Short
En bref

Regular exposure facilitates dual-mapping of Cantonese pronunciation variants

RACHEL SOO

Department of Linguistics, University of British Columbia, Vancouver, Canada
rachel.soo@ubc.ca

and

MOLLY BABEL

Department of Linguistics, University of British Columbia, Vancouver, Canada
molly.babel@ubc.ca

Abstract

The multilingual landscape of Canada creates opportunities for many heterogeneous bilingual communities to experience systematic phonetic variation within and across languages and dialects, and exposes listeners to different pronunciation variants. This paper examines phonetic variation through the lens of an ongoing sound change in Cantonese involving word-initial [n] and [l] across two primed lexical decision tasks (Experiment 1: Immediate repetition priming task, Experiment 2: Long-distance repetition priming task). Our main question is: How are sound change pronunciation variants recognized and represented in a Cantonese-English bilingual lexicon? The results of both experiments suggest that [n]- and [l]-initial variants facilitate processing in both short and long-term spoken word recognition. Thus, regular exposure to Cantonese endows bilingual listeners with the perceptual flexibility to dually and gradiently map pronunciation variants to a single lexical representation.

Keywords: mergers, Cantonese, multilingualism, word recognition

Résumé

Le paysage multilingue du Canada permet à de nombreuses communautés bilingues hétérogènes de faire l'expérience d'une variation phonétique systématique au sein de et entre les langues et les dialectes, et expose les auditeurs à différentes variantes de prononciation. Cet article examine la variation phonétique sous l'angle d'un changement de son en cours en cantonais impliquant les initiales du mot [n] et [l] dans deux tâches de décision lexicale avec amorçage (expérience 1 : tâche d'amorçage par répétition immédiate, expérience 2 : tâche d'amorçage par répétition à longue distance). Notre principale question est la suivante : comment les variantes de prononciation par changement de son sont-elles reconnues et représentées dans un lexique bilingue cantonais-anglais ? Les résultats des deux expériences suggèrent que les variantes initiales [n] et [l] facilitent le traitement dans la reconnaissance des mots parlés à court et à long terme. Ainsi, l'exposition régulière au cantonais confère aux auditeurs bilingues la flexibilité perceptive nécessaire à la mise en correspondance duale et progressive des variantes de prononciation avec une représentation lexicale unique.

Mots-clés: fusion, cantonais, multilinguisme, reconnaissance des mots

1.   INTRODUCTION

Phonetic variation abounds in spoken words. Variation may be irregular, as in the case of the occasional mispronunciation, or systematic, as in the case of variation that is conditioned by social and linguistic structures (e.g., Labov 1963, Clopper and Turnbull 2018). In the current study, we examine systematic variation and its effects on word recognition for Cantonese-English bilinguals in Canada by targeting a described diachronic change-in-progress involving the onset consonants /n/ and /l/ in Cantonese.

Listeners do not treat all experienced phonetic or phonological variation equivalently (e.g., Johnson 1997, Sumner and Kataoka 2013). For instance, mispronunciations may be recognizable, but they need not be encoded in long-term linguistic representations. At the same time, an adaptive linguistic system may need recourse for encoding more regular phonetic/phonological variation. It is important to separate the *recognition* of an item from its *encoding* in the lexicon. While the nature of linguistic representations and the process of encoding differ across models (see Pierrehumbert 2016 for a review and Todd et al. 2019 for a recent instantiation of the distinction between recognition and encoding), behaviour in different experimental manipulations has traditionally been used to delineate recognition from encoding. For instance, multiple conditioned pronunciation variants may facilitate short-term processing and recognition (McLennan et al. 2003, Sumner and Samuel 2009), as evidenced by immediate priming paradigms that probe listener recognition across a short interval of time (e.g., 500 ms). There is also evidence that typical pronunciations are encoded more robustly in lexical representations than atypical pronunciations (Nygaard et al. 2000) and that canonical pronunciations receive preferential encoding (Sumner and Samuel 2005) in tasks where primed listener recognition occurs even across a more extended interval of time (e.g., across experimental blocks). The notion of preferential encoding points to the fact that the forms that are encoded, or, perhaps, encoded most robustly, may be socially prestigious or

valuable, as opposed to encoding strength being simply related to frequency of occurrence (Sumner and Kataoka 2013, Sumner et al. 2014, Babel et al. 2019, Jones and Clopper 2019). More broadly, the idea that multiple forms may be encoded, but at gradient strengths that activate a lexical representation to varying degrees, is supported by work showing that monolingual and bilingual listeners are sensitive to social-dialectal information in word recognition (Drager 2010, McGowan 2015, Szakay et al. 2016, Jones and Clopper 2019).

Importantly, regular contextual exposure can also lead to more lenient or flexible speech processing strategies (Sumner and Samuel 2009). For example, Samuel and Larraza (2015) posit that Basque listeners are regularly exposed to Spanish-accented Basque, which allows for the dual-mapping of pronunciation variants – in this case, laminar [t͡s̺] and palato-alveolar [t͡ʃ] – to lexical items for the benefit of efficient word recognition. In this respect, regular exposure to Spanish-accented Basque can blur this affricate place-of-articulation contrast for word recognition, as both pronunciation variants map to a single lexical item. This dual-mapping can also be realized in within-dialect changes-in-progress. In New Zealand English, younger speakers are merging the EAR/AIR vowels, producing AIR words in a similar phonetic space to EAR words. However, given their regular exposure to older speakers, who maintain a more clear contrast in these lexical sets, listeners show a kind of dual-mapping where recognition of, for example, 'chair' is achieved by both pronunciation variants [t͡ʃɪə] and [t͡ʃeə] (Rae, Warren, et al. 2002).

1.1 Current study: Cantonese /n/ and /l/ variation

These findings for New Zealand English demonstrate that pronunciation variation can be rooted in diachronic change, where, from a synchronic perspective, a demographic group may be more or less advanced in a sound change-in-progress. This may be the case with a sound change in Cantonese, where word-initial /n/ is often produced as [l] (Wong 1941, Matthews and Yip 2013, To et al. 2015), a pattern that was initially observed in the first half of the twentieth century (Ball 1907). In this sound change, words such as *nou5*[1] 腦 'brain', historically pronounced with an initial /n/ (hereafter, historical variant) are pronounced with [l] (hereafter, innovative variant), making them homophonous with pre-existing /l/-initial words, such as *lou5* 老 'old'. Participation in this sound change may be higher in younger speakers (Yeung 1980, Bourgerie 1990, To et al. 2015, Zee 1999) though more recent work finds similarly high rates of [l] for /n/ between younger and older speakers in Hong Kong (Cheng et al. 2022). Additionally, early work by Pan (1981) provides some evidence that the sound change may be conditioned by stylistic context. In their study, young and old Cantonese speakers heard two passages (one bearing [n]-initial forms and another bearing [l]-initial forms) produced by the same speaker. While listeners were asked to pay attention to the similarity between speakers, they were ultimately also asked to decide which speaker had a higher socio-economic status. Overall, the speakers indicated that the speaker who produced the

---

[1]Cantonese items are romanized in Jyutping (Tang et al. 2002), which transparently indicates whether the intended forms are [n]- or [l]-initial.

passage with [n]-initial forms had a higher socioeconomic status, suggesting that historical forms denote prestige. The direction of this sound change combined with the social context may ultimately position [l]-initial pronunciations to be more frequently encountered in everyday, more casual speech processing and [n]-initial pronunciations to be more frequently encountered in more formal contexts.

While the historical progression of the sound change has been investigated from a sociolinguistic perspective, there are few phonetic studies that examine the consequences for perception or recognition. Cheng (2017) and later Cheng et al. (2022) represent the first attempts at carefully examining Cantonese [n] and [l] in perception. In Cheng (2017), older and younger groups of early Cantonese-English bilinguals recruited from Hong Kong and Vancouver took part in a two-alternative forced choice word identification task where listeners heard tokens from a synthesized [n]-to-[l] continuum. The results showed that younger speakers were more categorical in their perception of [n] and [l] than older speakers, suggesting a relative unmerging of [n] and [l]. Cheng (2017) notes that this unmerging is not surprising given the bilingual participants' contact with languages like English and Mandarin (where these consonants remain contrastive), as well as pressure from metalinguistic awareness about the 'correct' (i.e., historical) pronunciation variants.

More broadly, while Cantonese is spoken as the majority language in Hong Kong, Macau, and Guangzhou (Bauer and Benedict 1997), there are large diaspora communities in different parts of the world, including metropolitan cities across Canada. While Canada is traditionally recognized for its rich history of French-English bilingualism, the multilingual landscape of Canada has created relatively stable bilingual speech communities with other languages as well, such as Cantonese. By adulthood, early Cantonese-English bilinguals raised in Canada are typically more dominant in English or French, the dominant societal languages, but Cantonese speech communities across Canada are robust and part of the multicultural fabric of many Canadian communities (Yee 2006). In various population centres, Cantonese is a predominant language spoken and heard by talkers (e.g., Richmond, British Columbia, see Statistics Canada 2017).

The cultural mosaic of Canada is a rich resource for engaging in research questions about multilingualism in ways that acknowledge the long-standing history of languages, like Cantonese, in Canada (see also Wong and Babel 2017). We emphasize that English and Cantonese are both immigrant languages in British Columbia and that this linguistic landscape is an opportunity to consider the dynamics of Cantonese-English bilingualism in Canada (see discussions in Muysken 2020, Polinsky and Scontras 2020). As such, we consider this a proof-of-concept study that these psycholinguistic questions can be explored in a heterogeneous bilingual community in Canada.

As mentioned, dialectal variation provides evidence suggesting that listeners encode multiple phonetic forms in the lexicon (Sumner and Samuel 2009, Samuel and Larraza 2015), since experience with pronunciation variants may support their representation in the lexicon. Thus, the pronunciation variation experienced by Cantonese-English bilingual listeners may facilitate the recognition and representation of sound change variants.

In this study, we examine how these sound change variants are recognized and encoded in early Cantonese-English bilinguals across two experiments. Experiment 1 was an immediate repetition priming task designed to examine how [n] and [l] pronunciations affect short-term word recognition. In critical trials, [l]-initial target words (e.g., *lou5*) were preceded by primes that were identical (i.e., *lou5*) but physically distinct tokens, or by the historical [n]-initial pronunciation variant (e.g., *nou5*). Immediate priming tasks probe short-term word recognition as primes and targets are presented in a single trial and separated by an inter-stimulus interval (ISI). We use a 500 ms ISI that taps into "phonetic perception", following terms devised by Werker and Logan ([1985](#)). If the sound change in question plays a role in word recognition, we predict that both [n]- and [l]-initial pronunciation variants should facilitate the recognition of [l]-initial targets equivalently or at least that both would provide stronger priming than other phonologically similar forms. Thus, primes with a shared rime but with a different onset consonant not undergoing a sound change were also included as control trials. We extend the findings of Experiment 1 through Experiment 2, a long-distance repetition priming task designed to examine the long-term encoding of these sound change variants in lexical representations. In this task, primes and targets were separated into distinct blocks. Long-distance priming tasks tap into longer-term representations as primes and targets are separated across blocks (cf. immediate priming tasks, where primes and targets are presented within a trial). If the sound change has affected the longer-term representation of these items, we hypothesized that [n]- and [l]-initial primes would again prime [l]-initial targets equivalently. On the other hand, if lexical items are unaffected by the sound change in the longer-term, and underlying /n/-initial forms (like *nou5* 腦 'brain') are stored with an initial [n], we hypothesized that only [l]-initial primes would facilitate recognition of [l]-initial targets.

2. EXPERIMENTAL PROCEDURES

The general procedure for Experiments 1 and 2 was the same, with different participants for each task. First, participants completed the Bilingual Language Profile (BLP) questionnaire (Gertken et al. [2014](#)) through a Qualtrics ([2022](#)) survey (Qualtrics, Provo, UT). The BLP questionnaire computes a quantified measure of dominance between the two languages of a bilingual (in this case, Cantonese and English) through a series of questions about language history (e.g., "At what age did you start learning Cantonese/English?"), use (e.g., "In an average week, what percentage of the time do you use Cantonese/English with family?"), proficiency (e.g., "On a scale of 0–6, how well do you read Cantonese/English?"), and attitudes (e.g., "To what degree do you feel like yourself when speaking each language?"). Positive "dominance scores" indicate greater English dominance and negative dominance scores indicate greater Cantonese dominance (the valence is arbitrary on this scale). The dominance scores for the participants in Experiment 1 and 2 are provided in Figures 1 and 3 respectively. The dominance scores span a wide range; this heterogeneity is typical of the Cantonese-English speech community in our subject population.
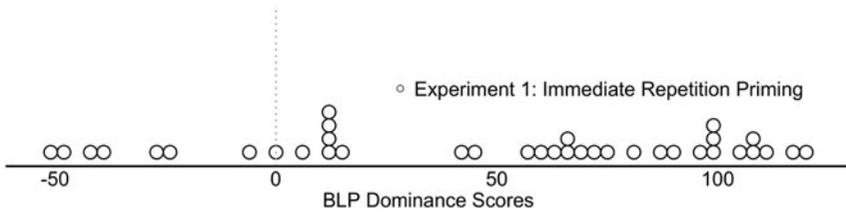
Figure 1: Distribution of dominance scores for participants in Experiment 1: Immediate Repetition Priming. Each dot represents one participant. The number of dots do not add up to the number of participants across experiments, as not all participants filled out the Bilingual Language Profile in full. Positive scores indicate greater English dominance and negative dominance scores indicate greater Cantonese dominance (the valence is arbitrary on this scale, which ranges from ±218).

The rest of the experiment was carried out online through jsPsych (de Leeuw and Motz 2016). After providing informed consent, listeners took part in an auditory headphone check (Woods et al. 2017). Next, participants listened to a Cantonese version of "The North Wind and The Sun", recorded by the same talker who produced the stimuli, to become familiarized with the talker's voice and prepare for a more Cantonese-centric language mode. Finally, participants proceeded to the main task: the immediate repetition priming task in Experiment 1 (see section 3), or the long-distance repetition priming task in Experiment 2 (see section 4). Each experiment took approximately 30–40 minutes to complete.

3.   EXPERIMENT 1: IMMEDIATE REPETITION PRIMING

Experiment 1 was an immediate repetition priming task designed to examine whether historically /n/-initial words pronounced as [n], the conservative and less common variant, or [l], the innovative and more common pronunciation variant, prime equivalently well. The design of Experiment 1 follows the methods described in a similar study by Sumner and Samuel (2009), who examined the representation of dialect variants in the lexicon. In each trial, participants were presented with an auditory prime, followed by the corresponding auditory target after a 500 ms ISI. Participants were instructed to indicate whether the second item (i.e., the target) was a real word of Cantonese or a nonword by pressing "1" or "0" on their keyboard (key order counterbalanced across listeners). The next trial proceeded after an inter-trial interval (ITI) of 1000 ms.

3.1   Stimuli and participants

Following Sumner and Samuel (2009), 40 [l]-initial target words were selected. As minimal pairs for /n/ and /l/ are already limited in Cantonese, variables like lexical category were not considered during stimulus selection. More broadly, a representative bilingual Cantonese-English corpus from which frequency counts could be extracted was not available at the time of the study and corpus statistics

Experiment 1 pair types	Pair	Prime	Target
Critical	Identity	*lou5* 老 'old'	*lou5* 老 'old'
	Historical	*nou5* 腦 'brain'	
Control	Rime	*pou5* 抱 'embrace'	
	Unrelated	*caa4* 茶 'tea'	
Filler	Identity	*daan2* 蛋 'eggs'	*daan2* 蛋 'eggs'
	Rime	*caan2* 剷 'shovel'	
	Unrelated	*fei1* 飛 'to fly'	
	Nonword	*gwem3	
Nonword	Identity	*su1	*su1
	Rime	*zu1	
	Unrelated	*gwei2	
	Real word	*coeng3* 唱 'sing'	

Table 1: Example stimuli sets in Experiment 1. See the Supplementary Materials for the full list of stimuli in each experiment.

derived from monolingual speaker sources are often unreliable measures for populations that diverge from traditional monolingual speaker norms.[2] Each [l]-initial target word (e.g., *lou5* 老 'old') was preceded by one of four different prime types, each representing different pair types (see Table 1). In identity pairs, the prime and target were identical words (though physically distinct tokens from the same speaker; e.g., a different token of *lou5* 老 'old'). In historical pairs, the prime was the historical and less frequent [n]-initial variant of the target (e.g., *nou5* 腦 'brain'). These critical pairs were matched with two control pairs. In rime pairs, the prime and target shared rimes and differed in their initial consonants (e.g., *pou5* 抱 'embrace'). Crucially, unlike the historical pairs the prime did not bear an initial consonant that is reported to be merging with [l]. In unrelated pairs, the prime and the target did not share any features (e.g., *caa4* 茶 'tea').

An additional 80 non-[l]-initial filler target words were selected (e.g., *daan2* 蛋 'egg'). As with the real-word targets, the real-word fillers were preceded by different primes, each representing different pair types: identity (e.g., a different token of *daan2* 蛋 'egg'), rime (e.g., *caan2* 鏟 'shovel'), unrelated (e.g., *fei1* 飛 'to fly'), and nonword primes (e.g., *gwem3). This produced a 1:2 ratio of critical to filler items in the experiment, and a total of 120 real-word targets.

To balance the number of real-word and nonword targets in the experiment, 120 nonword targets were created from systematic and accidental gaps in the language (Kirby and Yu 2007). Here again, the nonword targets (e.g., *su1) were preceded by different primes, each representing different pair types: identity (e.g., a different token of *su1), rime (e.g., *zu1), unrelated (e.g., *gwei2), and real-word primes

---

[2]To at least partially account for the potential effect of item frequency, the Bayesian mixed effects models for Experiment 1 and 2 both included by-item random intercepts.

Pair type	Mean reaction time (ms)
Identity	1063.23 (412.53)
Historical	1106.52 (410.21)
Rime	1182.27 (419.34)
Unrelated	1256.43 (409.09)

Table 2: Mean reaction times in milliseconds (ms) to critical (identity and historical) and control (rime and unrelated) pair types in Experiment 1. Standard deviations are provided in parentheses.

(e.g., *coeng3* 唱 'sing'). Forty of these 120 nonword targets were [l]-initial and bore [n]-initial nonword rime primes. These were included so that [n] and [l] were not exclusively represented in real-words.

Each of these prime-target pairs was counterbalanced across 4 lists. Thus, each list contained a total of 240 trials, in which there were 120 real-word targets (40 critical pairs and 80 filler pairs) and 120 nonword targets. All stimuli were produced by a linguistically-trained female early Cantonese-English bilingual (27 years old) who produced clear [n]- and [l]-initial items without the sound change. Items were recorded in a sound-attenuated cabin in Audacity through a Samson C03U USB microphone at a sampling rate of 44.1 kHz and 24-bit depth and pre-tested to ensure that the initial [n] and [l] segments were clearly produced.[3]

Sixty-three early Cantonese-English bilinguals who acquired Cantonese before the age of five completed the task. As in Sumner and Samuel (2009), participants with high error rates on filler trials were removed. Twenty-three participants with error rates greater than 15% on filler trials were removed (4.6% of the data), leaving the data from 40 early Cantonese-English bilinguals (mean age = 21 years) for further analysis. These 40 early Cantonese-English bilinguals all spent the majority of their lives in a Cantonese-speaking family (mean = 19 years). Additionally, on a scale of 0 to 6 (where 0 represents "not well at all" and 6 represents "very well") they reported high understanding of Cantonese (median = 5) and English (median = 6) and high speaking ability in Cantonese (median = 4) and English (median = 6).

3.2 Analysis and results

Following Balota et al. (2007), reaction times greater than 3000 ms or less than 200 ms were removed, as well as reaction times less than or greater than 2.5 standard deviations from each subject's mean (approximately 4.8% of the responses). The reaction times for correct responses of the data post-outlier removal are summarized in Table 2.

---

[3]The critical [n]- and [l]- initial auditory tokens were blinded and submitted to three research assistants who did not speak Cantonese or any other language where /n/ and /l/ are merging. The initial consonants of all real-word items were labeled as intended by all research assistants.

Experiment 1	$\hat{\beta}$ Estimate	95% CrI	PD	$\hat{R}$	Bulk ESS
Intercept (Historical)	6.95	[6.88, 7.02]	100%	1.00	2733
Pair (Identity vs. Historical)	−0.05	[−0.10, 0.01]	95.26%	1.00	12706
Pair (Rime vs. Historical)	0.07	[0.02, 0.11]	99.94%	1.00	18671
Pair (Unrelated vs. Historical)	0.14	[0.09, 0.18]	100%	1.00	17875

Table 3: Population-level or fixed-effect predictors summary of the Bayesian mixed-effects model carried out on log reaction times from Experiment 1. The $\hat{\beta}$ estimate, lower and upper bounds of the 95% Credible Interval (CrI), and Probability of Direction (PD) are reported. The $\hat{R}$ and bulk effective sample size (ESS) are also provided.

The reaction time data were analyzed using a Bayesian mixed-effects model fitted in Stan using the {brms} package (Bürkner 2017) in R (R Core Team 2021). Bayesian models shed light on the continuous probability distributions of parameter values, and, crucially, afford meaningful interpretation of a null result. More broadly, a Bayesian framework was most appropriate for the current set of experiments as the study aimed to see how well data would fit with theoretical predictions and to incorporate nuance to interpretations that are, in all likelihood, not black or white. In Bayesian models, one way to quantify and evaluate the strength of an effect is through a credible interval (CrI) and the probability of direction (PD). However, just as the 0.05 alpha level in Frequentist frameworks is an arbitrary selection, the use of a selected credible interval percentage is also arbitrary (Kruschke 2014, McElreath 2020). We follow the standards in Nicenboim and Vasishth (2016), wherein strong evidence for a non-null effect is represented by a 95% CrI that does not include 0, and weak evidence for an effect is represented by a 95% CrI that includes 0, with probabilities of at least 95%.

The model included Pair (Historical, Identity, Rime, Unrelated) as a dummy-coded fixed effect with Historical pairs as the reference level. Crucially, this allows us to compare whether Historical pronunciations prime equivalently well as Identity and Rime primes. The model also included by-item random intercepts and by-subject random slopes for the effect of Pair. The inclusion of subject in the random effects structure is also fundamental, given the heterogeneity that is expected within bilingual populations. We assumed a lognormal response distribution for reaction times with a mean of 6 and standard deviation of 1 for the intercept, following Ciaccio and Veríssimo (2020), who also analyzed lexical decision reaction times in a Bayesian framework. Regularizing, weakly informative priors were selected for the all population-level effects, and for the random intercepts and slopes, centered at 0 with a standard deviation of 0.25 (see the Supplementary Materials for a full list of prior distributions). Following Casillas (2020), the correlation parameter was an LKJ correlation prior with regularization set at 2. We used Hamiltonian Monte-Carlo sampling with four chains (each with 2000 iterations, 1000 warm-up) to draw samples from the posterior distribution. Table 3 provides the output of the model. The mean of the posterior distribution and the 95%
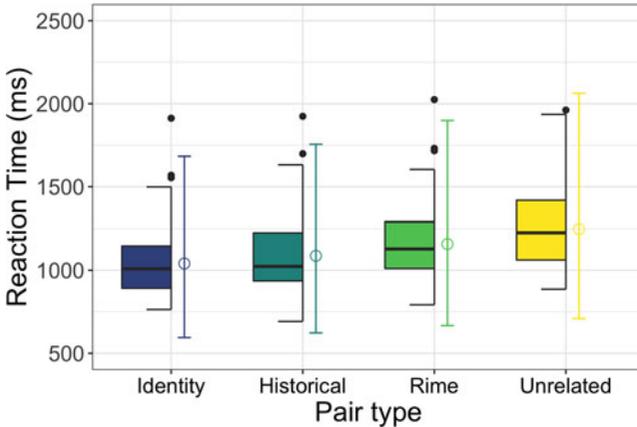
Figure 2: Half box-and-whisker plots of the raw reaction time (ms) for critical (identity and historical) and control (rime and unrelated) pair types in Experiment 1. The 95% range of the posterior predictive distributions for each population level coefficient is indicated by the coloured error bars.

credible interval (CrI) are given for each estimate. We also report the probability of direction (PD) for the levels in Table 3, which provides a measure of how probable the specific direction of the effect is, in addition to the bulk effective sample size (ESS) and $\hat{R}$ values to show that the model was well mixed.

Participants tended to respond faster to identity pairs compared to historical pairs (the reference level). The upper limit of the 95% CrI for this comparison encompasses 0 and the PD is approximately 95% (Identity pairs: $\beta = -0.05$, CrI = [−0.10, 0.01], PD = 95.26%), making this relatively weak evidence for identity pairs having a processing advantage over historical pairs. While these current data might, therefore, suggest a lack of a *strong* difference between historical and identity pairs, we note that the CrI is almost entirely negative. This is visualized in the posterior distributions for each population level coefficient in Figure 2. Thus, while the CrI encompasses 0, given the data the distribution of values do suggest that identity primes may be slightly more effective than historical primes.

In contrast, participants were significantly slower to respond to both rime and unrelated pairs compared to historical pairs. This is indicated by the positive beta estimates, the 95% credible intervals that do not encompass 0, and the strong probabilities of positive directions (Rime pairs: $\beta = 0.07$, CrI = [0.02, 0.11], PD = 99.94%, Unrelated pairs: $\beta = 0.14$, CrI = [0.09, 0.18], PD = 100%).

3.3   Interim discussion

Overall, these results suggest that listeners map the historical [n]-initial pronunciations and the innovative [l]-initial pronunciations to historically /n/-initial words. The lack of strong response time difference between historical and identity pairs suggests that the sound change *may* have rendered [n]-initial primes (in historical pairs)
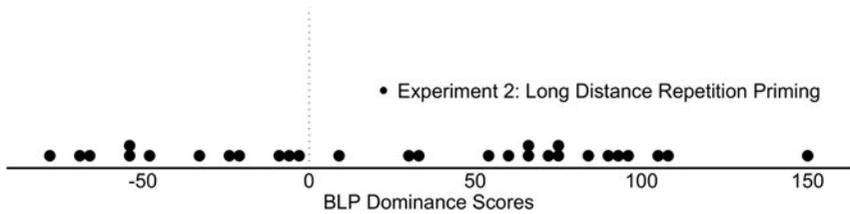
Figure 3: Distribution of dominance scores for participants in Experiment 2: Long-Distance Repetition Priming. Each dot represents one participant. The number of dots do not add up to the number of participants across experiments, as not all participants filled out the Bilingual Language Profile in full. Positive scores indicate greater English dominance and negative dominance scores indicate greater Cantonese dominance (the valence is arbitrary on this scale, which ranges from ±218).

and [l]-initial primes (in identity pairs) as recognition equivalents, though the data may ultimately suggest that identity pairs indeed have a processing advantage. Certainly, the priming observed for historical pairs is unlikely to be a product of shared rimes between the prime and target, as there was strong evidence for a difference between the historical and rime trials. If the priming effect of the historical primes was due to the shared rime, we would not expect to see such a difference between the historical and rime pairs.

Taken together, these data provide evidence that both historical [n]-initial primes and innovative [l]-initial primes facilitate recognition of /l/-initial targets compared to control trial types, and that the innovative identity prime may have a processing advantage.

These data may be accounted for by (at least) two explanations. First, listeners may simply be poor at distinguishing [n] from [l] at a phonetic level. That is, the difference between the initial [n] and [l] in the historical and identity primes, respectively, may simply not have been robustly perceptible. Work by Cheng (2017), however, suggests that this is unlikely to be the case, as the same population of early Cantonese-English bilinguals recruited from Vancouver in her study showed sensitivity to [n] and [l] pronunciation variants. In fact, as mentioned in section 1, the participants in Cheng (2017) demonstrated an apparent perceptual *unmerging* of [n] and [l], as for younger participants the category boundary between the two sounds was rather crisp in comparison to older bilinguals.

A second explanation is found at the lexical level. Listeners are able to perceive the difference between [n] and [l] and map the two phonetic codes to a single lexical representation (e.g., *lou5* and *nou5* both map to *lou5* 老 'old'). This dual-mapping explanation was proposed in Samuel and Larraza (2015) to account for the finding that very early, highly fluent Basque-Spanish bilinguals accept "erroneous" pronunciations of Basque /tʃ/ as [tʂ] at high rates in a lexical decision task. The Basque-Spanish bilinguals showed clear discrimination of these sounds at a phonetic level (in an AXB task). In a subsequent picture matching task, however, where listeners had to tap into their lexical knowledge to map the auditory utterance to the correct image, listeners showed tolerance for erroneous pronunciations (at even higher

rates than in the lexical decision task). The authors conclude that the acceptance of these mispronunciations reflects the mapping of two phonetic variants (a "correct" and "incorrect" pronunciation, following Samuel and Larraza's terms) to a single lexical representation. In the case of the Basque listeners, this variability in phonetic forms stems from regularly hearing Spanish-accented Basque, which often does not maintain the historical Basque pronunciation.

We further test the dual-mapping hypothesis in Experiment 2 with a long-distance repetition priming task where the prime and target are separated across experimental blocks. In this paradigm, minimal pairs are not expected to show priming across long-distances. As such, if [n]- and [l]-initial forms are indeed separately encoded variants, priming across a more extended time interval in a long-distance repetition priming task is not expected. On the other hand, if both pronunciations are fully encoded as acceptable variants of a single lexical item, long-distance priming is anticipated.

4. EXPERIMENT 2: LONG-DISTANCE REPETITION PRIMING

Experiment 2 was a long-distance repetition priming task modeled after the long-distance priming task described in Sumner and Samuel (2009). Unlike Experiment 1, where participants were presented with prime-target pairs separated by a 500 ms ISI on each trial, listeners were instead presented with individual auditory tokens on each trial and instructed to indicate whether the item was a real word of Cantonese or a nonword by pressing "1" or "0" on their keyboard (key order counterbalanced across listeners). The next trial proceeded after an ITI of 1000 ms. There were two blocks in Experiment 2 separated by a short, self-paced break and trials were randomized within each block.

4.1 Stimuli and participants

Experiment 2 utilized a subset of the stimuli from Experiment 1 (see Table 4). Specifically, the first block contained the primes from 10 critical identity pairs (e.g., *lou5* 老 'old'), the primes from 10 critical historical pairs (e.g., *naam4* 男

Experiment 2 pair types	Pair	Block 1	Block 2
Critical	Identity	*lou5* 老 'old'	*lou5* 老 'old'
	Historical	*naam4* 男 'male'	*laam4* 藍 'blue'
Control	Unmatched	–	*laai1* 拉 'pull'
Filler	Repeated	*tin1* 天 'sky'	*tin1* 天 'sky'
	Unrepeated	*haang4* 行 'walk'	*so2* 'lock'
Nonword	Repeated	**wen2*	**wen2*
	Unrepeated	**son3*	**tu2*

Table 4: Example stimuli sets in Experiment 2. See the Supplementary Materials for the full list of stimuli in each experiment.

'male'), 160 real-word filler items (e.g., *tin* 天 'sky'), and 240 nonword items (e.g., *wen2*). This produced a total of 420 trials in the first block, consisting of 180 words and 240 nonwords. The second block contained the 10 corresponding targets of the critical identity pairs (e.g., a different token of *lou5* 老 'old') and the 10 corresponding targets of the critical historical pairs (e.g., *laam4* 藍 'blue'). Twenty additional [l]-initial words that bore no corresponding prime in the preceding block were also included and represented unmatched controls (e.g., *laai1* 拉 'pull'). One hundred and sixty real-word filler items were again presented; 80 of these were words repeated from the preceding block (e.g., *tin1* 天 'sky'), while the other 80 were new filler items (e.g., *so2* 鎖 'lock'). Two hundred and forty nonwords were also included in the second block, with 120 words repeated from the first block (e.g., *wen2*), and 120 new nonwords (e.g., *tu2*). This produced a total of 420 trials in the second block, consisting of 200 words and 240 nonwords.

Fifty-two early Cantonese-English bilinguals who acquired Cantonese before the age of five completed the task. As in Sumner and Samuel (2009), participants with high error rates on filler trials were removed. Twenty-two participants with error rates greater than 15% on filler trials were removed (4.4% of the data), leaving the data from 30 early Cantonese-English (mean age = 21 years) for further analysis. These 30 early Cantonese-English bilinguals all spent the majority of their lives in a Cantonese-speaking family (mean = 19 years). Additionally, on a scale of 0 to 6 (where 0 represents "not well at all" and 6 represents "very well") they reported high understanding of Cantonese (median = 6) and English (median = 6) and high speaking ability in Cantonese (median = 5.5) and English (median = 6).

4.2 Analysis and results

In Experiment 2, we analyzed the reaction times for the critical pair types from Block 2 only. Reaction times greater than 3000 ms or less than 200 ms were removed, as well as reaction times less than or greater than 2.5 standard deviations from each subject's mean (approximately 4.4% of the responses). These remaining data are summarized in Table 5. We analyzed the reaction times from Block 2 using the same Bayesian model structure and priors as in Experiment 1 (see section 3.2). For the sake of brevity, we do not go over the details here again. Table 6 provides the output of the model, where the mean of the posterior distribution and the 95% CrI

Pair type	Mean reaction time (ms)
Identity	1104.26 (302.56)
Historical	1050.75 (302.06)
Unpaired	1158.58 (321.97)

Table 5: Mean reaction times (ms) to to critical (identity and historical) and control (unmatched) pair types in Block 2 of Experiment 2. Standard deviations are provided in parentheses.

Experiment 2	$\hat{\beta}$ Estimate	95% CrI	PD	$\hat{R}$	Bulk ESS
Intercept (Historical)	6.94	[6.86, 7.01]	100%	1.00	4982
Pair (Identity vs. Historical)	0.04	[−0.04, 0.13]	85.16%	1.00	6472
Pair (Unpaired vs. Historical)	0.08	[0.01, 0.16]	98.30%	1.00	6599

Table 6: Population-level or fixed-effect predictors summary of the Bayesian mixed-effects model carried out on log reaction times from Experiment 2. The $\hat{\beta}$ estimate, lower and upper bounds of the 95% Credible Interval (CrI), and Probability of Direction (PD) are reported. The $\hat{R}$ and bulk effective sample size (ESS) are also provided.
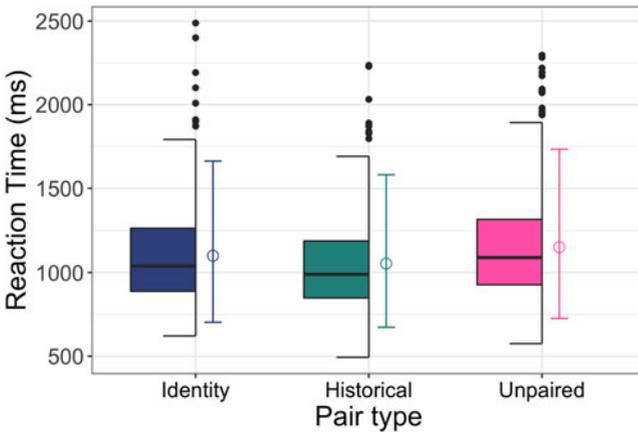

Figure 4: Half box-and-whisker plots of the raw reaction time (ms) for critical (identity and historical) and control (unpaired) pair types in Block 2 of Experiment 2. The 95% range of the posterior predictive distributions for each population level coefficient is indicated by the coloured error bars.

are given for each estimate as well as the PD, ESS, and $\hat{R}$ values. The empirical data and the posterior distributions for each pair type are visualized in Figure 4.

Participants were slower to respond to identity pairs compared to historical pairs (the reference level), as indicated by the positive beta estimate. However, the 95% CrI encompasses 0 and the probability of direction is approximately 85%, which, together, suggests there is little to no evidence that this difference is meaningful (Identity pairs: $\beta = 0.04$, CrI = [−0.04, 0.13], PD = 85.16%). Participants were also slower to respond to unpaired items compared to historical pairs, and the evidence for this difference was strong (Unpaired pairs: $\beta = 0.08$, CrI = [0.01, 0.16], PD = 98.30%). These results are generally in line with those of Experiment 1, and indicate that historical [n]-initial primes and innovative [l]-initial primes are at an advantage

over unpaired or control items at both short and long time lags when processing [l]-initial targets.

5. GENERAL DISCUSSION

In Experiment 1, [l]-initial targets were primed by both [n]- and [l]-initial primes in an immediate repetition priming paradigm. There was strong evidence that [n]- and [l]-initial targets prime better than control pairs, and weak evidence that the innovative identity prime [l]- had a priming advantage over the historical [n]- pronunciation for immediate word recognition. These data suggest that both variants of the sound change may be dually-mapped to the same lexical representation in short-term processing, but may potentially vary in activation strength. To investigate the dual-mapping hypothesis with these pronunciation variants more fully , we carried out a long-distance repetition priming task in Experiment 2 to assess lexical encoding. Similarly to Experiment 1, the data from this task revealed no major difference between the historical and identity pairs, with more convincing evidence that the [n]-initial and [l]-initial forms primed the [l]-initial targets equivalently well. Our Bayesian analysis allows for subtlety in interpretation: numerically speaking, the innovative [l]-initial variants prime to a slightly greater degree than [n]-initial variants in short-term processing, suggesting that the more frequently encountered innovative form may afford a processing advantage. However, overall, these findings suggest that pronunciation variants may be dually-mapped in immediate word recognition as well as in longer-term lexical encoding. These results complement the findings in Sumner and Samuel (2005) and Jones and Clopper (2019), which support gradient encoding in the lexicon. Additionally, these results suggest that in Cantonese, particular phonetic forms may be encoded to varying strengths, with, in this case, the more frequently encountered innovative form receiving some degree of preferential status in immediate recognition, though not in long-term encoding.

How do listeners develop the perceptual flexibility to include a separate pronunciation variant for a given lexical item? Similarly to the Basque-Spanish bilinguals in Samuel and Larraza (2015), the Cantonese-English bilingual listener population in Canada is exposed to variation in pronunciation. These populations reside in communities where they are regularly exposed to alternative phonetic variants – in some cases participating in this variation themselves – making it potentially beneficial for listeners to expand the set of acceptable pronunciations for a particular lexical item. Regular exposure to sound change variants in the community may simultaneously allow listeners to map two phonetic variants to a single lexical representation, and maintain a distinction between [n] and [l] at the phonetic level, regardless of the source of variation (Sumner and Samuel 2009, Samuel and Larraza 2015).

While the [n] and [l] merger has been traditionally described as a change-in-progress, it may also be the case that the variation has become enregistered (Agha 2005). As mentioned in section 1.1, work by Pan (1981) suggests that the historical form indexes prestige. In this regard, the variation between [n] and [l] may represent more stable stylistic variation. In such a case, most speakers would exhibit some amount of structured variation in their usage, adopting more [n]-like pronunciations

in more formal contexts and more [l]-like pronunciations in more casual contexts. This would mean that a dual-mapping of [n] and [l] pronunciation variants would nearly be required for effective speech processing, as all speakers would exhibit some amount of pronunciation variation based on social context. It may be that the [n] and [l] merger, which has been described as complete in previous work (To et al. 2015), may indeed be stabilizing as context-conditioned variation. This is also considered in the discussion in Cheng et al. (2022).

6.   CONCLUSION

An ongoing sound change in Cantonese sees word-initial /n/ produced as [l], rendering [n]-initial words and [l]-initial words homophononous. In the current study, we examined the effects of this sound change on word recognition in early Cantonese-English bilinguals across two experiments. Experiment 1 was an immediate repetition priming task designed to examine how [n] versus [l] pronunciations affect immediate word recognition while Experiment 2 was a long-distance repetition priming task designed to examine how these pronunciation variants are encoded in the longer term. The results of both experiments indicate that [n]- and [l]-initial pronunciations facilitate processing. The evidence in support of recognition equivalence (Sumner et al. 2014) of these pronunciation variants in long-term lexical encoding was clear, though there was some indication that the identity primes ([l]-initial pronunciations) may have had a slight advantage over the historical [n]-initial pronunciations in immediate word recognition. Collectively, these results are generally consistent with a dual-mapping explanation whereby regular exposure to pronunciation variants endows Cantonese-English bilinguals with the perceptual flexibility to map multiple phonetic forms to a single lexical representation.

SUPPLEMENTARY MATERIALS

To view supplementary material for this article, please visit https://doi.org/10.1017/cnj.2024.3.

REFERENCES

Agha, Asif. 2005. Voice, footing, enregisterment. *Journal of Linguistic Anthropology* 15(1): 38–59.

Babel, Molly, Brianne Senior, and Sophie Bishop. 2019. Do social preferences matter in lexical retuning? *Laboratory Phonology* 10(1): 1–22. DOI 10.5334/labphon.133.

Ball, James Dyer. 1907. *Cantonese made easy: A book of simple sentences in the Cantonese dialect, with free and literal translations, and directions for the rendering of English grammatical forms in Chinese.* Kelly & Walsh.

Balota, David A., Melvin J. Yap, Keith A. Hutchison, Michael J. Cortese, Brett Kessler, Bjorn Loftis, James H. Neely, Douglas L. Nelson, Greg B. Simpson, and Rebecca Treiman. 2007. The english lexicon project. *Behavior Research Methods* 39(3): 445–459.

Bauer, Robert S, and Paul K Benedict. 1997. *Modern Cantonese phonology.* Berlin: Mouton de Gruyter.

Bourgerie, D. S. 1990. A quantitative study of sociolinguistic variation in Cantonese. Doctoral Dissertation, Ohio State University.

Bürkner, Paul-Christian. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of statistical software* 80(1): 1–28.

Casillas, Joseph V.. 2020. The longitudinal development of fine-phonetic detail: Stop production in a domestic immersion program. *Language Learning* 70(3): 768–806.

Cheng, Lauretta. 2017. Mergers-in-Progress in Hong Kong and Vancouver Cantonese-English Bilinguals. Bachelor's Thesis, University of British Columbia.

Cheng, Lauretta S. P., Molly Babel, and Yao Yao. 2022. Production and perception across three Hong Kong Cantonese consonant mergers: Community-and individual-level perspectives. *Laboratory Phonology* 13(1). DOI 10.16995/labphon.6461.

Ciaccio, Laura Anna, and João Veríssimo. 2020. Investigating variability in morphological processing with bayesian distributional models. *Psychonomic Bulletin & Review*: 1–11.

Clopper, Cynthia G., and Rory Turnbull. 2018. Exploring variation in phonetic reduction: Linguistic, social, and cognitive factors. In *Rethinking reduction*, eds. Francesco Cangemi, Meghan Clayards, Oliver Niebuhr, Barbara Schuppler, and Margaret Zellers, 25–72. De Gruyter Mouton. DOI 10.1515/9783110524178-002.

Drager, Katie. 2010. Sociophonetic variation in speech perception. *Language and Linguistics Compass* 4(7): 473–480.

Gertken, Libby M., Mark Amengual, and David Birdsong. 2014. Assessing language dominance with the bilingual language profile. In *Measuring l2 proficiency: Perspectives from SLA*, eds. Pascale Leclercq, Amanda Edmonds, and Heather Hilton, 208–225. Bristol: Multilingual Matters Bristol.

Johnson, Keith. 1997. *Speech perception without speaker normalization: An exemplar model*. In *Talker variability in speech processing*, eds Keith Johnson and John W. Mullennix, 145–165. San Diego: Academic Press.

Jones, Zack, and Cynthia G. Clopper. 2019. Subphonemic variation and lexical processing: Social and stylistic factors. *Phonetica* 76(2–3): 163–178.

Kirby, James P., and Alan C. L. Yu. 2007. Lexical and phonotactic effects on wordlikeness judgments in Cantonese. In *Proceedings of the International congress of the Phonetic Sciences XVI*.

Kruschke, John. 2014. Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan.

Labov, William. 1963. The social motivation of a sound change. *Word* 19(3): 273–309.

de Leeuw, Joshua R., and Benjamin A. Motz. 2016. Psychophysics in a web browser? Comparing response times collected with JavaScript and Psychophysics Toolbox in a visual search task. *Behavior Research Methods* 48(1): 1–12.

Matthews, Stephen, and Virginia Yip. 2013. *Cantonese: A comprehensive grammar*. 2nd edn. Routledge.

McElreath, Richard. 2020. *Statistical rethinking: A Bayesian course with examples in R and Stan*. 2nd edn. Chapman & Hall/CRC Press.

McGowan, Kevin B.. 2015. Social expectation improves speech perception in noise. *Language and Speech* 58(4): 502–521.

McLennan, Conor T., Paul A. Luce, and Jan Charles-Luce. 2003. Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29(4): 539–553.

Muysken, Pieter. 2020. The case for contact induced-change in heritage languages. *Bilingualism: Language and Cognition* 23(1): 37–38.

Nicenboim, Bruno, and Shravan Vasishth. 2016. Statistical methods for linguistic research: Foundational ideas—part II. *Language and Linguistics Compass* 10(11): 591–613. DOI 10.1111/lnc3.12207.

Nygaard, Lynne C., S. Alexandra Burt, and Jennifer S. Queen. 2000. Surface form typicality and asymmetric transfer in episodic memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26(5): 1228–1244.

Pan, Peter G. 1981. Prestige forms and phonological variation in Hong Kong Cantonese speech. Master's Thesis, The University of Hong Kong. DOI 10.5353/th_b4389299.

Pierrehumbert, Janet B. 2016. Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics* 2: 33–52.

Polinsky, Maria, and Gregory Scontras. 2020. Understanding heritage languages. *Bilingualism: Language and Cognition* 23(1): 4–20.

Qualtrics. (2022). "Qualtrics," https://www.qualtrics.com (Last viewed March 22, 2022).

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna. URL https://www.R-project.org/.

Rae, Megan, Paul Warren, and Jen Hay 2002. Goldilocks and the three beers: Sound merger and word recognition in NZE. *New Zealand English Journal* 16: 33.

Samuel, Arthur G., and Saioa Larraza. 2015. Does listening to non-native speech impair speech perception? *Journal of Memory and Language* 81: 51–71.

Statistics Canada, Government of Canada. 2017. *Census Profile, 2016 Census: Richmond, CY [Census subdivision], British Columbia and Greater Vancouver, RD [Census division], British Columbia (table)*. Ottawa, Statistics Canada Catalogue no. 98-316-X2016001. Released November 29, 2017. Accessed March 7, 2022. URL https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E.

Sumner, Meghan, and Reiko Kataoka. 2013. Effects of phonetically-cued talker variation on semantic encoding. *The Journal of the Acoustical Society of America* 134(6): EL485–EL491.

Sumner, Meghan, Seung Kyung Kim, Ed King, and Kevin B. McGowan. 2014. The socially weighted encoding of spoken words: A dual-route approach to speech perception. *Frontiers in Psychology* 4: 1015. DOI 10.3389/fpsyg.2013.01015

Sumner, Meghan, and Arthur G. Samuel. 2005. Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language* 52(3): 322–338. DOI 10.1016/j.jml.2004.11.004.

Sumner, Meghan, and Arthur G. Samuel. 2009. The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language* 60(4): 487–501.

Szakay, Anita, Molly Babel, and Jeanette King. 2016. Social categories are shared across bilinguals' lexicons. *Journal of Phonetics* 59: 92–109.

Tang, Sze-Wing, Fan Kwok, Thomas Hun-Tak Lee, Caesar Lun, Kang Kwong Luke, Peter Tung, and Kwan Hin Cheung. 2002. *Guide to LSHK Cantonese romanization of chinese characters*. Hong Kong: Linguistic Society of Hong Kong.

To, Carol K. S., Sharynne McLeod, and Pamela S. P. Cheung. 2015. Phonetic variations and sound changes in Hong Kong Cantonese: Diachronic review, synchronic study and implications for speech sound assessment. *Clinical Linguistics & Phonetics* 29(5): 333–353.

Todd, Simon, Janet B. Pierrehumbert, and Jennifer Hay. 2019. Word frequency effects in sound change as a consequence of perceptual asymmetries: An exemplar-based model. *Cognition* 185: 1–20.

Werker, Janet F., and John S. Logan. 1985. Cross-language evidence for three factors in speech perception. *Perception & Psychophysics* 37(1): 35–44.

Wong, Phoebe, and Molly Babel. 2017. Perceptual identification of talker ethnicity in vancouver english. *Journal of Sociolinguistics* 21(5): 603–628.

Wong, Shek-ling. 1941. *A Chinese syllabary pronounced according to the dialect of Cantonese*. Canton: Lingnam University.

Woods, Kevin JP, Max H. Siegel, James Traer, and Josh H. McDermott. 2017. Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics* 79(7): 2064–2072.

Yee, Paul. 2006. *Saltwater city: an illustrated history of the chinese in vancouver.* Douglas & McIntyre.

Yeung, H. S.-W. 1980. Some aspects of phonological variations in the Cantonese spoken in Hong Kong. Master's Thesis, The University of Hong Kong. DOI 10.5353/th_b3194810.

Zee, Eric. 1999. Change and variation in the syllable-initial and syllable-final consonants in Hong Kong Cantonese. *Journal of Chinese Linguistics* 27(1): 120–167.