Acquisition of Standard Chinese Neutral Tone under Variation Theory: A comparison of Beijing Mandarin speakers and Cantonese speakers

Xuanxin Wu* and Akitaka Yamada†

Osaka University

Abstract. In this study, we explore dialect speakers' acquisition of the neutral tone (NT) in Standard Chinese using the lens of variation theory. The study has several goals: (a) to verify the difficulty of NT acquisition in a group of speakers whose dialectal acquisition has not been quantitatively examined in previous literature (i.e., Cantonese speakers from Guangdong Province), (b) to observe NT acquisition using a group of words that have not been examined in previous literature, (c) to determine the impact of intra- or extralinguistic factor on the pronunciation of NT-bearing words, and (d) to explore lexical and speaker variations in NT acquisition. Several results have been observed: (a) Cantonese speakers from Guangdong Province (a non-NT dialect) have more difficulty pronouncing NT-bearing words than speakers of Beijing Mandarin (an NT dialect). (b) Rule-governed NTs are easier to pronounce than those that do not follow any established rules. (c) Speakers' pronunciation of NT-bearing words is strongly related to the existence of explicit rules predicting the use of an NT in the given word. Pronunciation is also strongly related to whether the participant is a native speaker of Cantonese or of Beijing Mandarin. Pronunciation of NTs is also weakly related to the tone of the syllable preceding the NT in a given word. (d) Variations in NT pronunciation among lexical items and individual participants indicate that simple generalizations should be avoided. Based on these results, we argue that, overall, Cantonese speakers' acquisition of the NT follows tendencies identified in a previous study (Liu et al. 2016); however, individual differences are non-negligible. Moreover, the lexical rules of NT -bearing words and lexical idiosyncrasies significantly impact NT acquisition.

1 Introduction

Most existing work on variation theory examines variations in dialect acquisition using data on English and other European languages, with a particular focus on language contact (Kroch 1978; Payne 1980; Chambers 1992; Babel 2010). Research on variation in Asian languages is relatively scarce. By examining dialect speakers' acquisition of the neutral tone (NT), this study elucidates variations in speakers' acquisition of various Chinese dialects.

*Graduate School of Humanities, 1-8 Machikaneyama, Toyonaka, Osaka, 560-0043, Japan
Email: u989146c@ecs.osaka-u.ac.jp
†Graduate School of Humanities, 1-8 Machikaneyama, Toyonaka, Osaka, 560-0043, Japan
Email: a.yamada.hmt@osaka-u.ac.jp

This work was supported by JSPS Grant-in-Aid For Scientific Research (C) (22K00507) for the second author.

Although the NT is found in Standard Chinese, it is not present in all Chinese dialects; for example, the tonal system of Beijing Mandarin includes the NT, whereas the Cantonese dialect uses citation tones (CTs) in cases where Standard Chinese uses NTs. It is well known that speakers of dialects without NTs have difficulty in acquiring the NT when they learn Standard Chinese as a second language (Liu et al. 2016). Therefore, the current study compares second-language acquisition of NT by Cantonese speakers to Beijing Mandarin speakers' pronunciation of NTs. We ask the following research questions:

(1) a. Can the difficulty of NT acquisition by speakers of a non-NT dialect be verified in a group of speakers whose dialectal acquisition has not been quantitatively examined in the literature (i.e., Cantonese speakers from Guangdong Province)?
b. Can the difficulty of NT acquisition also be observed with a group of words that have not been examined in previous literature?
c. What, if any, are the impacts of intra- or extralinguistic factor on speakers' pronunciation of NT-bearing words?
d. Are there lexical and speaker variations in NT acquisition?

For each research question, the following conclusions have been reached.

(2) a. Cantonese speakers from Guangdong Province (a non-NT dialect) have more difficulty pronouncing NT-bearing words than speakers of Beijing Mandarin (an NT dialect).
b. Rule-governed NTs are easier to pronounce than NT-bearing words for which no rules has been established.
c. The pronunciation of NT-bearing words is strongly related to the existence of explicit rules predicting the use of NT in the given word. Pronunciation is also strongly impacted by whether the participant is a native speaker of Cantonese or of Beijing Mandarin. Pronunciation is weakly related to the tone of the syllable preceding the NT in a given word.
d. Variations in NT acquisition among lexical items and individuals indicate that simple generalizations should be avoided.

2 Previous literature : factors affecting the use of the neutral tone

Standard Chinese is a typical tone language; a syllable's pitch contour can distinguish lexical meanings. Full and weak syllables are differentiated: full syllables are stressed and carry one of the four citation tones (CTs): Tone 1 (high), Tone 2 (low-high), Tone 3 (low), or Tone 4 (high-low). Weak syllables are unstressed and carry an NT (Chao 1968; Duanmu 2007). Previous studies have identified several important characteristics of NT-bearing words.

First, an NT cannot be distributed in an initial or isolated position, and its acoustic features — such as pitch contour and length — depend on the tone of the preceding syllable (Lin and Jingzhu 1980; Lee 2003).

Second, many NT-bearing words are disyllabic (Shi 1992); according to the Modern Chinese Lexical Database, approximately 67% of all NT-bearing words are disyllabic.

Third, from a morphological perspective, some studies have proposed a distinction between rule-governed and non-rule-governed NT-bearing words (Lu 2001; Jing 2002) (cf., from a phonological perspective, see Zhang 2022) on intrinsic and derived NTs). For some words, the use of an NT syllable is predicted; these are rule-governed NT-bearing words. For example, when a word is

formed via reduplication and refers to a relative (e.g., ma1ma0 meaning 'mother' /ma1 ma0/), the second syllable must receive an NT in Standard Chinese. Similarly, words with the suffix *-zi* or the suffix *-tou* are also rule-governed NT-bearing words. However, not all NT-bearing words are governed by such rules; these unpredictable NT-bearing words are non-rule-governed NT-bearing words.

Finally, language policies can have a significant impact on language contact and language change (Spolsky 2004). Significant dialectal variation has led to a strong need for nationwide pronunciation norms in China. However, such norms were not successfully established until the second half of the last century, when the government promoted Standard Chinese, which is based on Beijing Mandarin. Despite this standardization, dialectal variation is non-negligible. For example, in acoustic and perceptual experiments, Liu et al. (2016) show that Hong Kong Cantonese speakers (unlike Beijing Mandarin speakers) do not acquire "short/weak" features of NTs. They examined three types of rule-governed NT-bearing words; the target words were assigned fixed positions in designed monologues. An analysis of acoustic features of NT (such as pitch contour and duration) demonstrated that Hong Kong Cantonese speakers produced NTs with a lower pitch register and a narrower pitch range than Beijing Mandarin speakers. However, some issues related to dialect speakers' NT acquisition still need to be examined. First, since Liu et al. (2016) only examined speakers of Hong Kong Cantonese, it is still unclear whether their findings can be generalized to other Cantonese speakers (e.g., those born and raised in Guangdong Province, the birthplace of Cantonese). Second, since Liu et al. (2016) focused exclusively on rule-governed NT-bearing words, additional research on non-rule-governed NT-bearing words is needed.

3 Data and Methods

To better understand the role of NT in dialect acquisition, we conducted an experiment to answer the aforementioned research questions in (1). Unfortunately, all the combinations of the aforementioned factors is beyond the scope of this paper. Thus, for practical purposes, this study focuses only on the following conditions.

First, our study is limited to disyllabic words; participants' pronunciation of 24 NT-bearing words and 24 corresponding CT-bearing words were compared. The NT-bearing words included in the study have an NT on the second syllable in Standard Chinese; all included CT-bearing words take a CT on the second syllable. Each NT-bearing word corresponds to a CT-bearing word with the same Chinese character in the second syllable (e.g., ma1ma0 meaning 'mother' /ma1 ma0/ versus gan1ma1 meaning 'Chinese godmothers' /kan1 ma1/). In all included words , the first syllable takes one of the four CTs and is treated as an independent variable.

Second, this study compares the pronunciation of Cantonese speakers to that of Beijing Mandarin speakers; speakers of other dialects were excluded. Eight participants (four Beijing Mandarin speakers and four Cantonese speakers) were asked to read sentences that included the target words.[1] For each word, the appropriateness of each speaker's recorded pronunciation of the relevant syllable was evaluated using a five-point Likert scale by native speakers of NT dialects (the dependent variable).

Third, to explore stylistic variation (Chambers and Schilling 2013), the present study exam-

[1]Cantonese speakers: one of them was born and raised in Hong Kong, the rest of three speakers were born and raised in Guangdong province.

ines participants' pronunciation of NT/CT words in two distinct speech situations: a dialogue (conversation) and a monologue. In a controlled experimental environment, the following independent variables were measured and then analyzed using an ordered logistic regression model. Parameters were estimated using `ordinal` on R.

(3) Fixed-effects variables:
   a. SPEECH STYLE (`SS`): a dummy variable indicating whether the word is embedded in a dialogue (a conversation) or in a monologue (a diary entry)
   b. RULE (`Rule Type`): a dummy variable indicating whether there is an explicit rule predicting the use of NT in the given word
   c. FIRST SYLLABLE (`TS1`): the tone of the first syllable of the given word
   d. DIALECT (`DL`): a dummy variable indicating whether the participant is a Cantonese speaker (born and raised in Guangdong Province or in Hong Kong) or a Beijing Mandarin speaker (born and raised in Beijing city)

(4) Random-effects variables:
   a. ITEM: a random effect representing the uniqueness of the given word
   b. PARTICIPANT: a random effect representing unique characteristics of the speaker
   c. EVALUATOR: a random effect representing the evaluator's unique characteristics

(5) Model

$$\text{logit}(P(y_{i(jkl)} \leq t)) = \theta_t - \beta_{ss}x_{ss,i(jkl)} - \beta_{\textit{rule type}}x_{\textit{rule type},i(jkl)} - \beta_{ts1}x_{ts1,i(jkl)} - \beta_{dl}x_{dl,i(jkl)} - \text{Item}_{0j} - \text{Participant}_{0k} - \text{Evaluator}_{0l}$$

In this model, $y_{i(jkl)}$ represents the *i*-th evaluation of the *j*-th item pronounced by the *k*-th participant and evaluated by the *l*-th evaluator (dependent variable). θ represents the threshold ($t \in \{1, 2, 3, 4\}$). x represents three fixed-effect variables that, along with their β coefficients, function as predictors. `Item`, `Participant` and `Evaluator` represent three random-effects variables.

4 Results

4.1 Fixed-effects variables

The estimated values for the fixed-effects parameters for NT-bearing words are shown in Figure 1. The negative value for `DL` shows that the evaluators tend to give lower scores for Cantonese speakers' pronunciation of NTs than Beijing Mandarin speakers' pronunciation, suggesting that speakers of non-NT dialects have more difficulty acquiring NT in a second language. This result aligns with those of Liu et al. (2016) regarding Hong Kong Cantonese speakers (*answering research question* 1-a).

However, this does not mean that Cantonese speakers never acquire this phonetic element, as several factors also affect the tendency.

First, when explicit rules are present (`Rule Type`), it is easier to pronounce the NT than in NT-bearing words for which no rule has been established (*answering research question* 1-b).

Second, speech style (`SS`) also influences NT pronunciation. The positive parameter value for `SS` indicates that the relevant syllable is more likely to be perceived as an NT in a monologue than

	β^	SE	Z.value	P value	95% CI (low)	95% CI (high)	OR
SS|Monologue	0.154819	0.047595	3.253	0.00114**	0.06	0.25	1.17
Rule Type|RW	2.616867	0.354657	7.379	1.60E-13***	1.92	3.31	13.69
Rule Type|Tou	2.022640	0.353343	5.724	1.04E-08***	1.33	2.72	7.56
Rule Type|Zi	2.473055	0.354131	6.983	2.88E-12***	1.78	3.17	11.86
TS1|T2	-0.816071	0.353918	-2.306	0.02112*	-1.51	-0.12	0.44
TS1|T3	0.069024	0.353078	0.195	0.84501	-0.62	0.76	1.07
TS1|T4	0.007522	0.352986	0.021	0.983	-0.68	0.70	1.01
DL|Cantonese	-1.501914	0.254793	-5.895	3.75E-09***	-2.00	-1.00	0.22

Table 1: Estimates of fixed-effects parameters for NT-bearing words

in a dialogue. However, while the P-value for this estimate is small, the effect size and the range of the CI is close to 0. This is interpreted to mean that the influence is marginal at best, indicating that the impact of `SS` on NT pronunciation is not as important as the other variables.

Third, the phonological environment affects NT pronunciation: when the preceding syllable carries Tone 2, the relevant syllable is less likely to be perceived as bearing an NT (`TS1|T2`).

To answer research question 1-c, we measured the impact of the morphological factor: whether there are explicit rules predicting the use of NT in a given word. This factor has a strong impact on NT pronunciation; the results show that all types of rules we choose (`Rule Type`) have large effect sizes. The extralinguistic factor, or whether the participant is a Cantonese speaker or a Beijing Mandarin speaker, also strongly influences NT pronunciation, as shown by the highly negative estimated value for `DL`. However, the studied phonological and sociolinguistic factors — the tone of the preceding first syllable (`TS1`) and speech styles (`SS`) — have a weak relationship with NT pronunciation, since the estimated values of these parameters are close to zero.

4.2 Random-effects variables

Lexical variations also affect NT pronunciation. Figure 1 summarizes the random effects for CT-bearing words and NT-bearing words. While most of these appear to be randomly distributed, there is one outlier: `bao1zi3` meaning 'spore' (/pɑu1 ts3/), which has a strong tendency to be perceived as an NT. Since the difference between Cantonese and Beijing Mandarin speakers is captured by the fixed effects, this tendency holds even for Cantonese speakers. In addition, some NT-bearing words are easily perceived as NT-bearing words even when pronounced by Cantonese speakers, as suggested by the left panel of Figure 1. One example is `bo4he0` meaning 'mint' (/puo4 xɤ0/). One individual's speech may also be evaluated differently in different cases: the left panel of Table 2 shows that the variance explained by `Evaluator` is as large as that explained by `Item`. However, on the right panel, the substantial value of `Evaluator` indicates high variance in the evaluation of CT-bearing words. Moreover, although the variance in `Participant` is not as large as that of the other two random-effects variables, one participating Cantonese speaker seems to be surprisingly good at articulating NTs. While it is generally assumed that phonetic features that do not exist in a speaker's first language are difficult to acquire (Face 2006; Fatemi

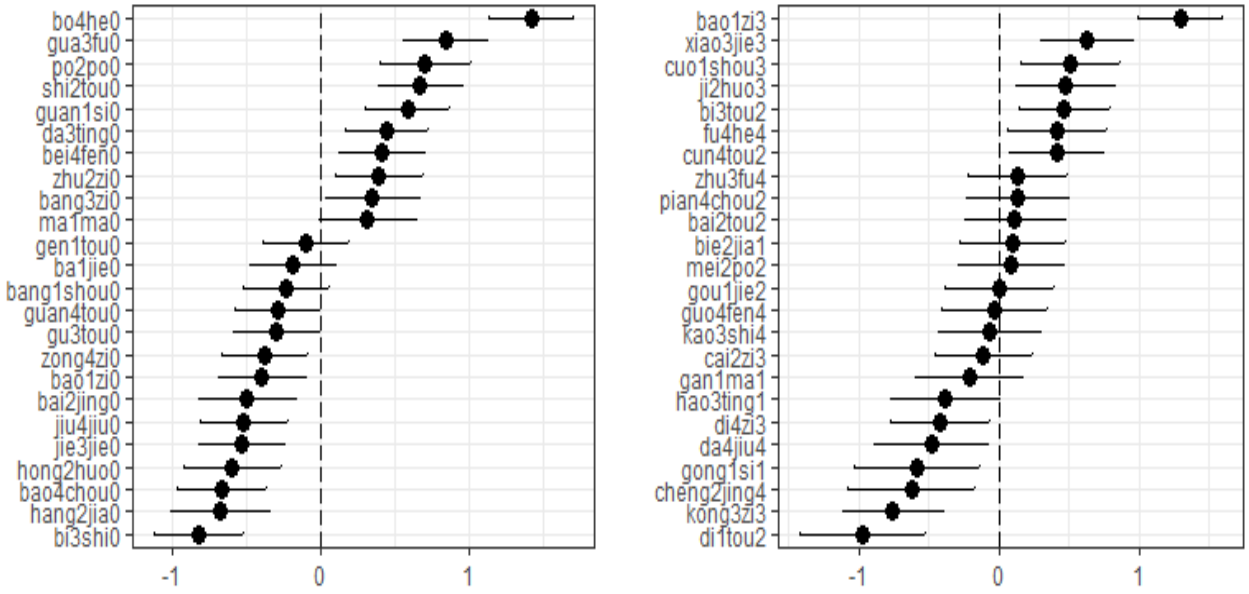

Figure 1: Estimates of random-effects variance (Left: NT-bearing words; Right: CT-bearing words)

	Variance
Item	0.36
Evaluator	0.30
Participant	0.12

	Variance
Item	0.29
Evaluator	1.49
Participant	0.07

Table 2: Random-effects variance (Left: NT-bearing words; Right: CT-bearing words)

et al. 2012), this finding suggests that some Cantonese speakers may have effectively acquired NTs, indicating that simple generalizations should be avoided (thus answering research question 1-d).

5 Conclusion and future directions

This study has addressed some gaps in previous studies of NT acquisition by speakers of Chinese dialects. More specifically, the present study has examined acquisition of non-rule-governed NTs by non-Hong-Kong Cantonese speakers. By analyzing a broad range of factors, this study provides more detailed information about NT acquisition by Chinese dialect speakers. Several conclusions may be drawn from the present findings. First, although some overall similarities and differences in performance have been observed, we cannot draw an arbitrary line between Cantonese speakers' and Beijing Mandarin speakers' pronunciation of NTs. Second, dialect speakers' NT acquisition is influenced by lexical rules and lexical idiosyncrasies. However, the cause of these lexical idiosyncrasies is still unclear. Moreover, some other factors that may also influence NT acquisition still need to be investigated in more depth.

The present study has some limitations as well. First, we have only examined two types of speech, monologues and dialogues; other speech styles should be examined in future research.

Second, although we have found that T2 on the first syllable impacts the ease of NT pronunciation, we have not presented a theoretical explanation for this exception. To develop a better understanding of NT acquisition, future studies should explore these issues as well.

References

Babel, Molly (2010) "Dialect divergence and convergence in New Zealand English," *Language in Society*, Vol. 39, pp. 437–456.

Chambers, J. K. (1992) "Dialect acquisition," *Dialect Acquisition*, Vol. 68, No. 4, pp. 637–705.

Chambers, J. K. and Natalie Schilling (2013) *The handbook of language variation and change*, Chichester, England: John Wiley & Sons, 2nd edition.

Chao, Yuen Ren (1968) *A grammar of spoken Chinese*, Berkeley: University of California Press.

Duanmu, San (2007) *The phonology of standard Chinese*, Oxford: OUP.

Face, Timothy L (2006) "Intervocalic rhotic pronunciation by adult learners of Spanish as a second language," in *Selected proceedings of the 7th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages*, pp. 47–58.

Fatemi, M. Ali, Atefe Sobhani, and Hamzeh Abolhasani (2012) "Difficulties of Persian learners of English in pronouncing some English consonant clusters," *World Journal English Language*, Vol. 2, No. 4, pp. 69–75.

Jing, Song (2002) *Xian Dai Han Yu Qing Sheng Dong Tai Yan Jiu [Dynamic study of Neutral Tone in modern Chinese]*, Beijing: Min Zu Chu Ban She [Publishing House of Nationalities].

Kroch, Anthony S. (1978) "Towrad a theory of social dialect variation," *Language in Society*, Vol. 7, No. 1, pp. 17–36.

Lee, Wai-Sum (2003) "A phonetic study of the neutral tone in Beijing Mandarin," in *Proceedings of the 15th International Congress of Phonetic Sciences (ICPHS 2003), Barcelona*, pp. 1121–1124.

Lin, maocan and Yan Jingzhu (1980) "Bei Jing Hua Qing Sheng De Sheng Xue Xing Zhi [The acoustic properties of the Neutral Tone in Beijing Mandarin]," *Fang Yan [Dialect]*, No. 3, pp. 166–178.

Liu, Lei, Huang Nan, and Gu Wentao (2016) "Mandarin neutral tone by native speakers and Cantonese L2 learners," in *2016 10th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, pp. 1–5, IEEE.

Lu, Yunzhong (2001) *Qing Sheng He Er Hua [Neutral Tone and r-ending retroflexion]*, Beijing: Shang Wu Yin Shu Guan [The Commercial Press].

Payne, Arvilla C. (1980) "Factors controlling the acquisition of the Philadelphia dialect by out-of-state children," in *Locating language in time and space*, pp. 179–218, New York: Academic Press.

Shi, Dingguo (1992) "Pu Tong Hua Zhong Bi Du De Qing Sheng Ci [Words that must be pronounced as Neutral Tone in Putonghua]," *Yu Wen Jian She [Language Planning]*, No. 6, pp. 28–34.

Spolsky, Bernard (2004) *Language policy*, Cambridge: Cambridge University Press.

Zhang, Yixin (2022) "Neutral Tone in Mandarin: representation and interaction with utterance-level prosody," Ph.D. dissertation, University of Cambridge, Cambridge.