels into higher front vowels' spaces creates a situation of merging: a phonological change by which sound A becomes indistinguishable from sound B. Examples of NZE mergers are informally referred to as the TRAP/DRESS ([æ]~[ɛ]), FLEECE/DRESS ([i]~[ɛ]), and HERE/HAIR ([iə]~[eə]) mergers, respectively; and movement of [ɪ] into central position, [ɪ̄], is referred to as the KIT chain. Gordon and Maclagan (2001: 231) describe several types of mergers: approximation, transfer, and expansion. In approximation vs. expansion, the combined phonological space following merging is either smaller or larger, respectively. Merger by approximation takes longer to reach completion than expansion (ibid.). However, examples like the HERE/HAIR merger (Gordon and Maclagan 2001, Maclagan and Hay 2007, Woods 1997) appear to suggest that post-merge, NZE is retaining small pockets of distinguishable space, a situation reminiscent of merger by approximation (Gordon and Maclagan 2001: 232).

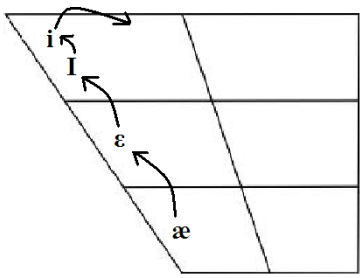

Figure 1. Phonological front-vowel space of NZE.

2 Previous research on NZE

Several trends have emerged from sociolinguistic investigations of NZE. Research has revealed progressive raising of front-mid monophthongs and diphthongs, a situation which is leading toward (and may have already produced) merging between the front-mid and front-high vowels. Furthermore, relationships

have been consistently demonstrated between raising and age, sex, social class, and style. Refer to Bayard 1995 and Gordon et al. 2004 for a review.

Recently, Maclagan and Hay (2007) investigated the phonetics of the NZE front vowels [i] and [e]. Eighty individuals were selected from the Origins of New Zealand English (ONZE) corpus and were grouped according to age, sex, and social class, and thus represent speakers from all combinations of social backgrounds. The authors used speech samples from the Word List component of the ONZE and analyzed individual and group F1/F2 vowel spaces. It was found that young speakers and female speakers had higher [e] frequencies than did older speakers and male speakers. These results neatly follow Labov's principles of change (Cheshire 2004), in that young women are leaders in language change. In some extreme cases (usually involving young women), the frequency for [e] was even higher than [i]! In these situations, diphthongization of [i] was more likely to occur. Furthermore, the authors found that short [i] (i.e. preceding a voiceless consonant) was more likely to be diphthongized. Maclagan and Hay (2007) argue that diphthongization is a response to merging of [i] and [e] within the same vowel space. In order to maintain contrastivity between the two phonemes, speakers are beginning to diphthongize [i]. Diphthongization in the context of short [i] further supports this theory: because short [i] is intricately embedded within the vowel and durational space of [e], it is more likely to be diphthongized than long [i], a response which accords with contrast maintenance. Hence, Maclagan and Hay's (2007) study covers much ground: it demonstrates a sociolinguistic connection between age/sex and raising of [e] and provides evidence that the mid-front and high-front vowels of NZE are merging.

Research on front vowel diphthongs in NZE has uncovered similar results. Gordon and Maclagan (2001) discuss the history of the diphthongs [eə] and [iə], pointing out that although raising of [eə] was present in the 1940's and has increased steadily since, speakers only became aware of this phonological change recently, since the late 1970's (Gordon and Maclagan 2001: 216-217). Such progress is typical of change from below. To determine the character of these changes in real-time, Gordon and Maclagan (2001) conducted a longitudinal study of the diphthongs [eə] and [iə] in the phonology of 14-year-old students. The authors constructed a list of test sentences, each containing an example of each diphthong (e.g. *Come HERE and I'll brush your HAIR*). The list was administered to a sample of students on four occasions, each five years apart, between 1983 and 1998. Over time, variation between the realization of [eə] or [iə] reduced significantly, and in the latest sample, the occurrences of the [eə] variant were nearly negligible. Furthermore, the direction of the merger had clearly shifted from [eə] to [iə]. These results indicate that raising has increased over time and that the change is occurring rapidly. Furthermore, the loss of distinction between the two diphthongs is evidence that they are merging. Like Maclagan and Hay (2007), Gordon and Maclagan (2001) identified an effect of sex, such that the female students began merging the diphthongs, and did so more frequently, before the male students. A small effect of social class was also observed, which the authors attribute to a recent influx of Polynesian students. In

addition, several internal factors were found to contribute to raising of [eə]. For example, when the context is stressed, the realization is more likely to be the raised [iə] variant (Gordon and Maclagan 2001: 229). In all, it appears that the front vowel diphthongs follow the same processes of change and sociolinguistic relationships as the front vowel monophthongs.

3 Current study

The above research paints a comprehensive picture of NZE. However, few studies have investigated the form of NZE in media like television and radio. Although one would assume that programs produced in New Zealand would employ some form of NZE (i.e. either a vernacular or formal style), the situation at hand is in fact more complicated. Radio hosts and news anchors have historically reserved a Received Pronunciation (RP) for broadcasts (Glass 2007). This linguistic style was likely adopted from the long English tradition of using RP, also recognized as the Queen's English, on the BBC. Furthermore, New Zealand imports vast amounts of culture from England, Australia, and America. Bayard points out that, "[o]ur own accent [of NZE] features in only about a quarter of the programmes screened (Bayard 1995: 209)." However, now that phonological changes in NZE have reached the level of awareness, speakers are becoming prouder of their vernacular and more tolerant of the use of NZE on television. For example, news programs are allowing Americanized pronunciations of words like *harassment* (Bayard 1995: 208).

A conflict thus emerges in the use of vernacular versus Standard (i.e. NZE versus RP) on television. Despite increased recognition of and nationalism toward NZE, New Zealanders may still be self-conscious about their speech. Babel (2010) demonstrated that in a repetition task, New Zealanders will use fewer NZE forms if they believe that the speaker holds anti-New Zealand views. These results support Communication Accommodation Theory (cf. Giles et al. 1991), such that New Zealanders will suppress their vernacular depending upon the social context.

When it comes to entertainment programs, on the other hand, the rules are slightly different. Tagliamonte and Roberts (2005) investigated the use of intensifiers (amplifying adverbs such as *so*, *very*, *really*) on the American television comedy, *Friends*. The authors discovered that the frequency of intensifiers closely reflected vernacular speech. Summarizing their results, the authors write, "This evidence leads us to the conclusion that media language actually does reflect what is going on in language… (2005: 296)." This finding is indeed striking, and it opens up the possibility that performers in New Zealand-produced entertainment programs may be using a more vernacular-sounding phonology.

To explore the character of NZE on televised media, I have investigated the use of NZE on the medical drama *Shortland Street*, a program which is particularly suitable for a sociolinguistic investigation. First, *Shortland Street* began its broadcast history in 1992. Since then, it has become New Zealand's

longest running television program and one of its most popular. Because of its status, *Shortland Street*, like *Friends*, is in a unique position to influence its audience both culturally and linguistically. Second, the premiere of *Shortland Street* coincides with front vowel raising having been largely established and widely recognized. Furthermore, because the performers are much closer to the tail-end of language change – the adult cast being born in the 1970's and the adolescent cast being born even more recently, in the 1990's – front vowel raising should be a prominent feature of their speech. Hence, the use of NZE forms is expected. Third, *Shortland Street* is produced by New Zealanders, for New Zealanders. The majority of the cast was born in New Zealand. As such, the vernacular speech of the Pakeha (non-indigenous) actors is NZE. Fourth, *Shortland Street*'s popularity has expanded beyond New Zealand, finding a market in Australia and the UK. International broadcasting thus gives *Shortland Street* another opportunity to exhibit New Zealand culture and language. Finally, because it is a soap opera, *Shortland Street* aims to achieve a balance between fantasy and reality. Story lines are plausible, albeit occasionally embellished, and the characters are meant to be faithful, although idealized, portrayals of everyday people. As such, the actors' performances should appropriately reflect the age, gender, and social status of the characters they are portraying.

My research questions are thus the following (repeated from above):
(i) How accurately does New Zealand televised media – specifically, the drama *Shortland Street* – reflect NZE vernacular?
(ii) What is the rate of front vowel raising, and the status of merging, on *Shortland Street*?

4 Methodology

4.1 Participants

Two female actors, Olivia Tennet (b. 1991, playing *Tuesday Warner*) and Amanda Billing (b. 1976, playing *Dr. Sarah Potts*) were selected for this study. The possibility of following two male actors in addition to the females was considered. However, the phonology of male characters was largely categorical and so was not pursued in this study. Although this observation was not quantified, it will be discussed further in the results section.

Olivia appeared on *Shortland Street* for a brief but memorable run in early 2008 as the niece of the longest-serving and most popular protagonist, *Chris Warner*. For this reason, speech samples from both characters were transcribed from episodes stretching between January and March, 2008.

4.2 Circumscribing the variable context

Following the Principle of Accountability (Feagin 2004) and the methodology of previous studies (Gordon and Maclagan 2001, Maclagan and Hay 2007, Woods 1997), all instances of underlying front-mid monophthongs and front-mid

diphthongs were transcribed. As seen in (3), front-mid monophthongs are realized as the variant [ɛ] and the raised variant [ɪ], and front-mid diphthongs as the variant [eə] and the raised variant [iə].

(3) a. Realizations of /ɛ/:

/ɛ/
/\
[ɛ] [ɪ]

b. Realizations of /eə/:

/eə/
/\
[eə] [iə]

All transcriptions were coded impressionistically. When a monophthong was not discernable between [ɛ] or [ɪ], a situation which only occurred in unstressed, reduced environments, the variable was coded as [ə].

Several contexts posed empirical issues for both transcription and analysis and were excluded from the investigation:

(i) *Post-vocalic* [*r*]. In the case of monophthongs, not only is this context indiscernible between [ɛ] or [ɪ], but it may be an instance of a different vowel altogether. In fact, the sound appears to be an invariable rhotic schwa.

(ii) *Contractions,* e.g. *we're, we'll, he'll, they're,* etc. In addition to the fact that these contexts include the excluded post-vocalic [r], it is unclear if some additional phonological effect is occurring. For example, it is possible that the vowel is reduced from [i], and hence, the context would not even be an instance of merging.

(iii) *Auxiliary 'have'* and (iv) *determiner 'the'*. Even if a phonological effect takes place in unstressed environments, the variable rule would be changing of [a] → [ɛ] in the case of *have* and changing of [ʌ] → [ə] in the case of *the*, which is not the same process as the variable context.

(v) *Any-X*. The first sound in words such as *anyway, anyone, again* is categorically [ə]. This position is invariably unstressed, and in the possible case that it is stressed, it would likely be a low mid vowel, neither [ɛ] nor [ɪ].

(vi) *Syllabic nuclei,* [*r, l, n*]. Not only is it unclear whether these contexts take a preceding vowel, but the question of a vowel's exact identity in this context is also beyond the scope of this article.

4.3 Coding and analysis

Transcriptions were coded for conversion into an input file for Goldvarb X software (Sankoff et al. 2005). The following independent variables were included as possible factors influencing front vowel raising (the dependent variable):
(i) *Speaker and age*. Both Olivia and Amanda were coded independently. Because Olivia was born closer to the tail-end of change, and since she is younger than Amanda (17 and 32, respectively, at time of filming), it is predicted that Olivia will show more instances of front vowel raising.
(ii) *Preceding* and (iii) *Following phonological context*. The internal factor of phonological context may condition front vowel raising. It is possible, for example, that raising begins in one context before spreading to others. Such a situation would be evidence for a grammaticalization process (Tagliamonte and Roberts 2005).
(iv) S*tress*. As discussed above, Gordon and Maclagan (2001) discovered a relationship between stress and raising, which may be relevant for my investigation as well.
(v) *Listener*. The listener is the character whom Olivia or Amanda is conversing with. Olivia's character, *Tuesday*, takes on many roles over the course of her time on *Shortland Street*: she becomes a matriarch when caring for her drug-addicted father, yet she is torn between these duties and her identity as a teenager. It is possible, then, that Olivia speaks differently depending upon her conversant, who brings out one of her many "roles." The same may be true for Amanda. (Due to the dialogue-based design of soap operas, characters are rarely speaking to more than one person at a time; as such, the listener is always one person.)

5 Results

5.1 Distributional analysis

Approximately 800 tokens in total were collected: 383 from Olivia and 401 from Amanda. In addition to the excluded environments discussed in sec. 5.2, additional contexts removed from analysis included the first syllable in the words *refuse* (V), *results*, *remember*, *prescribed*, *before*, and *behind*, as it was unclear whether the syllable was an instance of raising or an underlying high vowel. Verbs taking epenthetic [ə] when forming the past tense (e.g. *voted treated*, *unwarranted*, *teleported*) and *pretty* were excluded as well, since these vowels are (likely) invariably reduced. On the other hand, words ending in *–ment* or *–ent* were not excluded as they constituted a variable context.

Tokens containing a raised variant are listed for each speaker in (4):

(4) a. *Olivia's* tokens containing raising ([ɪ]): *yes, then, let's, accident, best, everything(/'s), everyone, parents, friends, yesterday, different, hundred, exercise, honestly, government, argument, whatever, meant, message, checked, planet, instruments, then, never, get, sense, perfect*.

b. *Amanda's* tokens containing raising ([ɪ]): *sickness, arrangements, attempt, teleported, everyone, parenting, sense*.

Against expectation, all tokens in the case of diphthongs were categorical. Conversely, a modest variation was found for the sample of monophthongs. See summaries in tables 1 and 2.

Monophthongs and diphthongs will be analyzed separately in the next two sections.

Table 1. Overall frequencies of variation in front vowel monophthongs, [ɪ] and [ɛ].

	Olivia	*Amanda*		
n [ɪ]	33	8		
n [ɛ]	350	393		
N	383	401	N_{tot}	704
% [ɪ]	9.7	2.2	$\%_{tot}$ [ɪ]	5.95

Table 2. Overall frequencies of front vowel diphthongs, [iə] and [eə]. (Absence of raising in /eə/ variable contexts is indicated as the proportion of [iə] variants over [eə] variants.)

	Olivia	*Amanda*		
n [iə]	25	11		
n [eə]	18	26		
N	43	38	N_{tot}	80
% [iə]/[eə]	0	0	$\%_{tot}$ [iə]/[eə]	0

5.1.1 Front vowel monophthongs

5.1.1.1 External variables. The rate of front monophthong raising was 6%: 9.7% of Olivia's vowels were raised as opposed to 2.2% of Amanda's. Frequency of raising is significantly less than in previous studies, where raising reached levels between 50-60% (Bayard 1995: 71, Gordon et al. 2004, Maclagan and Hay 2007: 2). Indeed, raising is far less than what has been found in conversational speech, and furthermore, than what would be expected for women of the same age and social class. However, it appears that Olivia raises more frequently than Amanda, a finding consistent with previous research (ibid.) and with Labov's Principles of Change (Cheshire 2004).

The effect of listener on raising is unclear. Olivia exhibited an instance of raising with all listeners she conversed with, save one. In fig. 2 (and cf. table A1 for numerical values), we see that Olivia's tokens are distributed widely. Amanda, on the other hand, raises sparsely, only raising in conversations with a quarter of her listeners (fig. 3). Even in these cases, there is often only a single raised token (table A1). It may appear, then, that Olivia uses raising with all speakers (as would be expected, given her age and gender), whereas Amanda raises only in the presence of specific speakers. However, due to the paucity of raising in Amanda's speech, her differential use of the variable according to her listener cannot be confirmed.

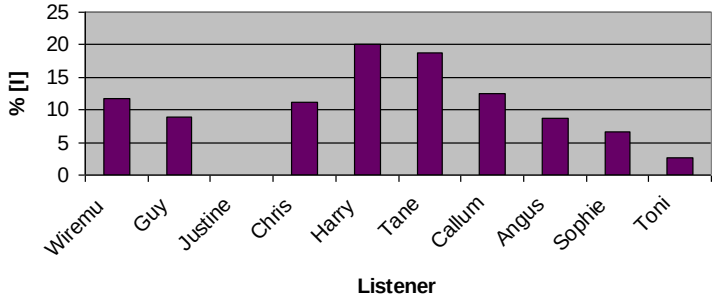

Figure 2. Olivia's proportions of front vowel monophthong [ɪ] (%) by listener. (No bar indicates no [ɪ] variants; see table A1 for numerical values.)

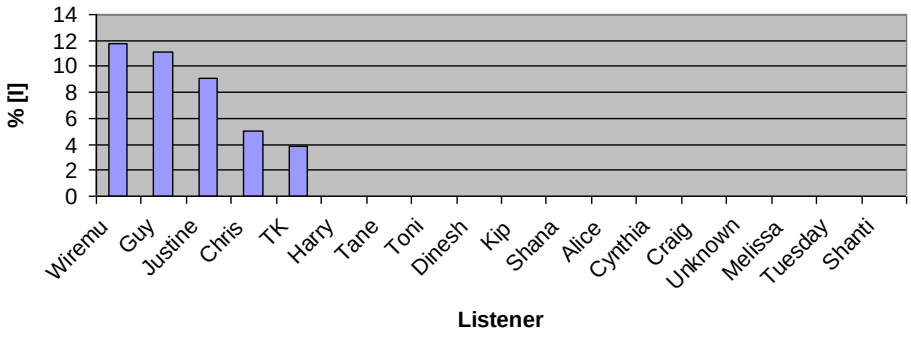

Figure 3. Amanda's proportions of front vowel monophthong [ɪ] (%) by listener. (No bar indicates no [ɪ] variants; see table A1 for numerical values.)

5.1.1.2 Internal variables. Realization of front vowel monophthongs varies across most phonological environments (table 3). Of particular interest is the distribution of preceding [j]. Since this context includes the frequent *yes*, characteristically and (almost) uniformly pronounced as [jɪs] by New Zealanders, it is striking that so few tokens are raised. This observation applies to other frequent words, such as *better*, *best*, and *then*. Of note as well is the categorical distribution of following liquids: both [l] (alveolar) and [ɬ] (palatal, also known as

dark) invariably take preceding [ɛ]. Since 111 tokens of following [l]/[ɬ] appear in the data, this finding is particularly prominent.

Unstressed tokens totalled 287, and stressed tokens totalled 417. Of these, 16 unstressed and 25 stressed tokens of [ɪ] were found, accounting for 5.6% and 6.0% of the data, respectively. This difference is very modest. Tokens are distributed across many words with no single word dominating the distribution. The most instances of [ɪ] occurred in the word *everything* with a frequency of five, all of which were stressed.

Table 3: Overall frequencies of front vowel monophthongs by *preceding* and *following* phonological environment.

Context	*Preceding*		*Following*	
	n [ɪ]	*n* [ɛ]	*n* [ɪ]	*n* [ɛ]
Stop/Affr	8	189	9	215
Nasal	12	65	17	185
Vowel	2	16	n/a	n/a
Fricative	6	158	15	146
[*h*]	0	33	n/a	n/a
Glide	3	110	0	2
[*j*]	3	26	0	2
Liquid	9	114	0	111
[*l*]	3	42	0	51
[*ɬ*]	0	0	0	60
rhotic	6	72	n/a	n/a
Word edge	1	9	0	2

5.1.2 Front vowel diphthongs

All instances of mid and high diphthongs were categorically [eə] and [iə], respectively. This finding is atypical, albeit it may be revealing of the underlying use of language on *Shortland Street*. Despite the smaller number of tokens found in the data ($N = 80$; see table 4), the total absence of raising is unlikely, especially given frequencies ranging between 50-80% in other contexts (Bayard 1995: 66, Gordon et al. 2004, Gordon and Maclagan 2001). In addition to the findings above, it appears that raising in mid-vowel diphthongs is being resisted by the speakers. This suggestion will be taken up in detail in the discussion section.

Table 4. Frequency of words containing diphthongs for Olivia (*O*) and Amanda (*A*). (Words containing underlying mid diphthongs are highlighted. *N* = 80.)

Frequency	*Tokens*	*O*	*A*
23	*Here*	17	6
16	*There*	7	9
14	*Where*	7	7
5	*Care*	3	2
3	*Year*	2	1
3	*Nearly*	1	2
2	*Cheer(/ed)*	2	0
2	*Hear*	2	0
2	*Anywhere*	1	1

Frequency	*Tokens*	*O*	*A*
2	*Somewhere*	0	1
1	*Nowhere*	0	2
1	*Carelessness*	0	1
1	*Hair*	0	1
1	*Despair*	0	1
1	*Nightmare*	0	1
1	*Fierce*	1	0
1	*Ear*	0	1
1	*Nearest*	0	1

5.2 Multi-variate analysis

In order to identify possible phonological patterns, *preceding* and *following phonological environments* were grouped into classes (*stop*, *fricative*, etc.), split between voiced and voiceless. For *following phonological environment*, several sonorants (*liquid*, *nasal*, *glide*) were accompanied by categorical use of the [ɛ] variant, thus rendering a variationist analysis unworkable. Hence, these factors were grouped together under one category, *sonorant*. Narrowing the factor of *preceding environment* to *sonorant* did not change the analysis.

There was a significant difference in raising between Olivia and Amanda (table 5). Note that the reported factor weights express a relative calculation between the two speakers: given a context where raising is possible, Olivia will be inclined to raise 69% of the time, and Amanda, 31% of the time. This result reflects both speakers' age and character, with Olivia (a young female) more likely to raise than Amanda (a female adult).

Table 5: Multi-variate analysis of front vowel monophthong raising (realization of [ɪ]) (*N* = 702).

Corrected mean: 0.046

Factor	**FW**	**% raising [ɪ]/([ɪ]+[ɛ])**	**$N_{[ɪ]+[ɛ]}$**
Speaker			
Olivia	0.69	9.7	338
Amanda	0.31	2.2	364
Range	38		

When represented in terms of into natural classes, *preceding* and *following phonological environment* do not condition raising. Instead, Olivia's and Amanda's speech appears to resist variation as innovative forms (the raised [ɪ]) are avoided.

Stress was not selected as significant, in contrast to Gordon and Maclagan's (2001: 229) findings. It is interesting to note that preliminary results

suggested that raising is more likely in unstressed position, again contradicting Gordon and Maclagan (ibid.) This possibility is explored in the *Discussion*.

Finally, Individual analysis of Olivia's tokens revealed no significant factors conditioning raising. An analysis of Amanda's, on the other hand, revealed an effect of *listener*. However, as discussed above, this finding is likely an effect of the skewed distribution and the few tokens of [ɪ] in Amanda's speech. For instance, in the 82 tokens of conversation with the character *Kip Denton*, no single instance of raising was recorded. This observation is even more striking given that the majority of these exchanges were emotionally charged, a situation which favours vernacular speech. Further evidence is required before concluding that Amanda raises selectively in conversation with these speakers.

6 Discussion

6.1 Overall findings

Sociolinguistic investigations of NZE have quantified the unique characteristics of its phonology and have illuminated the relationships between vowel realization and social variables such as age, sex, and social class. Raising in front mid vowels has increased considerably in the last 30 years such that today, in conversational speech, raising accounts for the majority of speakers' tokens. On the other hand, Olivia Tennet and Amanda Billing, when speaking on the New Zealand television series *Shortland Street*, appear to resist mid vowel raising as much as possible. Off-camera, Olivia speaks a typical NZE[1]. It seems unrealistic for her, then, to sound like an adult on *Shortland Street*. Furthermore, it seems unreasonable for both women to resist using NZE forms. This resistance is even more striking in the case of male characters, whose instances of raising are near-negligible. Although Labov's Principles aptly predict that men will raise less frequently than women, the near-absence of raising suggests that some external force is acting upon the linguistic system. What, then, is driving this resistance to raising?

To begin, it should be recognized that the medium of performance is both formal and highly crafted. Accordingly, not only will linguistic forms be close to Standard (Schilling-Estes 2004), but speech will be thoroughly rehearsed. As such, speakers become extremely aware of how they sound, and more importantly, how they are supposed to sound when they take on the role of a character. This dichotomy creates a tension between the use of vernacular "real speech" and formal "performance speech." In the latter context, careful attention to speech is reflected in a speaker's linguistic forms. For example, Olivia is often careful to avoid dark [ɬ] pronunciations, as seen in the high number of alveolar [l] in coda position (table 3). It is probable, then, that Olivia and Amanda are aware

[1] Olivia Tennet – actor in Othello. (2011, April 25). *Youtube*. Retrieved from http://www.youtube.com/ watch?v=36Fs-Gk_Pyo.

of raising in NZE and are actively avoiding using raised variants, resisting the vernacular realization as much as possible.

Because speech on *Shortland Street* is conveyed through performance, many possible explanations for resistance to raising can be explored, and others can be ruled out:

(i) Olivia and Amanda are established actors, and so, it is unlikely that they are avoiding raising for reasons of linguistic security.

(ii) If the producers of *Shortland Street* were mindful only of a New Zealand audience, it would be unlikely for the performers to use RP for purposes of conveying a standardized, "universally comprehensible" accent, specifically in light of NZE being largely homogeneous (Aside from some regions in the south, e.g. Southland; cf. Bayard 1995: 45).

(iii) As discussed above, New Zealand media has historically favoured the use of RP (Glass 2007). Apparently, this tradition is embedded so deeply in New Zealand televised media that actors on *Shortland Street* are trained to avoid raising and embody an RP style of speech.

(iv) Because *Shortland Street* is a drama, the performers will speak formally. Indeed, many of the actors on the series (including Olivia Tennet) have classical training and demonstrate this professional experience in their performance.

(v) Despite *Shortland Street* being a New Zealand production, it is also marketed and enjoyed internationally. During an interview with Canadian journalist George Stroumboulopoulos, famous Australian actor Hugh Jackman explained that unlike the United States, which exports vast amounts of entertainment media globally, countries like Australia do not have the resources to support such an industry[2]. Even though many blockbusters are filmed in New Zealand (e.g. *The Lord of the Rings* trilogy), the situation for homegrown entertainment is likely the same in New Zealand as it is in Australia. To make the series appealing for an international audience, it is possible that *Shortland Street* actors are pressured into avoiding raising in order to eliminate any possible barriers to communication (or, perhaps, any negative attitudes about NZE).

In all, many forces conspire in pressuring Olivia and Amanda to avoid raising. Even though their speech retains some flavour of NZE, it is heavily modified in order to meet the standards of a dramatic performance and the demands of an international market. In the end, then, it appears that NZE in televised media is not the same as NZE in conversational speech.

6.2 Empirical findings

Because she is a young woman, it is expected that Olivia will push the forefront of change further than Amanda. Furthermore, despite Olivia's lower frequency of monophthong raising (9.7%) as opposed to the frequencies witnessed elsewhere (50-80%; see above), her probability of raising is higher compared to Amanda's

[2] George Tonight: Hugh Jackman. (2011, Oct. 12). *Youtube*. http://www.youtube.com/watch?v=SnFI2y1d7hw&feature=results_main&playnext=1&list=PLEA4C37686FEA2A2E. At 7:40.

(0.69 against 0.31, respectively). This result indicates that, despite the apparent resistance to raising observed in this sample, the amount of raising in Olivia's phonology is still relevant. On the other hand, when taken in isolation, Olivia is not likely to use raised variants. Instead, her phonology comes off as unexpected and inappropriate for her age. Performance style is thus masking Olivia's vernacular.

Gordon and Maclagan (2001: 229) argue that unstressed positions are more likely to take an [ɛ] realization. The authors propose (ibid.) that because unstressed positions are more likely to take reduced pronunciation, the realization is more likely to be the open variant, [ɛ]. In the case of Olivia and Amanda, when raising "slips" into speech, one would expect it to occur in a position which takes less focus; specifically, the unstressed position. This behaviour would provide further evidence that speakers on *Shortland Street* are resisting raising and are hyper-aware of their speech. Further investigation is required, however, since (i) it has been shown that performance speech is not the same as vernacular, and since (ii) the proportion of unstressed and stressed [ɪ] tokens is largely the same (see sec. 6.1).

7 Conclusion

In this paper, I investigated the rate of front vowel raising in the speech of Olivia Tennet and Amanda Billing on the television program *Shortland Street*. Because *Shortland Street* is produced in New Zealand and stars New Zealanders, it was possible that Olivia and Amanda's speech would exhibit front vowel raising, a characteristic of NZE which has developed rapidly in the past century. In the end, Olivia and Amanda showed moderate monophthong raising (approximately 6%) and categorical diphthong realization. Hence, *Shortland Street* is not an accurate portrayal of NZE; its unique aspects are actively suppressed in a performance, where many factors pressure a speaker to augment his/her phonological choices. However, the results also suggest that Olivia and Amanda must be hyper-aware of their own vernacular, and that occasionally, differences in each actor's rate of raising can emerge.

It would be revealing to explore how performers sound when reading a script for the first time compared to the final product. From this perspective, the amount of augmentation to speech can be calculated and examined. Furthermore, the results from *Shortland Street* should be compared to the rate of raising found in other broadcast material, such as different genres of television (e.g. comedy, talk show, newscast), stage, and radio (e.g. impromptu call-in programs). Even though phonological merging in NZE has yet to expand into televised media, we have an exciting opportunity to witness this development from the very beginning.

Acknowledgements

I would like to thank Jeff Tennant for supervision and feedback as well as Alex D'Arcy.

References

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society, 39*, 437-456.

Bayard, D. (1995). *Kiwitalk: Sociolinguistics and New Zealand Society*. Palmerston, NZ: Dunmore Press Ltd.

Cheshire, J. (2004). Sex and gender in variationist research. *The Handbook of Language Variation and Change*. Chambers, J.K., Trudgill, P., & Schilling-Estes, N., (Ed.). Hoboken, NJ: Blackwell Publishing, 423-443.

Feagin, C. (2004). Entering the community: Fieldwork. *The Handbook of Language Variation and Change*. Chambers, J.K., Trudgill, P., & Schilling-Estes, N., (Ed.). Hoboken, NJ: Blackwell Publishing, 20-39.

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequences. *Contexts of Accommodation*. Giles, H., Coupland, J., & Coupland, N., (Ed.) Cambridge, NY: Cambridge University Press, 1-68.

Glass, L. (2007, May 24). *One News* [Television broadcast]. Wellington, NZ: TVNZ. http://www.youtube.com/watch?v=0ePwKYJcEOo, first accessed Sept. 26, 2011.

Gordon, E., & Maclagan, M. (2001). 'Capturing a sound change': A real time study over 15 years of the NEAR/SQUARE diphthong merger in New Zealand English. *Australian Journal of linguistics, 21*(2), 215-238.

Gordon, E., Campbell, L., Hay, J., Maclagan, M., Sudbury, A., & Trudgill, P. (2004). *New Zealand English: Its Origins and Evolution*. Cambridge, UK: Cambridge University Press.

Maclagan, M. & Hay, J. (2007). Getting *fed* up with our *feet*: Contrast maintenance and the New Zealand English "short" front vowel shift. *Language Variation and Change, 19*, 1-25.

Sankoff, D., Tagliamonte, S., & Smith, E. (2005). Goldvarb X: A variable rule application for Macintosh and Windows. Department of Linguistics, University of Toronto.

Schilling-Estes, N. (2004). Investigating stylistic variation. *The Handbook of Language Variation and Change*. Chambers, J.K., Trudgill, P., & Schilling-Estes, N., (Ed.). Hoboken, NJ: Blackwell Publishing, 375-401.

Tagliamonte, S. & Roberts, C. (2005). *So* weird; *so* cool; *so* innovative: The use of intensifiers in the television series *Friends*. *American Speech, 80*(3): 280-300.

Woods, N.J. (1997). The formation and development of New Zealand English: Interaction of gender-related variation and linguistic change. *Journal of Sociolinguistics, 1*(1), 95-125.

Appendix A

Table A1: Front vowel monophthong raising by listener. Displayed are the proportion and frequency of [ɪ], the raised variant of the mid-front vowel monophthong. (Complement to fig. 2 and 3)

Listener	*Olivia*		
	% [ɪ]	N [ɪ]	N_{tot}
Harry	25	1	4
Tane	20	3	15
Callum	16.7	1	6
Wiremu	13.3	2	15
Chris	12.7	8	63
Guy	10	11	99
Angus	9.4	3	32
Sophie	7	2	29
Toni	3	2	67
		33	

Listener	*Amanda*		
	% [ɪ]	N [ɪ]	N_{tot}
Wiremu	13.6	2	17
Guy	11	1	8
Justine	9.1	1	11
Chris	5	1	20
TK	3.8	3	78
		8	

On being *happier* but not *more happy*: Comparative alternation in speech data

Alexandra D'Arcy
University of Victoria
adarcy@uvic.ca

English adjective comparison is increasingly the focus of corpus linguistic research, but it is much less studied in the variationist framework. These two traditions converge, however, in revealing robust variation between historical inflection (*happier/happiest*) and newer periphrasis (*more/most happy*). However, our understanding of the strategies for comparison comes from written genres. In contrast, very little is known about comparison in vernacular speech. Since periphrastic comparison emerged as a change from above, the lack of spoken evidence proves a critical gap in our knowledge. To address this gap, this paper examines comparison strategies in New Zealand English, drawing on the whole of the Origins of New Zealand English Archive (Gordon et al. 2007). Analysis of 1400 tokens reveals a striking result. Consistent with reports elsewhere, inflection is the preferred mode of comparison. However, consideration by lexical item reveals a system that is not, in fact, variable. Rather, across the history of this variety (speakers born 1851-1982), individual adjectives pattern one way (inflection) or the other (periphrasis); in speech, the form of comparison has consistently been lexically conditioned, and by extension, invariant. This paper explores a number of explanations (e.g. variation is genre-specific or variety-specific, or may only be visible in extremely large corpora), and ultimately concludes that in speech, historical variation resulted in the full 'regularization of a confused situation' (Bauer 1994:60).
Keywords: adjective comparison; vernacular speech; variation; New Zealand English

1 A showcase of grammatical variation

By all accounts, comparative alternation in English is robust and long established as a variable feature of the language. Historical inflection, as in (1a,c) competes with 'innovative' periphrasis (1b,c), a layering of forms that has been attested since the Middle English period (Pound 1901; Mitchell 1985; Kytö 1996). This competition may also result in the combination of the two strategies, creating a hybrid such as is exemplified in (1d). This final strategy, an unmitigated minority option in Present Day English, was in Middle and Early Modern English a viable option for comparison, both comparative and superlative (Kytö 1996:124; see also Schlüter 2001; Kytö & Romaine 2006).

(1) a. *Subtler* and more successful methods are sometimes tried. (WWC/G45/124)[1]
b. Sondra ...was *more subtle*. (WWC/K87/121)
c. It was the *most silly*-- it was the *silliest* thing I've ever heard of. (V. Sheehy, b.1896)
d. You know the *most silliest* things you know. (fyn01-5b)

Although linguistically (i.e. referentially) equivalent, however, inflection and periphrasis are not in all instances socially (i.e. ideologically) equivalent. This lack of parity is most clear in the context of second language learning, but it also finds voice in less formalized metadiscourse, both online and otherwise (e.g. Urban Dictionary). While the former likely arises from pedagogical imperatives to teach 'standard' language practice, the latter is rooted in the nuances of communicative competence, determined by community-internal norms and sociolinguistic cultures. Regardless, it is clear that adjective comparison is an area of the grammar that is rife with notions of 'right' and 'wrong'.

Consider what seems a straightforward example, an adjective such as *happy*. On a language forum (WordReference.com; November 27 2007), use of the periphrastic comparative in 'They are more happy than the mean people' spurred two threads, both of which derided the construction in favour of inflection. Ultimately, *more happy* was deemed colloquial and better avoided in a range of formal situations. An English second language speaker concludes one of the threads with the assertion that 'for someone whose english is not the mother tongue, […] it's worth not using it.' On a different forum (*Yahoo!*Answers; January 2011), a user asked which of the two options in (2) 'make [sic] more sense'. The answer, prescriptive in tone, refutes (2b): both 'make sense', but *happier* is 'correct' and *more happy* 'is not'.

(2) a. I'm happier than ever before.
b. I'm more happy than ever before.

Metadiscourse of this kind is particularly insightful from a sociolinguistic perspective. As with other rule-based, prescriptive ideologies of language usage, commentary typically is not reflected in practice (see also D'Arcy & Tagliamonte 2010). Thus, despite overt notions of correct and incorrect usage (e.g. *when a word ends with <y>, drop the <y> and add <ier>*), a simple Google search quickly reveals not only a plethora of variation, as illustrated by the examples in

[1] The parenthetical information following examples from corpora provides details of providence. For examples from the Wellington Corpus of Written New Zealand English (WWC), the category and transcript number and the individual line number are given; for examples from the Origins of New Zealand English Archive (ONZE), the speaker details are provided. This takes the format of either a name or a code, depending on the protocol of the collection from which the example is drawn.

(3), but also creative and seemingly willful manipulation of the choice between inflection and periphrasis.

(3) a. The first step to becoming *more happy* is to realize you don't know how to make yourself *happier*. (Happy Deviant; Positive psychology: How to construct a happier life – Part 1; May 15 2011)

b. When asked if they'd like to be *more happy* the majority of people shout 'yes' because 'who wouldn't want to be *happier*?'' (Jodi Lee – Life Designer; Can you recognize what makes you happy?; November 5 2010)

c. *Happier* people tend to be healthier. The *more happy* you feel, the less prone to illness you will be. (Anti-Aging Web Magazine; How growing old makes you happier!)

d. I was happy with my stock '08 SG […] added a 6" long ride shield and was *more happy* […] added aileron grips and again I was *more happy* than ever […]. If I get any *happier* I don't think I'll be able to live with myself. (HDForums.com; Happy to be more happy (happier?); November 15 2009)

The examples in (3), all of which evince variation within connected stretches of writing, provide illustration of the robust nature of the choice mechanism that is operative in adjective comparison. They also illustrate the prosaic possibilities that this variation enables. In (3c), for example, the deployment of both the inflectional and the periphrastic constructions derives in each case from the stylistic drive for structural symmetry (*happier…healthier*; *the more…the less*). In other words, the language itself allows overt manipulation of comparative forms when it would be entirely possible (indeed, 'correct') to use only inflectional comparison for adjectives such as *happy*. To this end, the example in (3d) is striking not only for the variation in the text but also for the title of the post, which openly plays with the alternation between comparative strategies. As with (3c), the author opts for symmetry (in this case, bracketing the title with identical elements, *happy…happy*) while at the same time highlighting the fact that another option exists.

We therefore have a situation whereby adjective comparison is the subject of metalinguistic discussion among language learners, grammarians and language mavens, and the different strategies can be overtly manipulated to achieve a range of stylistic and pragmatic effects. Linguists tend to target known features, those that are treated in language grammars and by other researchers (Cheshire 1999). It therefore follows that a large body of academic work targets comparison. From a paradigmatic perspective, much of this is based in the corpus linguistic tradition (e.g. Bauer 1994; Kytö 1996; Kytö & Romaine 1997, 2000, 2006; Leech & Culpepper 1997; Mondorf 2003; etc.), the results of which suggest that comparative alternation is a robust and well-attested phenomenon. If

scholarly output is any evidence, however, then it is clear that variationists have been much less keen to investigate this feature (cf. Hilpert 2008; Scrivner 2010).

That the variationist tradition has not weighed in on comparative alternation is striking. By all accounts, variation between inflection and periphrasis is highly constrained. All levels of linguistic analysis are implicated (phonology, morphology, syntax, semantics, pragmatics, the lexicon, etc.), a fact which led Mondorf (2009:1) to describe this feature as 'a showcase of grammatical variation'. In sum, comparative alternation presents a complex phenomenon, and so it seems (intuitively anyway) to be an ideal candidate for variationist analysis, particularly if one is interested in mechanisms and pathways of change.

2 Historical perspective

Competition between inflection and periphrasis has been operative in English comparison since roughly the thirteenth century (Pound 1901; Mitchell 1985; Kytö 1996; Kytö & Romaine 1997, 2000, 2006). In other words, the spread of periphrasis within the sector constitutes a longstanding change with modern-day reflexes. However, the diachronic picture is atypical because adjective comparison does not exhibit the normal trajectory of replacement (i.e. A > B).

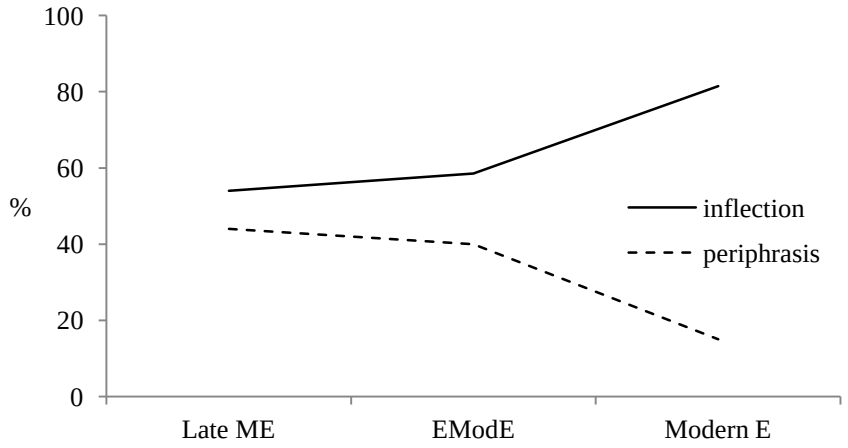

Figure 1. Historical trajectories of inflection and periphrasis, comparative (from Kytö & Romaine 1997:336, Fig.1)

Historically, the periphrastic forms are innovations; Old English of course was inflectional. When the periphrastic option for comparison first developed, it diffused relatively slowly until approximately the end of the fourteenth century, when it began to increase quite steadily (Mitchell 1985). This initial trajectory follows the established template for innovations, which generally entail introduction, incipient variation, and subsequent increase (i.e. the S-curve of

linguistic change; Weinreich et al. 1968; Bailey 1973; Altmann et al. 1983; Kroch 1989; Labov 1994; though see Denison 2003). However, periphrasis did not continue to spread throughout the sector, systematically pushing out or marginalizing inflection. Instead, periphrasis seems to have peaked during the Late Middle English period (Pound 1901); since that point, as illustrated in Figure 1, inflection has been busily 'reasserting itself' (Kytö & Romaine 2000:172). Contemporaneously, the majority of comparison (both comparative and superlative) is inflectional, with present day distributions having been achieved during the Late Modern English period (i.e. post 1710; Kytö & Romaine 1997).

The vast majority of our understanding concerning the history, development, and patterning of comparative alternation in English is based on evidence from written genres. From a historical perspective, this is perfectly valid; it is also a pragmatic necessity. However, it is also important to consider the mechanism by which periphrasis entered the field of comparison. The answer is particularly germane to the question of variation and change, since changes from above (adaptive) have sociolinguistic properties that are distinct from those associated with changes from below (evolutive).

Despite the evolutionary tendency for English to be moving from synthetic to periphrastic syntax, it has been suggested that periphrastic comparison is not a language-internal development. Instead, it likely emerged under the prestige influence of Latin primarily, but also, to a lesser extent, French (Mustanoja 1960:279). In other words, periphrasis in this instance is a historical change from above. In this case then, what happens in speech —and in unscripted, casual, vernacular speech in particular— presents a potentially rich source of empirical evidence to add to our understanding of this grammatical feature.

3 Data and method

The data for this study come from the Origins of New Zealand English Archive (ONZE), one of the largest repositories of longitudinal spoken English data available for sociolinguistic analysis. Consisting of three collections (the Mobile Unit, the Intermediate Archive, and the Canterbury Corpus), ONZE includes over 1000 hours of casual speech from more than 700 individuals, covering the history of New Zealand English, 1850 to the present (for full details, see Gordon et al. 2004 and Gordon et al. 2007).

Following the *principle of accountability* (Labov 1966, 1972), the cornerstone of variationist methodology, the recordings were exhaustively searched and all instances of comparison (comparative and superlative) were extracted. All told, these materials include 2621 tokens, 1221 suppletive (e.g. *good*, *better*, *best*) and 1400 non-suppletive (e.g. *nice*, *nicer ~ more nice*, *nicest ~ most nice*). The focus of this analysis is the non-suppletive group, as this is the site of variation; suppletive comparison is invariant.

The details of the analysis are provided in Table 1, which provides the breakdown by collection within the Archive.

Table 1. ONZE sample and data (non-suppletive adjectives only)

corpus	*speaker years of birth*	*N speakers*	*N tokens*
Mobile Unit	1860-1919	32	197
Intermediate	1891-1963	56	449
Canterbury Corpus	1922-1982	151	754
total N		239	1400

4 Results

In terms of overall results, the data conform to expectation in that the majority of comparison is inflectional (68.5%). They are also typical in that inflection is more frequent in the comparative (71.9%, N = 983) than it is in the superlative (60.4%, N = 417). Moreover, even though the data represent colloquial speech, hybrid forms such as (1d) are exceptionally rare: They account for less than 1% of the data overall (0.79%; N = 11).

The picture presented by these aggregate results is thus consistent with that presented by analyses based on written data. During the Modern Period, comparison is largely (but far from exclusively) an inflectional phenomenon, and conflation of the strategies is rare. However, given that the ONZE materials span more than a century in apparent time, and because they capture the history of the variety, they have the potential to be insightful with respect to the establishment of the present-day New Zealand system.

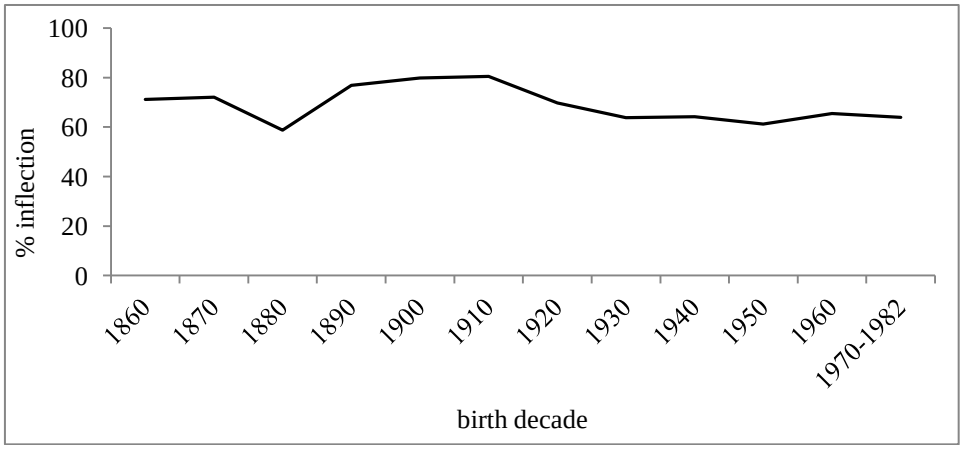

Figure 2. Overall distribution of inflectional comparison across time in ONZE

If the temporal dimension of the data is examined, as in Figure 2, the distributional workload appears stable across time. Indeed, there is no correlation between time (as a factor of speaker birth year) and the overall frequency of inflection in these data (r = –0.325632, p = .167739). This result strongly suggests that comparative alternation was fixed prior to the formative period of New Zealand English; the variety underwent no distributional re-organization

subsequent to permanent British settlement in the colony. Thus, despite the diachronic depth of these speech data, time is not implicated in the variation whatsoever. As it happens, there is another critical respect in which this is the case as well. This point will be illustrated shortly. First, however, there are grammar-internal factors to explore, factors which may constrain the degree of comparative variation that is possible within any given dataset.

A key aspect of adjective comparison is that the window for variation is fairly circumscribed: Not all adjectives may be compared using both inflection and periphrais. This is because over time, the modes have specialized in certain linguistic environments. Ultimately, periphrasis has successfully ousted the older inflectional means from some contexts, but in others it is the inflectional type that has triumphed. For example, adjectives consisting of four or more syllables categorically require periphrasis (e.g. *academic*, *democratic*, *profitable*, *apologetic*, *enthusiastic*, *sophisticated*). The same is largely true of trisyllabic adjectives as well. Thus, a form such as *important* can only be modified as *more* or *most important*; **importanter* is not grammatical (descriptively) in native, unmonitored, adult language. This is quite distinct from the behaviour of adjectives such as *happy*, where both modes of comparison are possible and 'grammaticality' is a matter of prescription (i.e. the structure of the language allows variation).

This kind of categorical patterning has ramifications for variationist analysis, since the aim is to focus on those forms for which variation is possible. In the ONZE materials, there are more than 130 types of multi-syllabic adjectives that categorically take periphrasis to mark comparison. Such tokens are thus of little interest to a discussion of variation.

Where the bulk of variation putatively occurs is among bisyllable adjectives (e.g. *bolshy*, *clever*, *deadly*, *mature*, *narrow*, *pleasant*, *quiet*, *recent*, *wealthy*, *vivid*, *yellow*, etc.). Certainly with forms such as *bolshy*, *clever*, and *deadly*, inflection or periphrasis are both acceptable. However, there are forms that cannot vary. With adjectives such as *alive*, *carefree*, *complex*, *correct*, *human*, *nervous*, *open*, *passive*, *peaceful*, *private*, *senior*, *unfair*, *unjust*, *upset*, and *useful*, for example, inflection is not possible. As with multi-syllabic adjectives, the sole grammatical mode of comparison is periphrasis. What is particularly notable about the exceptional forms in the bisyllabic category, however, is that structurally (syllabically and segmentally), they are not distinct from those that are variable: As outlined in Figure 3, a large proportion of coda or final segments (a putative condition for inflection or periphrasis, cf. /–i/) overlaps the categories. This is important because it means that the determinants of (non) variability are not predictable. Having invariant adjectives also means that the window of variation is further restricted: The focus for quantitative modelling are the forms that can alternate, not those that cannot and which, by extension, do not.

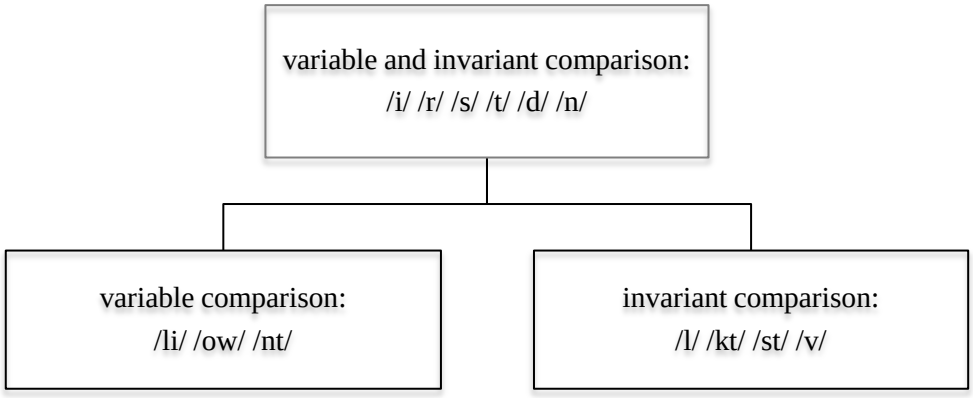

Figure 3. Coda or final segments of bi-syllabic adjectives in ONZE, according to mode of comparison

The consequence of categoricity is that tokens for which comparison is achieved strictly via inflection or strictly via periphrasis must be excluded from the analysis. In the ONZE materials, this category contains 838 tokens. Put in broader perspective, invariant comparison accounts for 60% of non-suppletive adjectives in the dataset. That is a huge proportion of forms that are simply not able to evince variation. Moreover, of the initial 1400 tokens, only 562 tokens remain in the analysis. These forms can alternate—at least, in principle they can.

As it happens, in these unscripted, casual speech data, there are contexts which, despite being *able* to vary, do not. Most notable among these are trisyllabic adjectives. In the normal case these are compared using *more* or *most*, but there are certain adjectives (e.g. *beautiful*, *dangerous*, *expensive*) for which cases of inflection are imaginable and acceptable. Indeed, such instances are easily located on the Web (e.g. 'The online world is becoming *dangerouser* and *dangerouser*', http://s1.zetaboards.com/H_P_Hummingbird/topic/3809578/1/). In speech-based ONZE, however, comparison of trisyllabic adjectives is strictly periphrastic.

At the other end of the syllable scale are monosyllabic adjectives; the 'rule' for these is that comparison is inflectional. Of course, there is a huge number of single-syllable adjectives that take variable comparison. Consider just the *b*s: *blue*, *blunt*, *bold*, *brash*, *brave*, *bright*, *broad*, etc. Despite being able to host both inflectional and periphrastic comparison, in ONZE these forms are compared 'as they should be', that is, inflectionally. Indeed, at over 95% (N = 427/449), the rate is nearly categorical. There is simply no variation in this group either.

Monosyllabic adjectives are also the largest syllable type within the dataset; they account for almost 80% of the remaining data. There is no reason to suspect that this distribution is atypical and that other analyses have not largely been based on monosyllabic forms as well. If such is the case, then it is these forms that are driving the higher overall rates of inflection that are regularly reported in the literature (e.g. Kytö & Romaine 1997, 2000). Note that this generalization holds within the full ONZE dataset as well, where monosyllabic

adjective comprise 68.7% of tokens (N = 962/1400) and 98.2% of inflection (N = 942/959). Moreover, the New Zealand results suggest that the nearly categorical status of inflectional comparison on monosyllabic adjectives is well entrenched: There is zero variability across time in ONZE. *Cheap* is always *cheaper*, never *more cheap*. *Dark* is always *darker*, *fast* is always *faster*, *hard* is always *harder*. This is a dataset that extends to the mid 19th century. For a form such as *cheap*, for example, there are tokens from speakers born in the period from 1893 through to 1980 but there is not one instance of periphrastic comparison (N = 45); *cheaper* or *cheapest* are the only forms attested.

In fact, once individual forms are considered, it quickly becomes apparent that there is no variation within lexical items in ONZE. The complete list of forms that exhibit variable comparison is given in (4): *clever* (N = 3), *common* (N = 2), *cool* (N = 8) and *silly* (N = 3). Of these, only two vary within a paradigm: *clever* in the comparative (4a) and *silly* in the superlative (4d). With the exception of (4d) (a correction), none varies either within a single utterance or within the speech of any single individual. In other words, there is little evidence that this variation is regular or systematic (i.e. it may be idiosyncratic).

(4) a. They were *cleverer* than we were. (W. Oliver, b.1907)
I was *more clever* at that than anything else. (A. Shacklock b.1891)
b. The *commonest* type of nickname they got was... (J. Marin, b.1900)
It seems to be getting more and *more common*. (L. Algie, b.1926)
c. Sam was a little *more cool*. (mop94-4)
South Intermediate, which was the *coolest* school ever... (fyn00-7)
d. It was the *most silly*-- it was the *silliest* thing I've ever heard of.
(V. Sheehy, b.1896)

Given that lexical items do not vary in this dataset, the complex, interwoven net of constraints that has been reported in the literature becomes problematic. Syllable structure, phonological structure of the root-final coda, stress, haplology, complement type, semantic status, and end-weight are just some of the constraints that have been reported to operate on comparative alternation (Kytö 1997; Leech & Culpepper 1997; Lindquist 1998, 2000; Mondorf 2003, 2009; Hilpert 2008, etc.). However, where these studies report robust variation, the ONZE materials do not provide any substantive evidence for it. These factors are thus neither relevant nor applicable in ONZE.

5 Resolving an analytical conundrum

These materials thus present an analytical conundrum. Contrary to other analyses of English comparative alternation, individual adjectives are invariant in their comparative strategy. This is consistent across the ONZE Archive, in both temporal and social space. There are three possible explanations: variety, corpus size, and genre or register.

5.1 Variety

The current data are from New Zealand English. Other research is based on either British English or American English. It is possible that New Zealand English is distinct from other World Englishes —and from the two primary Inner Circle varieties in particular— in not exhibiting (robust) comparative alternation. As the youngest of the colonial varieties, for example, New Zealand English has most recently undergone dialect formation. That such is the case allows for the possibility that when it coalesced as a distinct regional variety, the variation that characterizes adjective comparison in British English levelled (on *levelling* in new dialect formation, see Trudgill 2004).

One way to test for the role of language variety is to examine another corpus of New Zealand English data. One such source is the Wellington Corpus of Written New Zealand English (WWC; Bauer 1993). This is a 1-million-word corpus, modelled on the Brown and Lancaster-Oslo-Bergen (LOB) corpora; it contains 500 2,000-word samples from a range of written genres.

The most comprehensive variationist analysis of English comparative alternation is that of Hilpert (2008), who drew on the British National Corpus (BNC). Hilpert provided a list of 247 adjective types that alternated in the BNC materials. Using that list, the WWC was systematically searched, resulting in 1149 tokens of adjective comparison (comparative and superlative), compiled from 113 adjective types. The data are qualitatively different from those extracted from ONZE: They are variable at the level of individual lexical items— in some cases, quite robustly.

Table 2 presents a representative sampling of the WWC data in terms of raw occurrences (right), including a comparison with Hilpert's (2008) BNC findings (left). Most notable about the comparison is the similarity between the two sets of results with respect to the tendencies exhibited by individual forms.

The first five adjectives in the table occur in relatively high numbers in the WWC, and they are invariant. As such, it is tempting to conclude that New Zealand English is exceptional with respect to adjective comparison (i.e. that variation is not a feature of the variety). But such a conclusion would be premature. The token counts for these particular adjectives are extremely high in the BNC. The least frequent has a raw occurrence rate of over 1500. Of these adjectives, however, the most 'robustly' variable is *broad*, which takes periphrasis at a rate of just 0.4%. For all intents and purposes, these adjectives— despite being categorized as variable— pattern categorically. In other words, for forms such as these, variation is highly infrequent and exceptional.

In contrast, variation is very much in evidence for the forms illustrated in the bottom half of Table 2, both in the BNC and, most notably, in the WWC. Moreover, the New Zealand data exhibit variation despite the low frequencies at which these individual forms are attested in the WWC. This result contrasts starkly with that from ONZE, where these very same adjectives are categorically inflected (i.e. invariant).

Table 2. A sample comparison of results from the BNC and the WWC

	BNC		WWC	
Adjective	*-er*	*more*	*-er*	*more*
big	4466	1	39	0
broad	1588	7	12	0
great	15936	1	188	0
hard	1745	1	30	0
small	8816	5	62	0
just	9	8	2	1
mature	14	141	1	2
rare	231	22	3	1
shallow	125	6	6	2
simple	1115	60	8	1
stable	9	86	1	1
subtle	114	339	4	4
sweet	157	2	2	1

To summarize, the results once again present a confound. On the one hand, ONZE provides data from over 120 years of spoken New Zealand English and no variation at the level of the lexicon is evident: Individual adjectives pattern one way (inflection) or the other (periphrasis). On the other hand, the WWC provides data from just 4 years of written New Zealand English and among individual adjectives there is fairly robust variation (e.g., *rare*, *shallow*: 75% inflection; *sweet*, 66% inflection; *subtle*, 50% inflection).

This finding *ipso facto* rules out variety. New Zealand English does not have special status with respect to adjective comparison. Written New Zealand data exhibit the same variation as do written British data and written American data.

5.2 Corpus size

The next hypothesis to consider in explaining the anomalous New Zealand findings concerns corpus size. ONZE, while one of the largest extant English speech corpora, is nonetheless a significantly smaller corpus than what is generally used when investigating adjective comparison (the BNC, for example, contains 100 million words). Although 1400 non-suppletive tokens were initially extracted from ONZE, perhaps thousands more are required to uncover variation. This explanation, however, is not supported by the evidence. At 1 million words, the WWC is but one-hundredth the size of the BNC. Despite being radically smaller than the BNC, the WWC provides clear evidence of variability. It therefore cannot be that ONZE is too small to capture variation. There are over 1000 hours of unscripted speech in the Archive, and, at over 1.4 million words, ONZE is in fact a considerably larger corpus than is the WWC.

5.3 Genre

The remaining difference between the ONZE data and the data used in the vast majority of research on English comparative alternation concerns genre. ONZE is a spoken corpus, which is part of what makes it such a rich resource in English historical sociolinguistics. With one notable exception, to which I return shortly, all other work on comparative alternation has considered written evidence. Some details are provided in Table 3.

Some of these works provide a list of the adjectives used in the analyses; many do not. Hilpert's (2008) contribution is particularly valuable because it includes not only a listing of the token types but also a break-down of these types by comparative mode, as either inflectional or periphrastic. As discussed above, many of the adjectives that were included in his analysis as variable were, at best, only marginally so (standard variationist methodology sets the cut-off point for categorical behaviour at ≤ 95.0%; see Guy 1988). As it happens, however, this is particularly true of the data used in Scrivner (2010). This is a critical point, because Scrivner (2010) is the only other study to have focused exclusively on speech. Her data were drawn from the spoken component of the Corpus of Contemporary American English (COCA), specifically, from those sections that contained unscripted speech. These consisted of talk shows and news programs (e.g., *Good Morning America*, *The Today Show*, *60 Minutes*, etc.).

Table 3. English datasets used in previous analyses of comparative alternation

Work	Corpus
Bauer 1994	The Times, The New York Times
Kytö 1996	Helsinki Corpus
Kytö & Romaine 1997	ARCHER, BNC
Kytö & Romaine 2000	Corpus of Early American Texts, ARCHER
Kytö & Romaine 2006	CONCE
Mondorf 2009	newspapers, fiction, BNC
Hilpert 2008	BNC
Scrivner 2010	COCA

Identical to the methodology used by Hilpert (2008) and following standard variationist practice, Scrivner (2010) included only adjectives that alternated modes of comparison. In the spoken subset of COCA mined by Scrivner, there were 90 such types. However, of these 90 variable adjectives, a full 43% (representing 39 types) exhibited variability at levels below 5%. In fact, most were variable well below this level. Crucially, this is not an issue of token numbers. The least frequent of these adjectives occurs 33 times, but most of the virtually categorical forms occur at frequencies in the hundreds and even thousands.

Within the variationist paradigm, distributions above the 95%/5% threshold are considered 'nearly categorical', 'exceptional', and 'statistically problematic' (see, e.g., Guy 1988). For this reason it is standard practice to remove them from

the quantitative model. The concern here, however, is not statistical validity but something much more fundamental. A non-trivial proportion of Scrivner's data is 'virtually categorical'; variation is hard to find. What I would like to suggest, therefore, is that variation in adjective comparison is crucially affected by genre. In writing it remains a robust phenomenon, but in speech, it is marginalized. Such a hypothesis accounts not only for the marked differences between the ONZE results and those of previous analysis of comparative alternation in English, it also accounts for the stark contrast between the findings from ONZE and those from the WWC.

6 Discussion and conclusion

Speaking of comparative alternation, Bauer (1994:60) once suggested that 'change' in the 20th century was not a matter of strategy *per se* but of the 'regularization of a confused situation'. Ultimately, comparison became 'more predictable': Periphrasis and inflection specialized. The added insight provided by the ONZE data (and, arguably, by the COCA data as well) is that nowhere is this more apparent than in speech.

In the normal, unmarked case —in speech as in writing— adjectives with four or more syllables require periphrasis. In speech, however, this is extended to adjectives with three syllables. The category of unmarked comparison is also extended in speech to include inflection on monosyllabic forms.

What of bisyllabic adjectives? These are the bastions of variation in writing, where stylistic factors such as symmetry and word play, pragmatic factors such as the desire to be witty, and structural factors such as syntactic, semantic, and pragmatic complexity all exert their effects. This is uncontentious. In speech, however, comparison has regularized. What is notable about bisyllabic adjectives is that while generalizations can be made about phonological effects (for example, final /l/ favours periphrastic comparison), these generalizations do not apply across the board. Exceptions exist (e.g. *narrower* vs. *more mellow*). This is strongly suggestive that the effects are lexical (likely deriving from frequency effects; see, e.g., Braun 1982; Quirk et al. 1985; Hilpert 2008). They are not predictable on structural grounds.

Indeed, this raises the question, unasked before now, of the extent to which speech was ever variable. The answer is unclear, but ONZE, a rare source of diachronic evidence, provides potentially rich insight. As summarized by McCarthy (1991:143*f*):

> We do not know enough about the acceptable norms of grammar in speech since, up to now, our grammar books have been largely formulated from introspective and written data. A good grammar of spoken English might well contain a few surprises.

The written language may have always been more variable, particularly if periphrastic comparison entered as a change from above. To that end, consider

another well-known case of a prestige borrowing, the *wh-* relative pronouns. Romaine (1982:212) made the now famous observation that 'the infiltration of *wh-* … can be seen as completed in the modern written language … but it has not really affected the spoken language.' A similar argument could be made for comparative alternation. In this case, it is not that the periphrastic forms have not infiltrated speech. Clearly they have. Rather, for speech the options have specialized, conditioned by individual adjectives. The fall-out from this change from above remains visible in writing, where variation abounds for mono- and bisyllabic adjectives in particular, but it is much less of a factor for these same forms in speech.

To conclude, the possibility for variation in adjective comparison is not ruled out in speech. Instead, I would like to suggest that it be recognized for what it is: The exception rather than the norm, a locus of variation that is not, in fact, particularly variable. In writing it is possible to be *more happy*, but in speech, it is generally the case that *happier* is the way to be.

Acknowledgements

I gratefully acknowledge the members of the Origins and Development of New Zealand English team at the University of Canterbury, who prepared, transcribed, and time-aligned the ONZE materials. I thank Alesia Malec, my Research Assistant on this project, whose position was funded by the University of Victoria Work Study Program (Student Awards and Financial Aid). Appreciation is also extended to Julie Auger, Heidi Quinn, and Olga Scrivner for their help at various stages in the conduct of this research.

References

Altmann, G., von Buttlar, H., Rott, W. & Strauß, U. (1983). A law of change in language. In B. Brainerd (Ed.), *Historical linguistics* (104-115). Bochum: Brockmeyer.

Bailey, C.J. (1973). *Variation and linguistic theory*. Washington, DC: Centre for Applied Linguistics.

Bauer, L. (1993). *Manual of information to accompany the Wellington Corpus of Written New Zealand English*. Wellington: Department of Linguistics, Victoria University of Wellington.

Bauer, L. (1994). *Watching English change*. Harlow: Longman.

Braun, A. (1982). *Studien zur Syntax und Morphologie der Steigerungsformen im Englischen*. Bern: Francke.

Cheshire, J. (1999). Taming the vernacular: Some repercussions for the study of syntactic variation and spoken grammar. *Cuadernos de Filologia Inglesa*, 8, 59-80.

D'Arcy, A. & Tagliamonte, S.A. (2010). The prestige legacy of relative *who*. *Language in Society*, 39(3), 389-410.

Denison, D. (2003). Log(ist)ic and simplistic S-curves. In R. Hickey (Ed.), *Motives for language change* (54-70). Cambridge: Cambridge University Press.

Gordon, E., Campbell, L., Hay, J., Maclagan, M., Sudbury, A. & Trudgill, P. (2004). *New Zealand English: Its origins and evolution*. Cambridge: Cambridge University Press.

Gordon, E., Hay, J. & Maclagan, M. (2007). The ONZE corpus. In J.C. Beal, K.P. Corrigan & H.L. Moisl (Eds.) *Creating and digitizing language corpora*, Vol. 2 (82-104). New York: Palgrave Macmillan.

Guy, G. (1988). Advanced Varbrul analysis. In K. Ferrara, B. Brown, K. Walters & J. Baugh (Eds.), *Linguistic change and contact* (124-136). Austin: University of Texas at Austin.

Hilpert, M. (2008). The comparative: Language structure and language use. *English Language and Linguistics*, 12(3), 395-417.

Kroch, A.S. (1989). Reflexes of grammar in patterns of language change. *Language Variation and Change*, 1(1), 199-244.

Kytö, M. (1996). 'The best and most excellentest way': The rivalling forms of adjective comparison in Late Middle and Early Modern English. In J. Svartvik (Ed.), *Words: Proceedings of an international symposium* (123-144). Stockholm: Kungl. Vitterhets Historie och Antikvitets Akademien.

Kytö, M. & Romaine, S. (1997). Competing forms of adjective comparison in Modern English: What could be *more quicker* and *easier* and *more effective*? In T. Nevalainen & L. Kahlas-Tarkka (Eds.), *To explain the present: Studies in the changing English language in honour of Matti Rissanen* (329-352). Helsinki: Société Néophilologique.

Kytö, M. & Romaine, S. (2000). Adjective comparison and standardisation processes in American and British English from 1620 to the present. In L. Wright (Ed.), *The development of Standard English, 1300-1800: Theories, descriptions, conflicts* (171-194). Cambridge: Cambridge University Press.

Kytö, M. & Romaine, S. (2006). Adjective comparison in nineteenth-century English. In M. Kytö, M. Rydén & E. Smitterberg (Eds.), *Nineteenth-century English: Stability and change* (194-214). Cambridge: Cambridge University Press.

Labov, W. (1966). *The Social Stratification of English in New York City*. Washington: Center for Applied Linguistics.

Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, W. (1994). *Principles of linguistic change*, Vol.1, *Internal factors*. Cambridge and Oxford: Blackwell.

Leech, G. & Culpepper, J. (1997). The comparison of adjectives in recent British English. In T. Nevalainen & L. Kahlas-Tarkka (Eds.), *To explain the present: Studies in the changing English language in honour of Matti Rissanen* (353-373). Helsinki: Société Néophilologique.

Lindquist, H. (1998). The comparison of English disyllabic adjectives in *-y* and *-ly* in Present-day British and American English. In H. Lindquist, S.

Klintborg, M. Levin & M. Estling (Eds.), *The major varieties of English. Papers from MAVEN 97* (205-212). Växjö: Acta Wexionensia.

Lindquist, H. (2000). *Livelier or more lively*? Syntactic and contextual factors influencing the comparison of disyllabic adjectives. In J.M. Kirk (Ed.), *Corpora galore: Analyses and techniques in describing English* (125-132). Amsterdam: Rodopi.

McCarthy, M. (1991). *Discourse analysis for language teachers*. Cambridge: Cambridge University Press.

Mitchell, B. (1985). *Old English syntax*, Vol. 1. Oxford: Clarendon Press.

Mondorf, B. (2003). Support for *more*-support. In G. Rohdenburg & B. Mondorf (Eds.), *Determinants of grammatical variation in English* (252-304). Berlin: Mouton de Gruyter.

Mondorf, B. (2009). *More support for more-support. The role of processing constraints on the choice between synthetic and analytic comparative forms*. Amsterdam: John Benjamins.

Mustanoja, T.F. (1960). *A Middle English syntax*, Part I: *Parts of speech*. Helsinki: Société Néophilologique.

Pound, L. (1901). *The comparison of adjectives in the 15th and 16th centuries*. Heidelberg: Carl Winter.

Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. (1985). *A comprehensive grammar of the English language*. New York: Longman.

Romaine, S. (1982). *Socio-historical linguistics: Its status and methodology*. Cambridge: Cambridge University Press.

Schlüter, J. (2001). Why worser is better: The double comparative in 16th- to 17th-century English. *Language Variation and Change*, 13(2), 193-208.

Scrivner, O. (2010). The white is more firm, and the yolk is rounder: Comparative variation in American English. Paper presented at NWAV 39, University of Texas at San Antonio, 5 November.

Trudgill, P. (2004). *New dialect formation: On the inevitability of colonial Englishes*. Edinburgh: Edinburgh University Press.

Weinreich, U., Labov, W. & Herzog, M.I. (1968). Empirical foundations for a theory of language change. In W.P. Lehmann & Y. Malkiel (Eds.), *Directions for historical linguistics. A symposium* (95-195). Austin: University of Texas Press.