Addressing Outstanding Questions of the Mandarin Syllable

A Dissertation Presented

by

Yang Liu

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Linguistics

Stony Brook University

December 2022

Stony Brook University

The Graduate School

Yang Liu

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Lori Repetti – Dissertation Advisor
Chair, Professor, Department of Linguistics

Jeffrey Heinz - Chairperson of Defense
Professor, Department of Linguistics

Tatiana Luchkina
Assistant Professor, Department of Linguistics

Likan Zhan
Associate Professor, School of Communication Science
Beijing Language and Culture University

This dissertation is accepted by the Graduate School

Celia Marshik
Dean of the Graduate School

Abstract of the Dissertation

Addressing Outstanding Questions of the Mandarin Syllable

by

Yang Liu

Doctor of Philosophy

in

Linguistics

Stony Brook University

2022

The following questions central to Mandarin syllables remain controversial and are not yet answered. (a) What are the general phonotactic principles – both local and long-distance ones – that Mandarin syllables abide by? (b) What grammar can best account for native speakers' phonotactic intuition on what represents a well-formed syllable in Mandarin? (c) What is the status of the pre-vocalic glide [j] within the syllable: is it segmental (Cj) or, else, is it palatalized, i.e., realized as a secondary articulation (Cj)? This thesis seeks to contribute to the scientific understanding of the questions listed in (a-c), by implementing a series of experimental and computational methodologies to investigate syllables in Mandarin Chinese.

In this dissertation, I outline the phonological principles underlying Mandarin syllables, use a nonword acceptability judgment task to test linguistically naive listeners' sensitivity to constraint violations which account for systemic vs. accidental gaps in the syllable inventory, and model the acceptability judgment data using categorical and gradient grammars to find out what grammar can best account for how speakers rate Mandarin syllables. I question whether a

gradient grammar is more or less predictive of speakers' grammaticality judgments than a categorical grammar, and if a gradient grammar derived from the Hayes and Wilson Phonotactic Learner ("data driven") is more or less predictive of speakers' grammaticality judgements than one derived manually ("phonologically driven"). The results show that none of the four grammars can predict the ratings for highly ungrammatical tokens. Mandarin speakers are more sensitive to the distinctions among "more grammatical" nonwords.

Regarding the status of the prevocalic glide [j] in Mandarin, I examine the production data from 5 Mandarin native speakers, and I find that the status of the Mandarin glide is more likely to be palatalized, instead of segmental.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Generative phonologists (Cheng 1973; Lin 1989; Duanmu 1990, 2000, 2008) maintain that a Chinese syllable takes on the format (C) (G) V (X). In other words, a first onset consonant, a glide, a nuclear vowel and a coda consonant or an offglide coda together make up a maximum Chinese syllable. Among these four elements, all except the nuclear vowel are optional in a syllable. The topics studied by the previous literature on the Mandarin syllable include the relevant phonotactic constraints (Gong & Zhang, 2019, 2021; Duanmu: 2000, 2007; Duanmu & Yi, 2015), the experiments of nonword grammaticality judgment (Myers & Tsay, 2005, 2015; Gong & Zhang: 2018, 2019, 2021), the status of Mandarin glides (Wang, 1999; Chao, 1934; Duanmu: 2000, 2007), among many others. However, the following questions central to Mandarin syllables remain controversial and are not yet answered. (a) What are the general phonotactic principles – both local and long-distance ones – that Mandarin syllables abide by? (b) What grammar can best account for native speakers' phonotactic intuition on what represents a well-formed syllable in Mandarin? (c) What is the status of the pre-vocalic glide [j] within the syllable: is it segmental (Cj) or, else, is it palatalized, i.e., realized as a secondary articulation (C^j)? This thesis seeks to contribute to the scientific understanding of the questions listed in (a-c), by implementing a series of experimental and computational methodologies to investigate syllables in Mandarin Chinese.

More specifically, this dissertation does the following: (1) generalize the phonological principles underlying Mandarin syllables; (1.1) review the existing phonotactic constraints on syllable well-formedness in Mandarin; (1.2) propose novel CG and CV constraints consistent with Mandarin data, to account for certain systemic gaps in the existing syllables inventory; (2) using a syllable acceptability judgment task, test linguistically naive listeners' sensitivity to constraint violations which account for systemic vs. accidental gaps in the syllable inventory; (3) model the acceptability judgment data using categorical and gradient grammars to find out what grammar can best account for how speakers rate Mandarin syllables and their phonotactic knowledge on syllables. In probing this question, I tackle the theoretical issue of gradient vs. categorical grammars. Is gradient grammar more or less predictive of speakers' grammaticality judgments than categorical grammar? Is a gradient grammar derived from the Hayes and Wilson Phonotactic Learner (HWPL) ("data driven") more or less predictive of speakers' grammaticality judgements than one derived manually ("phonologically driven")? (4) explore the problematic status of the glide [j] in Mandarin syllables by determining whether it is segmental or palatalized as a secondary manner of articulation, by examining the production data from 5 Mandarin native speakers, with a special focus on a number of acoustic correlates.

This dissertation consists of seven sections organized as follows. In sections 1 and 2, I provide an overview of the research questions and previous literature. In §3, I conduct a corpus data analysis to explore all the local and distant systematic constraints of Mandarin syllables, and review the Mandarin phonotactic constraint discussed in Lin (1989), Duanmu (2000, 2007, 2015), Duanmu & Yi (2015), Gong & Zhang (2018, 2019), among many others. I follow the methodology established in these earlier studies and aim to summarize and further explore the general phonotactic principles of the Mandarin syllables. More specifically, I identify and critically analyze a number of constraints on Mandarin syllable well-formedness and conclude that they can be reduced to the following three principles: the OCP constraints NoLabLab and NoHiHi which prohibit adjacent segments with identical features (labial-labial or high-high), and the anti-OCP constraints on Backness Agreement which require adjacent segments to share the feature [back].

Of the 1254 possible (C)(G)V(X) syllables, only 402 are actually attested (32%) in everyday language use. The investigation presented in §3 aims to determine whether the missing Mandarin syllables are accidental gaps with 0 constraint violations, or systematic gaps with one or more violations. This question has been previously analyzed for other languages in studies by Halle (1962), Coetzee (2008), among many others. To illustrate, Halle (1962) defines systematic gaps as syllables which violate principled phonotactic constraints, while accidental gaps refer to other unattested forms. Coetzee (2008) also proposes that if the missing forms can be explained by general and natural phonological properties and constraints, they are systematic gaps (see also Frisch & Zawaydeh, 2001). Of the 852 unattested Mandarin syllables, 786 (92.25%) can be accounted for by the three principles identified in §3 (“systematic gaps”), with only 66 (7.75%) unattested syllables considered as “accidental gaps”.

In §4, I report on a nonword acceptability rating experiment to gauge speakers’ sensitivity to the phonotactic constraints examined in §3. In the previous studies examining the systematic gaps in the existing Mandarin syllable inventory (Gong & Zhang 2019, 2021, Myers & Tsay, 2005, 2015), the systematic gaps are all analyzed as being the same type. For instance, Gong & Zhang (2019, 2021) found the acceptability ratings of systematic gaps is lower than accidental gaps, allophonic gaps and tonal gaps; but they didn’t further divide systematic gaps into subcategories. Different from that, I exclude allophonic gaps and tonal gaps, and I include distinctions among the systematic gaps based on factors like the number of violated constraints and weight of constraints in the current experiment.

Some linguists like Frisch & Zawaydeh (2001) suggest that grammar is derived from lexical statistics. For instance, structures with high frequency are perceived to be more grammatical. Bybee (2001) also argues that the frequent structures could be automatically grammaticalized with no regard to phonetic factors. Coetzee (2008) maintains that statistical frequency is not enough to account for the speakers’ distinct ratings on structures with 0 occurrence. In his experiment, the missing word [spVp] is rated with lower acceptability than the missing word [skVk]. This shows that the constraint *spVp is ranked higher than *skVk in speakers’ phonological grammar, which is independent of frequency (see also Gong & Zhang, 2019). To probe this issue using Mandarin data, the nonword acceptability rating experiment is carried out in §4 to collect the perception data, which is then used to model the phonotactic knowledge of the Mandarin native speakers. In §5, I compare six different models of phonotactic grammars by analyzing their correlations with speakers' acceptability ratings, in order to determine which model can best account for the results of my syllable acceptability experiment. In Gong & Zhang (2019, 2021)'s study on Mandarin nonword acceptability ratings, they argue that non-word judgment is gradient: this means that syllables representing certain systematic gaps in the Mandarin inventory receive lower acceptability ratings than those representing accidental gaps (see also Myers & Tsay, 2005, 2015). I expand Gong & Zhang's investigation by further refining the categorization of the systematic gaps, and by comparing the various models of grammar to account for the experimental results reported in §4. To further explore what grammar best reflects native speakers' phonotactic judgment on syllable well-formedness in Mandarin Chinese, I compare: (a) two gradient grammars – a data-driven gradient grammar in which I calculate the penalty score of each Mandarin syllable via the UCLA Hayes and Wilson Phonotactic Learner (HWPL) (Hayes & Wilson 2008) and a manually constructed gradient grammar; as well as (b) two categorical grammars represented via "0" (ungrammatical, violating one or more constraint) and "1" (grammatical, violating 0 constraint), which are based on either the UCLA Hayes and Wilson Phonotactic Learner (HWPL) or the phonological generalizations defined in §3; and (c) two cumulative categorical grammar which counts the number of violated constraints, either via the UCLA Hayes and Wilson Phonotactic Learner (HWPL) or based on the phonological generalizations defined in §3. The correlational analyses between the values (described above) representing various grammars and the native speakers' acceptability ratings on syllables is carried out to determine which grammar model can best account for the acceptability ratings of the systematic vs. accidental gaps reported in §3, and thereby can serve as a proxy of the native speakers' phonotactic knowledge regarding what represents a well-formed syllable in Mandarin.

To shed light on the problematic status of the glide [j] in the Mandarin syllable, a phonetics experiment involving the Mandarin glide [j] is carried out in §6. Some of the earlier studies (Duanmu 2000: 28, 2007) treat CG as a single sound which stands in the same timing slot, while others (Wang, 1999: 128, 129) argue that C and G are two separate segments. Consistent with the previous work, two theoretical possibilities are entertained and critically evaluated in §6 of the present thesis: (1) the initial consonant and the following glide are regarded as the primary and secondary articulations respectively (C^j); or (2) the glide is segmental, i.e., should not be treated as a secondary manner of articulation, as has been previously reported for other languages like Russian (Ladefoged & Maddieson, 1996: 364; Suh & Hwang, 2016). To explore the question using phonetic evidence, following Suh & Hwang's approach (2016), the production data with respect to the Mandarin glide [j] was collected in an elicited (read) production study with 5 native Mandarin speakers. Four phonetic features of the Mandarin glide were measured and quantitatively analyzed: the duration, the F2 slope change "throughout the vocoid (glide + vowel)", the vowel-to-glide coarticulatory effect and the glide-to-vowel coarticulation. The experimental data was then compared with the results previously reported for Russian and Korean (Suh & Hwang, 2016) to evaluate the status of the Mandarin glide [j].

Chapter 2

Background

In sections §2.1, §2.2, §2.3 and §2.4 below, previous studies on Mandarin phonetic inventory and syllables, Mandarin phonotactic constraints, the experiments of acceptability ratings, categorical grammars as well as gradient grammars are discussed.

2.1 Inventory and Syllables in Mandarin Chinese

The format of a Mandarin syllable can be displayed as (C) (G) V (X) (Cheng 1973; Lin 1989; Duanmu 1990, 2000, 2008). A Mandarin syllable includes an optional initial onset consonant, an optional glide, a nuclear vowel, and an optional coda consonant or offglide coda. The following sections illustrate the consonants, vowels, and glides in Mandarin Chinese, as well as GV and VG sequences and syllable structure. (See Appendix 1 for a chart of the feature specifications adopted in this paper.)

2.1.1 Consonants

The consonant inventory in Mandarin Chinese is displayed in (1), which is adapted from Luo & Wang (1981), Chao (1968) and Duanmu (2000: 26, 49; 2000: 37, 50). I only include the features that are relevant for the discussion in this paper, namely the place features, the backness feature values and the continuancy feature values.

(1) Mandarin consonant inventory

	Place Manner	LAB	COR				DOR
		Labial	Alveolar	Dental	Retroflex	Palatal [+hi]	Velar [+hi]
		*Dor [0back]	*Dor [0back]	*Dor [0back]	*Dor [0back]	Cor, Dor [-back]	Dor [+back]
Obstruent	Stops [-cont]	pʰ, p	tʰ, t				kʰ, k
	Fricatives [+cont]	f		s	ʂ, ʐ	ɕ	x
	Affricates [±cont]			t͡sʰ, t͡s	ʈ͡ʂʰ, ʈ͡ʂ	t͡ɕʰ,t͡ɕ	
Sonorant	Nasals [-cont]	m	n				ŋ
	Liquids [+cont]		l				

Among the 22 consonants in Mandarin, the velar nasal [ŋ] is the only one that cannot occur at the onset position, which is similar to English. The nasals in Mandarin Chinese include labial [m], alveolar [n] and velar [ŋ]. There are six fricatives: labio-dental [f], dental [s], palatal [ç], velar [x] and the retroflex [ʂ, ʐ]. The stops are either aspirated: [pʰ, tʰ, kʰ], or unaspirated [p, t, k]. In

Mandarin, affricates include dentals [t͡s, t͡sʰ], palatals [t͡ɕʰ,t͡ɕ] and retroflexes [t͡ʂʰ, t͡ʂ], and I follow Riggle (2011) and label affricates as [±cont]. Besides nasals and vowels, the sonorants in Mandarin Chinese include one liquid: the alveolar lateral [l], although the lateral [l] can be missing in many regional varieties.

In the consonant inventory (1), the voiced retroflex sound is transcribed as the fricative [ʐ], which follows Karlgren's (1915: 26) and Dong's (1958: 39) analyses. Yet, in Duanmu's (2000: 26) study, this sound is transcribed as a retroflex liquid [r] instead of fricative [ʐ] (see also Fu 1956: 3; L. Wang, 1980). Duanmu (2000: 26)'s argument for adopting [r] rather than [ʐ] is that fricatives in Standard Chinese [f, s, ʂ, ɕ, x] are all [-voice], so it is "phonologically odd" to include [ʐ] as the only [+voice] one. However, the present paper maintains that the voiced fricative [ʐ] patterns with its voiceless counterpart [ʂ] in the Mandarin consonant inventory. The voiced/voiceless pairs [s/z, ʂ/ʐ] and the [±anterior] [s/ʂ, z/ʐ] exist in many Chinese varieties, like Shangfu. Besides, phonologically, the voiced retroflex behaves more like an obstruent than a sonorant; for instance, all initial sonorants in Mandarin can be followed by [j], yet the voiced retroflex + [j] is missing just like other fricatives + [j] sequences.

According to Odden (1991: 261, 262), the dorsal place node covers [high], [low] and [back] values (see also Clements 1985, Sagey 1986, McCarthy1988, Kenstowicz 1994), and we will see that the [back] plays a crucial role in Mandarin phonotactics. Mandarin palatals [t͡ɕʰ,t͡ɕ, ɕ] are both dorsal and coronal with [-back] features (Hayes 2011: 95-97, Riggle 2011), while the velars [kʰ, k, x ŋ] are dorsal with [+back] features. In (1), labials and non-palatal coronals do not carry a [back] value and are thus indicated as [0back].

2.1.2 Vowels

In Mandarin Chinese, there are 5 phonemic vowels (adapted from Duanmu 2000: 37, 50), more specifically, five monophthongs: [i, y, u, ə, a]. I adopt Gong & Zhang (2019)'s analysis and posit 8 allophonic vowels, which, in addition to the 3 high vowels [i, u, y], include 3 mid vowels [o, e, ə] and 2 low vowels [a, ɑ]. The distinctive features of the Mandarin vowels relevant for this study are given in (2). I follow Duanmu (2000: 47) and posit a dorsal place node for all vowels, a labial place node for the rounded vowels. Based on Hayes's feature chart, the vowels [a, ə] are both central vowels with [-front, -back] feature. Xu (1980: 183) proposes that the vowel [a] can have allophonic front, back or central variants. Wang (1993) leaves the frontness feature of the low vowel [a] unspecified when it is in a closed syllable. Lin (1989: 52) also argues that the underlying vowels [a, ə]'s backness and roundness feature are unspecified. Duammu (2007: 38-39) maintains that the Mandarin mid vowel [ə] and low vowel [a] can change their backness feature in different contexts, as displayed in the allophonic rules in (3). Therefore, the backness features of the vowels [a, ə] are left unspecified in Duanmu's paper (2007: 47). Following Hayes's feature chart and Duanmu's (2007) study, this study argues that the Mandarin vowels [a, ə] are central without a specification of their backness feature. We can account for the 8 allophonic vowels deriving from the 5 phonemic vowels as described in the 6 rules in (3), whereby /a/ and /ə/ are realized differently in different contexts.

(2) Distinctive Feature Chart of Mandarin Vowels (following Hayes, Duanmu 2007)

	a	e	o	ə	ɑ	i	u	y
LABIAL	-	-	+	-	-	-	+	+
round	0	0	+	0	0	0	+	+
CORONAL	-	-	-	-	-	-	-	-
DORSAL	+	+	+	+	+	+	+	+
high	-	-	-	-	-	+	+	+
low	+	-	-	-	+	-	-	-
back		-	+		+	-	+	-

(3) allophonic rules for vowels (adapted from Gong & Zhang, 2019)

a. /ə/ → [o] / w _ #, or _ u

b. /ə/ → [e] / j, ɥ _ #,

c. /ə/ → [ə] / ELSEWHERE

d. /a/ → [e] / j, ɥ _ n

e. /a/ → [ɑ] / _ u, ŋ

f. /a/ → [a] / ELSEWHERE

There are 3 allophonic variants of the schwa vowel /ə/, namely, [o, e, ə]. The vowel /ə/ varies the backness and roundness feature depending on the adjacent segment. In (3a), we see that if schwa is followed or preceded by the glide/vowel [w, u], it abides by roundness and backness harmony changing to the mid round back vowel [o]. However, if it is preceded by high front glides [j, ɥ], it is fronted to the front mid vowel [e] (3b). Elsewhere, the schwa stays the same (3c). This can be exemplified by the word “阔 kwo” (wide), “狗 gou” (dog) and “列 lje” (column), where the underlying vowel /ə/ is realized as [o] and [e] in the surface forms. Yet in the word “冷 ləŋ” (cold), the underlying vowel /ə/ stays the same in the surface form.

According to some previous analyses (Chao 1968, C. Cheng 1973, Xu 1980: 33, Duanmu 2000), the mid central vowel can take on the additional form [ɣ]. Fu (1956: 6) argues that [ɣ] appears after a labial sound, and Duanmu (2000: 39) maintains that after labials, both back vowel [o] and mid vowel [ɣ] can occur. However, forms like [mwo] (末 ‘end’) and [mwɣ] (末 ‘end’) are non-contrastive since the sound [ɣ] will be labialized after [w]. For the following analysis, I transcribe the vowels after a labial glide [w] or [ɥ], as one of the variants [o] without differentiation between [o, ɣ]

The low central vowel /a/ has three allophonic variations: [e, ɑ, a]. The vowel /a/ varies its backness and height features depending on the adjacent segments. If preceded by a palatal glide ([j] or [ɥ]) and followed by [n], it is fronted and raised to the front mid vowel [e] (3d). If followed by the segments [u, ŋ], it changes to the low back vowel [ɑ] (3e). The underlying /a/ does not change elsewhere (3f). For instance, in the words “年 njen” (year) and “劳 lɑu” (labor), the underlying vowel /a/ is realized as [e] and [ɑ] in the surface form, respectively. Yet in the word “排 pai” (row), the underlying vowel /a/ stays the same in the surface form.

2.1.3 Glides

The three high vowels [i, y, u] can be realized as glides (before or after the nuclear vowel) or monophthongs (as the nuclear vowel) in Standard Chinese; however, as glides, they are represented differently. Following the usual practice among linguists, the prevocalic glides are represented as [j, w, ɥ] while the postvocalic ones are represented as word-final vowels [i, u, y]. The three glides in Standard Chinese are exemplified in (4) (Duanmu 2000: 25).

(4)

编 [pjan]=[pian]	‘weave’
准 [tʂwən]=[tʂuən]	‘accurate’
学 [çɥe]=[çye]	‘study’
推 [tʰwəi]=[tʰuəi]	‘push’
丢 [tjou]=[tiou]	‘lose’

The features of glides [j, w, ɥ] in Mandarin are not contrastive with their high vowel counterparts [i, y, u]. The relevant features of glides and vowels are illustrated in (5).

(5) Glide-vowel correspondence table in Mandarin (adapted from Duanmu 2000)

glide ←→ high vowel		
[j]	[i]	[-back, +high, -ro]
[ɥ]	[y]	[-back, +high, +ro]
[w]	[u]	[+back, +high, +ro]

In fact, due to the non-contrastive correspondence between the high vowels [i, y, u] and the glides [j, ɥ, w] in terms of their features, some phonologists (Wang 1993, Pulleyblank 1983) propose that there are no high vowels in Mandarin Chinese, because their underlying representation are all glides. Pulleyblank (1983) even argues that the low vowel [a] is a pharyngeal glide as well. The present paper follows Duanmu (2000)’s argument that the glides in Mandarin include [j, ɥ, w], and are positional variants of high vowels.

(6) Distinctive Feature Chart of Mandarin Glides

	j	w	ɥ
sonorant	+	+	+
continuant	+	+	+
LABIAL	-	+	+
round	-	+	+
CORONAL	-	-	-
anterior	0	0	0
distributed	0	0	0
DORSAL	+	+	+
high	+	+	+
low	-	-	-
front	+	-	+
back	-	+	-

The place feature specification of glides and high vowels is debated. Riggle (2011) labels the glides [j, ɥ, w] as dorsal and non-coronal (see also Kenstowicz 1994, Duanmu 2000, Hayes 2011), while others specify the front glide [j] as coronal (Kochetov 2016; Meeussen 1959; Broselow & Niyondagara 1990; Hume 1990, 1994, 1996; Ntihirageza 1993). In this paper, I adopt the analysis of the glides as dorsal and non-coronal; however, an alternative analysis is investigated in §3.1 with the glides [j, ɥ] having a coronal place specification.

The articulator features of the three glides in Mandarin Chinese adopted in this paper are illustrated in (6). Since the high glides are all dorsal, the backness feature, which is relevant for the phonotactic analysis, is indicated. The rounded glides [ɥ, w] also have a labial feature. Although they are featurally identical to the high vowels, the glides behave differently from high vowels. For instance, the CV sequence [fu] is allowed while the CG group *[fw] is penalized.

2.1.4 VG and GV

Although [i, y, u] are all high vowels, only [i, u] can stand in the second slot in a diphthong (Duanmu 2000: 38), or function as the offglide at the end of a syllable. Based on Duanmu's (2000: 42) analysis, there are four falling diphthongs (VG) in Mandarin Chinese, namely, the mid ones [əi, əu(ou)] and the low ones [ai, au(ɑu)]. For instance, [tʰai]/[tʰaj] and [mai]/[maj] contain a falling diphthong with the high vowel [i] as the offglide.

Each of the high vowels [i, y, u] can occupy the first slot in a rising diphthong (glide position) occurring before the nuclear vowel (7). For example, the palatal approximant [j] stands before the nucleus [a] ([ja]), and the labial velar approximant [w] appears before the vowel [ə] ([wəi]).

(7) Consonant + Glide [j/ɥ/w] + Vowels:

	Pinyin	IPA	Chinese character	English gloss
a.	tian1	t + j + a + i	天	'day'
b.	jüe2	tɕ + ɥ + e	觉	'think'
c.	tuo1	t + w + o	拖	'drag'

2.1.5 Syllable structure

Mandarin syllable structure is (C)(G)V(X), which is consonant + glide + vowel + the ending segment. Only a nucleus vowel is required to form a Mandarin syllable, other segments are optional. Five phonemic vowels are included in Mandarin, among which [i, u, y] are high vowels, [ə, a] are non-high vowels without a backness value. Furthermore, there are 21 onset consonants in Mandarin, including 4 labials [p, pʰ, m, f], 11 coronals [t, tʰ, n, l, ts, tsʰ, s, tʂ, tʂʰ, ʂ, ʐ], 3 palatals [tɕ, tɕʰ, ɕ] and 3 velars [k, kʰ, x]. There are three prenuclear glides [j, w, ɥ] and 4 ending segments [i, u, n, ŋ]. The inventory of Mandarin onset consonants, glides, vowels and ending segments are listed in (8).

(8) Mandarin Inventory (see also Duanmu 2000, Gong & Zhang 2019):

21 Onset Consonants: [p, pʰ, m, f, t, tʰ, n, l, ts, tsʰ, s, tʂ, tʂʰ, ʂ, ʐ, tɕ, tɕʰ, ɕ, k, kʰ, x]

3 Glides: [j, w, ɥ]

5 Phonemic Vowels: [i, u, y, ə, a]

4 Ending Segments: [i, u, n, ŋ]

According to Gong & Zhang (2019)'s study, even though Mandarin syllable structure is relatively simple, native speakers' non-word acceptability judgment is still gradient; the acceptability ratings of allophonic gaps and tonal gaps are much higher than systematic gaps. In the present experimental study, only systematic gaps are included, which excludes the tonal gaps and allophonic gaps.

With the segments displayed in the inventory in (8), the number of all possible syllables in Mandarin, along with no realization of the optional components, is (21+1)*(3+1)*5*(4+1) = 2200. Since the syllable structure is (C)(G)V(X), the possible combinations include V, GV, VX, CV, GVX, CGV, CVX and CGVX, which are listed respectively in (9) below. The possible syllable forms include: (9a) 5 syllables consist of a single vowel (V): [a, ə, u, y, i]; (9b) 9 syllables consist of a glide followed by a vowel (GV): [ja, jə, ju, wa, wə, wi, ɥa, ɥə, ɥi]. Since there is no perceptual differentiation (see Gong & Zhang, 2019: 5) between [jy, wy, ɥy] and [y], [ji] and [i], [wu] and [u], these 5 diphthongs are not included. Furthermore, [ɥu] and [ju] are indistinguishable (see Gong & Zhang, 2019: 5), so we include [iu] only. Therefore, there are overall 3*5 – 6 = 9 possible GV syllables. (9c) 10 syllables consist of a nucleus vowel followed by an ending segment (VX): [ai, əi, au, əu, an, ən, aŋ, əŋ, iŋ, yŋ]. Since I consider glides to be positional variants of vowels (see (5)), a sequence such as [ji] is identical to [ij] as underlyingly they are both /ii/. We eliminate the following 10 VX sequences: [ii, yi, ui, iu, yu, uu] where there is a high vowel [i, y, u] followed by [i, u], because they are repetitions of (or not distinguishable from) GV forms [ji, ɥi, wi, ju, ɥu, wu]; as well as [in, un, yn, uŋ] which are not distinct from the GVX forms [iən, uən, yən, uəŋ]. Hence, the total number of possible VX syllables are 5*4 – 10 =10. The syllables in (9d) consist of a glide followed by a nucleus vowel and an ending segment (GVX): [iai, iəi, iui, uai, uəi, yai, yəi, iau, iəu, uau, uəu, uiu, yau, yəu, yiu, ian, iən, iun, uan, uən, uin, yan, yən, yin, iaŋ, iəŋ, iuŋ, uaŋ, uəŋ, uiŋ, yaŋ, yəŋ, yiŋ]. If the 9 acceptable GV sequences are followed by each of the 4 codas, we would expect 36 syllables, but we only find 33 (9*4 – 3 = 33). The 3 syllables that are eliminated include [uii, yii, iuu] (since the VX sequence is banned, and they are not distinguishable from GV sequences [ui, yi, iu). Last but not least, all the syllable forms in (9a) – (9d) can be preceded by a word-initial consonant, as shown in (9e). That is to say, the number of C(G)V(X) syllables is 21*(5 + 9 + 10 + 33) = 1197.

(9) Possible Syllable Combinations

(a) V: 5

[a, ə, u, y, i]

(b) GV: 3*5 – 6 = 9

[ia, iə, iu, ua, uə, ui, ya, yə, yi]

not included:

[iy, uy, yy]   [y]

[ii, uu]   [i, u]

[yu]   [iu]

(c) VX: 5*4 – 10 =10

[ai, əi, au, əu, an, ən, aŋ, əŋ, iŋ, yŋ]

not included:

[ii, uu]   [i, u]

[yu]   [iu]

[ii, yi, ui, iu, yu, uu] (GV repetition)

[in, un, yn, uŋ]   [iən, uən, yən, uəŋ] (GVX repetition)

(d) GVX: 9*4 – 3 = 33

[iai, iəi, iui, uai, uəi, yai, yəi,

iau, iəu, uau, uəu, uiu, yau, yəu, yiu,

ian, iun, uan, uin, yan, yin, iən, uən, yən,

iaŋ, iəŋ, iuŋ, uaŋ, uiŋ, yaŋ, yəŋ, yiŋ, uəŋ]

not included:

[uii, yii, iuu]   [ui, yi, iu] (GV repetition)

(e) C(G)V(X): 21*(5 + 9 + 10 + 33) = 1197

C-V

C-GV

C-VX

C-GVX

In summary, without repetitions and perceptual illusion, there are 5 + 9 + 10 + 33 + 1197 = 1254 possible syllables in Mandarin. Adding one extra rhotic syllable [ər] and interjection words [o "喔", io "唷"] in Mandarin, overall, there are 1257 syllables, among which, 402 are attested in written forms according to the online corpus: Mandarin Syllable Frequency Counts for Chinese Characters (Tsai, 2000, http://technology.chtsai.org/syllable/).

2.2 Previous studies on Mandarin phonotactic constraints

In this section, the phonotactic well-formedness constraints previously reported for Mandarin syllables are examined and critically evaluated. Following these studies, in §3, a number of new constraints are proposed to further account for CG and CV sequences.

Algeo (1978: 219) proposes that whether a cluster (a sequence which consists of more than one segment) is attested or whether it is permitted are two different questions: the former is "empirical decidable", while the latter needs a rule of permission. Gong & Zhang (2018) suggests that if unattested forms can be explained by general and natural phonological properties and constraints, they are "systematic gaps" ("not permitted" in Algeo's terms), different from accidental gaps which do not violate any systematic rules (see also Halle, 1962; Coetzee, 2008). Following Algeo (1978)'s and Gong & Zhang (2018)'s arguments, the present study (§4) examines the rules and constraints underlying the Mandarin systematic gaps. The phonological generalizations which account for Mandarin ungrammatical (not permitted) syllables are explored in this paper.

Mandarin phonotactic constraints which penalize the ungrammatical possible CGVX syllabic combinations have been proposed in previous studies (Gong & Zhang, 2019; Lin, 1989; Duanmu, 1990, 2000, 2003, 2008; Myers, 1995). They illustrate two basic principles: the Obligatory Contour Principle (OCP) ruling out adjacent forms with the same place features, and anti-OCP or Agreement constraints favoring adjacent segments that share a feature.

With respect to OCP, which states that "at the melodic level, adjacent identical elements are prohibited" (McCarthy, 1986: 208), Duanmu (2000) put forward to the Articulator Dissimilation principle: "Identical articulators cannot occur in succession" to explain the Mandarin missing consonant + glide (CG) combinations. For instance, the sequence [pw] is penalized in that the labial consonant [p] cannot be followed by a labial glide [w]. Similarly, Lin (1989: 251-271) argues that the OCP plays an important role in Taiwanese and proposes that the initial and final consonants cannot both be labial, for instance, *[pam, pap].

Furthermore, Duanmu (2007: 60) proposes the Rhyme-Harmony constraint to account for Mandarin phonotactics, which is an anti-OCP constraint, as shown in (10).

(10) Rhyme-Harmony: VX cannot have opposite values in [round] or [back]:

*[+back][−back], *[−back][+back]

*[+round][−round], *[−round][+round]

Following that, Duanmu & Yi (2015: 834-836) proposed three new constraints (cited below) for Lanzhou Chinese, which are also respected by Mandarin Chinese syllables. The constraints (11b) and (11c) below are long-distance constraints, and all abide by the Obligatory Contour Principle. Long-distance constraints have been studied in other languages like English as well. For instance, Coetzee (2009) studied English distant constraints *[skVk] and *[spVp], and he found that when subjects were asked to select one from a pair of ungrammatical forms, they prefer *[skVk] over *[spVp]. ([stVt] is fully grammatical.)

(11) (a) No [+high][+high]: No adjacent [+high] sounds are allowed.

(b) No [i]_[i]: [i] cannot occur in both glide and coda position.

(c) No [u]_[u]: [u] cannot occur in both glide and coda position.

Adapted from Duanmu & Yi (2015: 834-836)'s constraints in (11), Gong & Zhang (2018, 2019) additionally proposes three Mandarin (G)V(X) constraints, displayed in (12).

(12) a. *HH: The feature [+high] cannot occur in sequence.

b. *[Cor]_[Cor]: [Cor] cannot occur in both G and X.

c. *[Lab]_[Lab]: [Lab] cannot occur in both G and X.

Gong & Zhang (2018, 2019) further argues that because the front labial vowel/glide [y, ɥ] is marked and rare, it is missing in many syllables. The vowel [y] cannot be preceded by any non-palatal obstruent, for instance, *[py, ky, ty, sy]. Besides, the glide [ɥ] can only be followed by the allophonic vowel [e], which can be exemplified by *[ɥi, ɥu, ɥo, ɥa].

The present study adopts the constraints listed above and further proposes other constraints like CG and CV backness agreement (anti-OCP), which are discussed in detail in §3.

2.3 Nonword acceptability ratings

In this section, studies involved with acceptability rating experiments in different languages like English, Cantonese and Mandarin are reviewed. Acceptability ratings in cognitive experiments carried out by Armstrong, Gleitman & Gleitman (1983: 263-308) show that not only prototypic categories (like 'fruit', 'vehicle'), but also well-defined categories (like 'odd number') can yield graded responses. In the results of their experiment, well-defined categories (non-prototypical) were verified with longer time, while prototypical categories took shorter verification time. Similar experiments have been applied to grammatical judgements by Coetzee (2009) (See also Coetzee, 2008; Hayes & White, 2013). Coetzee (2009) found that participants in his word-likeness rating tasks assign equally high ratings to grammatical forms and equally low ratings to ungrammatical forms. However, when subjects were asked to select one from a pair of ungrammatical forms, they prefer *[skVk] over *[spVp]. He then maintains that both categorical grammar and gradient grammar are used due to the "inherent comparative character of an OT grammar". In Chomsky & Halle (1965)'s study, they argue that the acceptability decreases among the English nonwords "blick, bwick, bnick". For the gradience of acceptability of nonwords like "blick, bwick, bnick", Gong & Zhang (2019: 1) maintains that forms which violate phonotactic principles, for instance, Obligatory Contour Principle (McCarthy, 1986) and the Sonority Sequencing principle (Berent, Steriade, Lennertz, & Vaknin, 2007), are less grammatical than others (see also Myers, 2017; Sprouse, 2018; Chomsky & Halle, 1965).

Nonword acceptability judgment experiments have also been carried out for other languages like Cantonese and Mandarin. Based on the results of Kirby & Yu (2007)'s experiment, Cantonese nonwords which violate CV coronal co-occurrence rule are viewed as less acceptable than nonwords with labial co-occurrence in onset and coda positions. In terms of studies on Mandarin nonword acceptability, in Myers & Tsay (2015)'s syllable acceptability judgment experiment, Mandarin speakers judged whether the written syllables are grammatical or not, during which the reaction time and their judgment are collected. The results show that 13.42% responses to the nonword stimuli are labeled as grammatical (Myers & Tsay, 2015). In the experiment of Myers & Tsay (2005), participants assign acceptability ratings to both lexical and non-lexical CV(C) syllables, the results of which show that accidental gaps are more acceptable than systematic gaps (see also Myers, 2002). Apart from that, Gong & Zhang (2019, 2021) carried out a syllable acceptability judgment experiment to explore the nature of Mandarin speakers' phonotactic knowledge. Gong & Zhang (2019, 2021) divided Mandarin syllables into five types, namely, real words, accidental gaps, allophonic gaps, tonal gaps, and systematic gaps; they then carried out a Mandarin nonword judgment experiment and argued that non-word judgment is gradient: systematic gaps received lower acceptability ratings than accidental gaps, allophonic gaps and tonal gaps. Their experimental results indicate that the syllable acceptability ratings are influenced by not only systematic constraints, but also allophonic restrictions and suprasegmental restrictions. They also suggest that the acceptability ratings of accidental gaps are more related to lexical statistics than systematic gaps, in that systematic gaps are so obviously ungrammatical that the speakers do not need to resort to their frequencies. Similarly, in the study of Berent, Shimron & Vaknin (2001), they found that it takes longer for a more well-formed token to be rejected as non-word. Daland et al. (2011) also found that more grammatical words are more predictable from statistics.

2.4 Categorical and gradient grammars

In this section, the previous comparative studies concerning the categorical grammars and gradient grammars are introduced. The six grammars discussed in this paper include (13a-b) categorical grammars: (13a) a standard categorical grammar based on the systematic constraints and principles generalized in §3; (13b) a data-driven categorical grammar generated by the UCLA Hayes and Wilson Phonotactic Learner (HWPL) (Hayes & Wilson 2008); (13c-d) gradient grammars: (13c) a manually constructed gradient grammar which assigns an overall weight of violated constraints for each syllable; (13d) a data-driven gradient grammar including the calculated penalty score of each Mandarin syllable via the UCLA Hayes and Wilson Phonotactic Learner (HWPL) (Hayes & Wilson 2008); (13e-f) cumulative categorical grammars: (13e) a manually constructed cumulative categorical grammar based on the number of violated systematic constraints and principles; and (13f) a data-driven cumulative categorical grammar based on the number of violated constraints via Hayes and Wilson Phonotactic Learner (HWPL).

(13) Models of grammars

(a) standard categorical grammar based on the systematic constraints and principles

(b) data-driven categorical grammar generated by Hayes and Wilson Phonotactic Learner (HWPL)

(c) manually constructed gradient grammar which assigns an overall weight of violated constraints for each syllable

(d) data-driven gradient grammar including the calculated penalty score (HWPL) of each Mandarin syllable

(e) manually constructed cumulative categorical grammar based on the number of violated systematic constraints and principles

(f) data-driven cumulative categorical grammar based on the number of violated constraints via Hayes and Wilson Phonotactic Learner (HWPL)

Three grammars (13a), (13c) and (13e) are derived from manual phonotactic constraints and principles. In comparison, the grammars (13b), (13d) and (13f) via Hayes and Wilson Phonotactic Learner (HWPL) are completely data-driven and based on statistics. Although some linguists like Bybee (2001) suggest that grammar is derived from frequency and statistics, many studies (Shademan, 2007; Coetzee, 2008; Gong & Zhang, 2019, 2021) suggest that statistical factors are not enough to account for the nonword judgment ratings. Hayes & Wilson (2008) propose that the segmental features and natural classes are the bases of speakers' phonotactic knowledge. In contrast, Vitevitch & Luce (2004) suggests that the frequency and other statistical factors like the similarity of a nonword to the real words (Baily & Hahn, 2001, Gong & Zhang, 2019) could determine its acceptability. Myers & Tsay (2005) also argues that the higher neighborhood density, the higher the acceptability ratings. Gong & Zhang (2019, 2021) maintains that the speakers' phonotactic knowledge could come from phonological constraints and lexical statistics (frequencies). They then exemplified it with the nonword "lbick" which does not abide by the Sonority Sequencing Principle, and the nonword "bnick" which are not attested in the speaker's lexicon. Similarly, according to studies of White & Chiu (2017) and Daland et al. (2011), the phonotactic constraints and the statistical factors both influence native speakers' grammar judgment.

2.4.1 A data-driven step: Hayes and Wilson Phonotactic Learner (HWPL)

When it comes to the phonotactic knowledge from lexical statistics, a more data-driven approach to study Mandarin syllables is the UCLA Phonotactic Learner (HWPL) which was developed by

Hayes & Wilson (2008). Coetzee (2008) argues that lexical statistics and grammatical constraints both play a role in speakers' phonotactic knowledge. Similarly, Gong & Zhang (2018) argues that speakers' phonotactic knowledge are based on both statistics and grammatical principles. His analysis employed Hayes and Wilson Phonotactic Learner (HWPL) to show that the penalty score of Mandarin nonwords and the reaction time of native speakers' acceptability judgment are correlated: the higher the penalty score, the less time the participants would take, and fewer participants label the relevant syllable as acceptable. The Mandarin data with type frequency count in his analyses are taken from the online project Mandarin Syllable Frequency Counts for Chinese Characters (Tsai 2000, http://technology.chtsai.org/syllable/). In Gong & Zhang (2018)'s analyses, the data of grammaticality judgment tasks are taken from Tsai (2015)'s experiment where participants are asked to label the acceptability of each of the given Mandarin monosyllables using two response options: either grammatical or ungrammatical.

Different from Gong & Zhang (2018)'s analysis via Hayes and Wilson Phonotactic Learner (HWPL), the data of the speakers' acceptability ratings in the present study are gradient with a scale from 1 to 7, instead of binary (categorical). The learning data input (type frequency) of the Hayes and Wilson Phonotactic Learner (HWPL) comes from the online project Mandarin Syllable Frequency Counts for Chinese Characters (Tsai 2000, http://technology.chtsai.org/syllable/). Each of the Mandarin syllables, attested real words or unattested nonwords, receives a penalty score.

2.4.2 Previous comparative studies on grammar models

The previous comparative studies on categorical grammar and gradient grammar have been carried out by linguists like Coetzee (2009) and Gorman (2013) among many others. Coetzee

(2009) argues that both categorical grammar and gradient grammar are used by speakers due to the “inherent comparative character of an OT grammar”. However, based on Gorman’s (2013: 5) phonotactic study, like gradient models, a categorical model can also predict word-likeness judgements accurately, thus well-formedness and gradient models are not significantly correlated after “categorical effects are controlled for”. Durvasula (2020) also argues that type frequency and constraint weights are not more useful in phonotactic learning than categorial grammar, which stands in contrast to Hayes & Wilson’s study (2008). He maintains that a categorical phonotactic model causes gradient acceptability without gradient generalizations. Durvasula (2020) maintains that categorial grammars are easier to understand as well as probe generalizations. To further explore this question, the present study uses Mandarin data to compare categorical grammars and gradient grammars, the findings of which have significant implications on which grammar can best account for Mandarin speakers’ phonotactic judgment.

Chapter 3

Local and distant phonotactic constraints

In this section, the general phonotactic principles and constraints underlying Mandarin syllables are explored, which is one of the outstanding questions on Mandarin syllables in this thesis. The phonotactic constraints of Mandarin syllables are divided into local constraints and distant ones. Local constraints define the phonotactic constraints between segments which are immediately adjacent to each other; while distant constraints display the restrictions between segments which are not adjacent to each other. All of these constraints are driven by three underlying principles: 2 OCP constraints: NoLabLab, NoHiHi, and anti-OCP constraints (Backness Agreement). In §3.1 I discuss the bigram local constraints: §3.1.1 CG constraints, §3.1.2 CV local constraints, §3.1.3 GV constraints, §3.1.4 VX constraints. I then move on to the non-local constraints in §3.2, discussing G_X sequences. There are no distant constraints for C_V and C_X sequences in Mandarin syllables. In a nutshell, this section (i) adopts the existing constraints put forward to by Lin (1989), Duanmu (2000, 2007), Duanmu & Yi (2015) and Gong & Zhang (2019); (ii) further proposes three new CG backness agreement constraints (Yang & Repetti, 2019): Agree[back]:$C_{[-son, +cont]}$j, Agree[back]:$C_{[-son]}$ɥ and Agree[back]:DorG; as well as two new CV backness agreement constraints: Agree[back]:$C_{[-son]}$y and Agree[back]: $C_{[hi]}V_{[hi]}$. Following that, in §3.3, I then illustrated the calculated weight of each systematic constraint using the Maxent Grammar Tool (Hayes & Wilson, 2008).

Overall, there are 1254 (C)(G)V(X) possible combinations, among which 852 syllables are unattested. 786 of the 852 unattested syllables (92.25%) can be accounted for by the three principles, with only 66 (7.7%) unattested syllables categorized as accidental gaps.

3.1 Local constraints

The bigram local constraints include CG constraints (§3.1.1), CV constraints (§3.1.2), GV constraints (§3.1.3), and VX constraints (§3.1.4).

3.1.1 CG constraints

Some languages, like Korean, allow any CG combinations, while other languages, like Kirundi, avoid all CG sequences. Most languages have restrictions on their CG sequences. For example, Italian disallows palatals preceding [j] and [w]; some varieties of American English disallow coronals followed by the palatal glide [j]; Cantonese avoids the sequence [fj] with only a few loanword exceptions. Languages like Mandarin have a relatively complex CG system. Linguists have studied restrictions on CG sequences by focusing on articulator features: Kochetov (2016: 19) uses the articulator features Lab, Cor, Dor, and Lar in Kirundi CG constraints, and Duanmu (2000: 32) studies Mandarin CG combinations using three articulator features: Lab, Cor, Dor. Duanmu (2000: 32) argues that restrictions on Mandarin consonant plus glide combinations can be accounted for by the Articulator Dissimilation principle: “Identical articulators cannot occur in succession.” However, this proposal does not cover all CG data. For instance, Mandarin CG clusters like [fj] are ungrammatical even though the initial consonant [f] and the following glide [j] have different articulators. What constraints penalize [fj] in Mandarin? Another problem with

Duanmu's analysis (2000: 26, 28) is that the palatal consonants [t͡ɕ,t͡ɕʰ, ɕ] are left undiscussed with respect to Palatal + Glide combinations. The palatals carry both Dorsal and Coronal articulator features, which obscures the reasons for the grammatical CG combinations [t͡ɕʰj,t͡ɕj, ɕj,t͡ɕʰɥ, t͡ɕɥ, ɕɥ] and missing ones *[t͡ɕw, t͡ɕʰw, ɕw]. Despite Duanmu (2000)'s Articulator Dissimilation claim, further investigation reveals a much more intricate system.

This section is an extension of Duanmu's investigation into CG sequences in Mandarin. I show that all Mandarin CG data can be explained by two general principles: (14a) No Lab + Lab CG sequences, as predicted by Duanmu's Articulator Dissimilation; (14b-d) Backness agreement constraints account for the other CG clusters. I argue that both OCP (No Lab + Lab) and [back] agreement are needed to explain Mandarin CG clusters grammaticality. The CG constraints are defined in (14). Gong & Zhang (2018) maintains that *[fj] is an unnatural phonotactic constraint since it's the only labial that cannot occur before [j]. However, the present paper argues that it is a natural constraint which abides by backness agreement between the consonant and glide (Agree[back]:$C_{[-son, +cont]}$j).

(14) Mandarin CG constraints

(a)*CG-LabLab

A labial consonant cannot be followed by a labial glide [ɥ, w].

(b) Agree[back]: $C_{[-son, +cont]}$j

The fricatives/affricates and the following glide [j] need to have the same backness value.

(c) Agree[back]: $C_{[-son]}$ɥ

The obstruents and the following glide [ɥ] need to have the same backness value.

(d) Agree[back]: DorG

The dorsal consonants and the following glide [j, u, y] need to have the same backness value.

The markedness constraint *CG-LabLab (14a) originates from Duanmu (2000)'s articulator differentiation rule, which rules out all labials plus [ɥ, w] groups. Similarly, for Cantonese syllables, Yip (1988a) and Steriade (1987) proposed that a front rounded vowel and a labial consonant cannot be adjacent to each other, for example, [*py] (See also Lin, 1989: 273). As displayed in (15), the labial consonants [p, p^h, m, f] have the same labial place feature as the glide [ɥ], which violates OCP. Therefore, the sequences [pɥ, p^hɥ, mɥ, fɥ] are missing. For instance, the words 米 [mi] ("rice") and 旅 [lɥ] ("travel") are allowed, whereas words like *[mɥi] and *[pɥe] are missing in Mandarin. In the case of labials + velar glide [w], the roundness of the velar glide also plays a role requiring differentiation from the first onset consonant. In other words, repetition of the labial feature in onset position is disallowed. In Mandarin, the labial consonants [p, p^h, m, f] cannot be followed by a labial glide [w]: *[pw, p^hw, mw, fw].

(15)

*CG-LabLab		ɥ	w
Lab	p	*	*
	p^h	*	*
	m	*	*
	f	*	*

The next set of constraints are anti-OCP, or agreement, constraints whereby certain subsets of consonants must agree in backness with the following glide (14b-d): [j] and [ɥ] are both [-back], while [w] is [+back]. Since the [back] feature is on the dorsal node, these constraints can only be satisfied by consonants with a dorsal place, namely the palatals and velars. If a consonant fits the description of the consonant in the constraint (fricative/affricate (14b), obstruent (14c)), but it does not have a dorsal node (i.e., it is not a palatal or velar consonant), the agreement constraint is violated. In other words, the CG backness agreement constraints (14b-d) require the C and G to have the same backness value: both are [+ba] or both are [-ba]. CG combinations like [+ba, -ba], [-ba, +ba], [0ba, +ba] or [0ba, -ba] would not satisfy backness agreement. Note that the articulator feature nodes themselves are not explanatory enough.

Taking all the Cj sequences into account, we see that stop/liquid/nasal + j groups are grammatical, except the velar stops + j, which are accounted for below. Meanwhile, a fricative/affricate + j sequence is not allowed, except palatals + j. These exceptions can be accounted for with a single backness agreement constraint. When Mandarin consonants precede the glide [j], the affricates and fricatives are required to agree with the glide /j/ in backness value: [-back]. The anti-OCP constraint Agree[back]-Cj is defined in (14b). Among fricatives and affricates, only the palatals carry the same [-back] value as the glide /j/, and therefore these sequences are grammatical: [tɕʰj,tɕj, ɕj]. However, the labial fricative ([f]) and the non-palatal coronal fricatives and affricates ([ts, tsʰ, s, tʂ, tʂʰ, ʂ, ʐ]) which have no [back] feature, and the dorsal fricative ([x]) which has a [+back] value, cannot form an CG cluster with [j] which has a [-back] value.

(16)

Agree[back]: $C_{[-son, +cont]}$j		j [-ba]
Lab	f	*
Cor	ts	*
	ts^{h}	*
	s	*
	tʂ	*
	tʂ^{h}	*
	ʂ	*
	ʐ	*
Pal (Cor, Dor)	tɕ	1
	tɕ^{h}	1
	ɕ	1
Dor	x	*

(17)

Agree[back]: $C_{[-son]}$ɥ		ɥ [-ba]
Lab [0ba]	p	*
	p^h	*
	f	*
Cor [0ba]	t	*
	t^h	*
	ts	*
	ts^h	*
	s	*
	tʂ	*
	tʂ^h	*
	ʂ	*
	ʐ	*
Pal (Cor, Dor) [-ba]	tɕ	1
	tɕ^h	1
	ɕ	1
Dor [+ba]	k	*
	k^h	*
	x	*

When Mandarin consonants precede the glide [ɥ] in onset position, the obstruents are required to agree with the glide /ɥ/ in backness value: [-back] (14c). This anti-OCP constraint is illustrated in (17). Among the obstruents (stops, affricates, fricatives), only the palatals carry the same [-back] value as the glide /ɥ/. Therefore, the labials obstruents + [ɥ] ([pɥ, pʰɥ, fɥ]), the non-palatal coronal obstruents + [ɥ] ([tɥ, tʰɥ, tsɥ, tsʰɥ, sɥ, tʂɥ, tʂʰɥ, ʂɥ, ʐɥ, tʂɥ, tʂʰɥ, ʂɥ, ʐɥ]) and the dorsals + [ɥ] ([kɥ, kʰɥ, xɥ]) are penalized. (The constraint Agree[back]: DorG (14c) also rules out [kɥ, kʰɥ, xɥ], as discussed next.) This constraint does not apply to sonorant consonants + [ɥ] (mɥ, nɥ, lɥ), although the consonant and the following glide [ɥ] cannot both have the labial feature, ruling out [mɥ], because of the *CG-LabLab constraint (14a). Here again we see that backness agreement between initial obstruents and the glide [ɥ] plays a significant role in the grammaticality of CG sequences.

Another constraint on Mandarin CG sequences requires any dorsal consonant to agree with the following glide's backness value (14d). The anti-OCP constraint Agree[back]: DorG is shown in (18). The dorsal consonants [k, kʰ, x] are [+back] and have the opposite value of the [-back] glides [j ɥ]. Therefore, the sequences [kj, kʰj, xj, kɥ, kʰɥ, xɥ] are missing. Alternatively, the palatal consonants [t͡ɕʰ,t͡ɕ, ɕ] are [-back] and have the opposite value of the [+back] glide [w]. Therefore, the sequences*[t͡ɕʰw,t͡ɕw, ɕw] are missing. The anti-OCP constraint Agree[back]-DorG as displayed in (18) is obeyed by Mandarin CG combinations.

(18)

Agree[back]: DorG		j [-ba]	ɥ [-ba]	w [+ba]
Pal (Cor, Dor) [-ba]	t͡ɕ	1	1	*
	t͡ɕʰ	1	1	*
	ɕ	1	1	*
Dor [+ba]	k	*	*	1
	kʰ	*	*	1
	x	*	*	1

Taking all Mandarin CG sequences into consideration, there are four necessary phonotactic constraints: Agree-DorG, Agree-Cɥ, Agree-Cj and *CG-LabLab. Among the four constraints, *CG-LabLab abides by OCP whereas the other three are against OCP. The backness agreement constraints require specific consonant + glide sequences to have the same backness value: both are [+ba] or [-ba]. In summary, backness agreement between the initial consonants and glides play a significant role in the grammaticality of Mandarin CG sequences.

3.1.2 CV local constraints

There are four CV local constraints in Mandarin syllables, which are displayed in (19). These are similar, but not identical to the CG constraints discussed in §3.1.1.

(19) Mandarin CV constraints

(a) *Lab[y]

The labial consonant cannot be followed by a labial vowel [y].

(b) Agree[back]: $C_{[-son]}y$

The word-initial obstruents and the following vowel [y] need to have the same backness value.

(c) Agree[back]: $C_{[hi]}V_{[hi]}$

(i) A velar consonant ([+hi]) and the following high vowel need to have the same backness value. This means that only velar C + [u] is permitted, since they are both [+ba], but not velar C + [y]/[i].

(ii) A palatal consonant ([+hi]) and the following high vowel need to have the same backness value. This means that only palatal C + [i, y] is permitted, since they share the [-ba] specification.

With respect to CV combinations, a labial consonant cannot be followed by the labial vowel [y]. This is an OCP constraint between the initial consonant and its following vowel. For instance, syllables like [py], [p^hy], [myn] and [fye] do not exist in Mandarin. Note that this constraint is more restricted than the *CG-LabLab constraint discussed above (14a) since it targets [y] only, but not [u]. Gong & Zhang (2018) argues that the front labial vowel [y] is marked and rare in Mandarin syllables, and the constraint (19a) shows that the vowel [y] is more restricted than other labial vowels.

(20)

*CV-Lab[y]		y
Lab	p	*
	p^h	*
	m	*
	f	*

Similar to the Agree[back]-$C_{[-son]}$ɥ constraint (14c), word-initial obstruents and the following vowel [y] need to have the same backness value as well. As shown in (21), all the obstruents including stops, fricatives and affricates need to have the same backness value as the vowel [y]: [-back]. Therefore, only the palatal obstruents [t͡ɕ, t͡ɕʰ, ɕ], but no other obstruents can be combined with the front vowel [y].

Corresponding to the Agree[ba]-$C_{[hi]}V_{[hi]}$ constraint (19c-i), backness agreement is also required of CV sequences that share a [hi] feature specification. The only consonants with a feature specification for height are the dorsals, and they are all [+hi]. So this constraint only refers to dorsal consonants and high vowels. For velar consonant [k, k^h, x] + high vowel sequences, they must share the [+back] value since the velar consonants carry [+back] value (22).

(21)

Agree[ba]: $C_{[-son]}y$		y [-ba]
Lab [0ba]	p	*
	pʰ	*
	f	*
Cor [0ba]	t	*
	tʰ	*
	t͡s	*
	t͡sʰ	*
	s	*
	t͡ʂ	*
	t͡ʂʰ	*
	ʂ	*
	ʐ	*
Pal (Cor, Dor) [-ba]	t͡ɕ	1
	t͡ɕʰ	1
	ɕ	1
Dor [+ba]	k	*
	kʰ	*
	x	*

(22)

Agree[ba]: $C_{[hi]}V_{[hi]}$ (19c-i)		i [-ba]	y [-ba]	u [+ba]
Velar [+ba]	k	*	*	1
	k^h	*	*	1
	x	*	*	1

The CV constraint Agree[ba]-$C_{[hi]}V_{[hi]}$ (19c-ii) requires a palatal consonant [t͡ɕʰ, t͡ɕ, ɕ] to be followed by a high front vowel [i, y] ([-back]) since the palatal consonants carry [-back] value (23). In other words, the palatal consonants cannot precede the high [+back] vowel [u].

(23)

Agree[ba]: $C_{[hi]}V_{[hi]}$ (19c-ii)		i	y	u
Pal [-ba]	t͡ɕ	1	1	*
	t͡ɕʰ	1	1	*
	ɕ	1	1	*

3.1.3 Mandarin GV phonotactic constraints

There are three GV local constraints that apply to Mandarin syllables, which are displayed in (24). Duanmu & Yi (2015: 834-836) proposes the OCP constraint No [+high][+high], which refers to the fact that no adjacent [+high] sounds are allowed in Mandarin syllables. In (24a), the

OCP constraint *HH indicates that the adjacent two vowels/glides cannot be both high, which include GV-*HH (24a) and VX-*HH (28a). In terms of the GV anti-OCP constraints, Duanmu (2007: 52, 60) argues that the glide and nucleus vowel cannot have opposite backness and rounding values (see also Lin, 1997). The rounding agreement follows the allophonic rules in §2.1 so does not need to be accounted for by these constraints. The constraint (24b) is adapted from Duanmu (2000, 2007), but does not involve central vowels [a, ə] since their backness values are left undefined; therefore, clusters like [wa, wə] are not restricted and are grammatical, while all other vowels must have the same backness as the preceding [w], namely [+back]. The constraint in (24c) requires agreement on the backness value in the GV cluster [ɥV] so that the V must be [-ba]. In other words, (24b) and (24c) require the relevant GV sequence to have the same backness value: both are [+ba] or both are [-ba]. GV combinations like [+ba, -ba], [-ba, +ba], [0ba, +ba] or [0ba, -ba] would not satisfy backness agreement.

(24) Mandarin GV constraints

(a) *HH

The glide and its following vowel cannot be both high.

(b) Agree[back]: [w]V

The glide [w] must have the same backness values with its following non-central vowel.

(c) Agree[back]: [ɥ]V

The front glide [ɥ] must have the same backness values with its following vowel.

One of the GV constraints is *HH (24a), which means the glide [j, ɥ, w] and its following vowel cannot be both high. Combinations like *[ji, jy, ju, ɥi, ɥy, ɥu, wi, wy, wu] are forbidden, as illustrated in (25). For instance, words like *[twy], *[kɥu] and *[pju] do not exist.

(25)

GV-*HH	i	y	u
j	*	*	*
ɥ	*	*	*
w	*	*	*

The anti-OCP constraint Agree[back]: [w]V (24b) refers to the phenomenon that the labial glide [w] must have the same backness values with its following non-central vowel. In (26), the back glide [w] cannot be followed by the front vowel [i, y, e]. For instance, words like *[twi], *[pwy] and *[lwe] are not grammatical. The shaded cells indicate that the relevant sequences do not violate this constraint but might violate other constraints.

(26)

Agree[ba]: [w]V	i [-ba]	y [-ba]	e [-ba]	ɑ [+ba]	u [+ba]	o [+ba]	ə [central]	a [central]
w [+ba]	*	*	*	1	1	1	*	*

The last GV constraint Agree[back]: [ɥ]V (24c) indicates that the front glide [ɥ] must have the same backness values with its following vowel. It penalizes sequences where the front glide [ɥ] is followed by non-front vowels [a, ɑ, o, u] or schwa. As displayed in (27), the combinations *[ɥɑ], *[ɥo], *[ɥa] and *[ɥə] are not allowed by this constraint. The shaded cells indicate that the relevant sequences do not violate this constraint but might violate other constraints.

(27)

Agree[ba]:[ɥ]V	i [-ba]	y [-ba]	e [-ba]	ɑ [+ba]	u [+ba]	o [+ba]	ə [central]	a [central]
ɥ [-ba]	1	1	1	*	*	*	*	*

3.1.4 Mandarin VX phonotactic constraints

According to Duanmu (2007: 52, 60) and Lin (1997)'s studies, the vowel and its following coda cannot have opposite backness and rounding values. While I adopt the former, I do not adopt the latter since the rounding agreement follows the allophonic rules identified in §2.1. Based on Duanmu (2007)'s and Lin (1997)'s studies, I posit two VX constraints for Mandarin syllables: *HH and Agree[back]: VX.

(28) Mandarin VX constraint

(a) *HH:

The vowel and its following offglide/vowel cannot both be high.

(b) Agree[back]: $\mathbf{VX}_{[DOR]}$

The vowel must have the same backness values as a syllable-final dorsal segment.

Rhymes must abide by the OCP constraint *HH, which means that the vowel and its following offglide cannot be both high. In (29), the [+hi][+hi] sequences like [ii, iu, ɥi, ɥu, ui, uu] are disallowed. For instance, syllables like *[tiu], *[xui] and *[nɥi] don't exist in Mandarin.

(29)

VX-*HH	i	u
i	*	*
ɥ	*	*
u	*	*

The constraint (28b) requiring backness agreement is not involved with central vowels [a, ə], since they are neither back nor front and are not relevant to this constraint; therefore, clusters like [a, ə] + X are not restricted and all are grammatical. Likewise, the word-final consonant [n] is not Dorsal, so it is not relevant to this constraint. In a VX cluster, the non-central vowel and its following word-final dorsal segment are required to have the same backness value: both are [+ba] or both are [-ba]. VX combinations like [+ba, -ba], [-ba, +ba] would not satisfy backness agreement. Therefore, the front vowels [i, y, e] cannot be followed by [+ba] segments [u, ŋ]. In other words, sequences like *[iu, yu, eu, iŋ, yŋ, eŋ] are penalized. Likewise, the back vowels [o, u, ɑ] cannot precede the front word-final segment [i]. Therefore, VX combinations like [oi, ui, ɑi] are ungrammatical.

(30)

Agree[back]: VX	i [-ba]	u [+ba]	ŋ [+ba]
i [-ba]	1	*	*
y [-ba]	1	*	*
e [-ba]	1	*	*
o [+ba]	*	1	1
ɑ [+ba]	*	1	1
u [+ba]	*	1	1

3.2 Distant G_X Phonotactic Constraints

There are two distant G_X constraints on Mandarin syllables, as defined in (31). Based on Lin (1989: 251-271)'s analyses, the pre-nuclear glide and the post-nuclear segment cannot both be rounded, for instance, [*uau]. Similarly, Duanmu & Yi (2015: 834-836) proposes two long-distance constraints: No [i]_[i]: [i] cannot occur in both glide and coda position; and No [u]_[u]: [u] cannot occur in both glide and coda position. Following them, Gong & Zhang (2018) defines the two constraints as: *[Cor]_[Cor]: [Cor] cannot occur in both G and X; and

*[Lab]_[Lab]: [Lab] cannot occur in both G and X. The two G_X constraints I propose are adapted from Duanmu & Yi (2015: 834-836) and Gong & Zhang (2018).

(31) Distant constraints

(a) *Lab_Lab (G_X)

The pre-nuclear glide and the word-final X segment cannot both be labials.

(b) *Hi$_{[-ba]}$_Hi$_{[-ba]}$ (G_X)

The glide and the word-final X segment cannot both be high and front (i. e., palatal).

The non-local G_X constraint *Lab_Lab (31a) refers to the phenomenon that the glide and the post-vocalic segment with a vowel in between cannot both be labials, as illustrated in (32). Therefore, sequences like [ɥ_u, w_u] are penalized.

(32)

*Lab_Lab (G_X)	u
ɥ	*
w	*

(33)

*Hi$_{[-ba]}$_Hi$_{[-ba]}$ (G_X)	i
j	*
ɥ	*

Another OCP constraint on non-local G_X, $*\mathrm{Hi}_{[\text{-ba}]}\_\mathrm{Hi}_{[\text{-ba}]}$ (31b) refers to the phenomenon that the glide and the post-vocalic segment cannot both be palatal [i, y]. Hence, the front glides [i, y] cannot appear in the same syllable as the word-final high front vowel [i] in X position, ruling out words like *[tɥoi] and *[nɥei].

3.3 Weight of constraints

In previous studies like Gong & Zhang (2018) and Durvasula (2020), Maxent Grammar Tool and Hayes and Wilson Phonotactic Learner (HWPL) are regarded as a gradient grammar to measure how good a word is. The weight of each Mandarin systematic constraint is calculated via Maxent Grammar Tool (Hayes & Wilson, 2008) and an overall summed weight of violated constraints is assigned to each syllable as the gradient grammar. The data is taken from Mandarin Syllable Frequency Counts for Chinese Characters (Tsai, 2000, http://technology.chtsai.org/syllable/) to calculate the weight of constraints. In this corpus, 13,060 Chinese characters are included, among which there are 2 characters with four pronunciations, 47 characters with three pronunciations, 491 characters with two pronunciations, 12,484 characters with a single pronunciation, as well as 36 characters with no pronunciation information (2*4 + 47*3 + 491*2 + 12484 - 36 = 13615 character-pinyin types). Therefore, overall, 13615 types are included and calculated for type frequencies for Mandarin syllables. Hayes and Wilson (2007: 19) maintains that token frequencies, compared to type frequencies, appear to yield “slightly less accurate results” in modeling phonological intuitions (see also Bybee 1995, 2001, Pierrehumbert 2001a, Albright 2002a, Albright and Hayes 2003, Hayes and Londe 2006, and Goldwater 2007). In the current analyses on the weights of Mandarin constraints, type frequencies, instead of token frequencies are adopted. By means of Hayes & Wilson’s Maxent Grammar Tool (Hayes & Wilson, 2008), the constraints are weighted and displayed in (34) below.

(34) Weights of Mandarin systematic constraints

Constraint	weight(Maxent)
*Lab_Lab (G_X)	6.27
*CG: LabLab	6.38
*Lab[y]	3.45
*$Hi_{[-ba]}$_$Hi_{[-ba]}$ (G_X)	5.94
*HH (GV)	2.46
Agree[back]: $C_{[-son]}$ɥ	6.09
Agree[back]: $C_{[-son]}$y	5.54
Agree[back]: $C_{[-son, +cont]}$j	6.93
Agree[back]: DorG	6
Agree[back]: $C_{[hi]}V_{[hi]}$	6.75
Agree[back]: [w]V	4.19
Agree[back]: VX	6.03
Agree[back]: [ɥ]V	5.54

3.4 Summary

In summary, this section puts forward five original constraints, which include three CG backness agreement constraints: Agree[back]: $C_{[-son, +cont]}$j, Agree[back]: $C_{[-son]}$ɥ, Agree[back]: DorG; as well as two CV constraints: Agree[back]: $C_{[-son]}$y and Agree[back]: $C_{[hi]}V_{[hi]}$. Meanwhile, adopted from the previous studies of Lin (1989), Duanmu (2000, 2007), Duanmu & Yi (2015) and Gong & Zhang (2019), two GV constraints (Agree[ba]: [w]V, Agree[ba]: [ɥ]V), one VX constraint (Agree[ba]: VX), three *LabLab constraints (*CG-LabLab, *CV-Lab[y], *G_X-Lab_Lab) and two *HH constraints (*GV-HH, *G_X-*$Hi_{[-ba]}$_$Hi_{[-ba]}$) are illustrated in (35) below.

(35)

	Anti-OCP	**OCP**	
	Agree[ba]	***HH**	***LabLab**
CG	Agree[ba]: $C_{[-son, +cont]}$j Agree[ba]: $C_{[-son]}$ɥ Agree[ba]: DorG		*LabLab
GV	Agree[ba]: [w]V Agree[ba]: [ɥ]V	*HH	
CV	Agree[ba]: $C_{[-son]}$y Agree[ba]: $C_{[+hi]}V_{[+hi]}$		*Lab[y]
VX	Agree[ba]: VX		
G_X		*$Hi_{[-ba]}$_$Hi_{[-ba]}$	*Lab_Lab

Both the local and long-distance phonotactic constraints of Mandarin syllables are displayed in (35). The local constraints on CG sequences include *CG: LabLab (OCP) and the anti-OCP backness agreement constraints: Agree[ba]: $C_{[-son, +cont]}$j, Agree[ba]: $C_{[-son]}$ɥ, Agree[ba]: DorG. Similarly, the CV sequences follow the OCP constraint *Lab[y] as well as backness agreement constraints: Agree[ba]: $C_{[-son]}$y and Agree[ba]: $C_{[+hi]}V_{[+hi].}$ Both GV and VX sequences follow the OCP constraint *HH. They also abide by the anti-OCP backness agreement: Agree[back]: [w]V, Agree[ba]: [ɥ]V and Agree[ba]: VX. Lastly, the G_X long-distance OCP constraints include *Lab_Lab and $*Hi_{[-ba]}\_Hi_{[-ba]}$. As illustrated in (35), all the systematic constraints of Mandarin syllables are guided by three principles: backness agreement, *HH and *LabLab, among which the backness agreement constraints are anti-OCP, while *HH and *LabLab abide by the OCP.

Chapter 4

Syllable Acceptability Judgment Experiment

The systematic constraints and phonological generalizations discussed in §3 above play a role in Mandarin phonotactics. We examine how to best model their role by investigating whether a syllable violates any constraints, the number of constraint violations, as well as the summed weight of all violated constraints. In §2, the different types of categorical and gradient grammars were compared, which raises the question: which grammar best models the phonotactic knowledge and well-formedness criteria of Mandarin native speakers? Which one of these three types of grammars (categorical grammars, cumulative categorical grammars, and gradient grammars) best accounts for Mandarin phonotactics?

The present section reports on a nonword acceptability rating experiment carried out to check how the native speakers' phonotactic intuition can be accounted for by six different grammars: two categorical grammars (with constraints determined manually or by the Hayes and Wilson Phonotactic Learner (HWPL)), two cumulative categorical grammars (with the number of violated constraints determined as above), as well as two gradient grammars (with the weight of constraints determined by the MaxEnt Grammar Tool (Hayes & Wilson 2008) or the penalty score (HWPL) generated by the Hayes and Wilson Phonotactic Learner (2008). Allophonic constraints and tonal variations are not included as factors in the current experiment.

The acceptability ratings of all the possible (C)(G)V(X) sequences represent Mandarin grammatical and ungrammatical syllables and are collected from native speakers. I carried out comparative analyses to check the differences among different syllable types: attested syllables and unattested syllables, including attested syllables with constraint violation and attested syllables without constraint violation, accidental gaps, and systematic gaps, as well as different types of systematic gaps. The data collected in the present experiment is then used for the comparative analyses of gradient grammars, cumulative categorical grammars and categorical grammars in §5.

In this chapter, an overview of the Mandarin syllable acceptability judgment task is introduced in §4.1 below. Following that, §4.2 describes the experimental methods. In §4.3, I carry out correlation tests among four factors, i.e., number of violated phonological constraints, summed weight of phonological constraints, number of violated constraints in Hayes and Wilson Phonotactic Learner (HWPL), and penalty score (HWPL). The results indicate that they are highly correlated with each other. Subsequently, §4.4 further analyzes the correlations between syllable types and acceptability ratings from Mandarin speakers by checking whether syllable types can significantly influence Mandarin speakers' acceptability ratings on syllables. In §4.5, the interactions between syllable type and other factors are statistically modeled using a mixed effects linear regression model. Lastly, §4.6 summarizes the data and initial analyses carried out in this chapter.

4.1 Overview of the experiment

Mandarin attested syllables and nonwords are both included in the syllable acceptability judgment task for native speakers. The ratings on the made-up nonwords which speakers have not encountered before were solicited in order to tap into the underlying phonotactic grammar acquired by Mandarin speakers. This approach for conducting syllable acceptability judgment experiments was also adopted in other studies, including Gong & Zhang (2018, 2019, 2021), Myers (2017) and Sprouse (2018), among many others.

However, the previous studies didn't further divide the Mandarin systematic gaps and does not differentiate the grammaticality within systematic gaps. Whether the degree of acceptability of different systematic gaps is gradient remains unknown. For example, the syllable [ty] is a systematic gap which violates one constraint and the syllable [gyɑu] is a systematic gap violating 4 constraints. Whether the two words which are systematic gaps in the lexicon – [ty] and [gyɑu] – are viewed with different acceptability ratings was not answered in Gong & Zhang (2018, 2019)'s studies. To further explore this question, the present experiment divides the systematic gaps based on their number of constraint violations, the weight of constraints and the penalty score (HWPL) of syllables. In this experiment, tonal gaps and allophonic gaps are not included in the stimuli. The present study is intended to examine the acceptability ratings of different systematic gaps as well as accidental gaps by linguistically native speakers of Mandarin Chinese.

The methodology of the current experiment follows Gong & Zhang (2019, 2021: 250-251). Nonword acceptability judgments are used to test whether the speakers' ratings and the grammaticality of nonwords are positively correlated (see §5). If the data show that, among the systematic gaps, the higher the number of constraint violations, the higher the weight or penalty of violated constraint, the less acceptable the nonword is, then the cumulative categorical grammar and the gradient grammar are sufficient to account for Mandarin phonotactics. We need to differentiate not only between accidental gaps and systematic gaps, but also to make further distinctions among the different kinds of systematic gaps. If, instead, the data show that the acceptability ratings do not change with the number of violated constraints, the overall weight of violated constraints and penalty scores (HWPL), then the cumulative categorical grammar and gradient grammar are not more predictive of speakers' acceptability judgment than categorical grammar. If the categorical and gradient grammar account for the data equally well, then one has no advantage over the other on empirical grounds.

4.2 Methods

4.2.1 Participants

There were 51 Participants who took part in this experiment (N= 51, 26 males, 25 females, age range: 21-54 (mean = 29.68)). All the participants were native speakers of Mandarin Chinese and were recruited online and volunteered to participate in this experiment. On average, the task lasted around 20 – 25 minutes.

4.2.2 Stimuli

The stimuli in this experiment were 1257 (C)(G)V(X) combinations in Mandarin listed in §2.1.5. The stimuli were constructed following the method adopted by Gong & Zhang (2019, 2021) and Myers & Tsay (2015), which tested possible (C)(G)V(X) Mandarin syllables. The stimuli of the present experiment are also checked and ranked with respect to their number of constraint violations and summed weight. Since the present study focuses on the distinctions among systematic gaps, allophonic distinctions and tonal distinctions were not included. Among the 1257 syllables, 402 are attested in written forms according to the online corpus (Mandarin Syllable Frequency Counts for Chinese Characters). Among the 402 attested syllables, 2 syllables [jai] and [lyen] are not found in the online Xinhua Mandarin dictionary, which includes over 30,000 Chinese characters. The syllable [jai2] (崖, 啀, 娾, 睚) is pronounced as [ja2] and [ai2] according to Xinhua Mandarin dictionary. The syllable [lyen2] (孿, 攣, 孌) is pronounced as [luan] according to Xinhua Mandarin dictionary, as displayed in (36) below. Due to possible participant confusion caused by geographical variations in pronunciation of the aforementioned tokens, they were excluded from data analyses. Therefore, although the 7 character-pinyin types relevant to [yai] and [lyen] in (36) are in the corpus (Mandarin Syllable Frequency Counts for Chinese Characters), they are not included in the present analyses. Excluding the two (C)(G)V(X) syllables [jai] and [lyen], there are overall 1257 - 2 = 1255 monosyllabic stimuli types in this experiment.

(36) Variations of the two syllables [jai, lyen]

[jai] 4 types

崖 1 ㄧㄞˊ jai2 3 ->ja

啀 1 ㄧㄞˊ jai2 1 -->ai

娾 1 ㄧㄞˊ jai2 1 -->ai

睚 1 ㄧㄞˊ jai2 2 -->ja

[lyen] 3 types

孿 1 ㄌㄩㄢˊ lyen2 3 -->luan

攣 1 ㄌㄩㄢˊ lyen2 2 -->luan

孌 2 ㄌㄩㄢˇ lyen3 2 ㄌㄨㄢˇ luan3 1 -->luan

As displayed in Table (37) below, each participant was presented with 73 syllables from eight stimuli sets A-H (10 + 10 + 3 + 10 + 10 + 10 + 10+ 10), as well as 8 tokens in the practice trial (one from each of the 8 stimuli sets). Overall, each participant needed to provide an acceptability rating for 81 monosyllabic tokens. In the first column of Table (37), the acceptability judgment task tokens are categorized as “attested”, “accidental gaps”, or “systematic gaps”, based on whether they violate any of the manually constructed constraints in discussed §3. Most attested syllables and accidental gaps are grammatical with respect to categorical grammar since they do not violate any constraint. However, systematic gaps and seven attested syllables that violate one constraint are ungrammatical with respect to categorical grammar.

(37) Experiment Stimuli set A-H

type of syllables	No. of violated constraints	weight range	experiment stimuli set	No. of syllables	Practice trial	Random sample	No. of rated syllable types
attested syllables	0	0	A	396	1	10	270
accidental gaps		0	B	66	1	10	66
attested syllables	1	2.46	C	4	1	3	4
systematic gaps		2.46-6.93	D	337	1	10	270
systematic gaps	2	6.65-13.02	E	218	1	10	196
	3	12.68-18.87	F	146	1	10	141
	4	18.95-24.28	G	74	1	10	74
	5	25.33-27.36	H	14	1	10	14
TOTAL				1255	8	73	
					81		

In column 3 of Table (37), the weight range for each group is calculated based on Hayes & Wilson's Maxent Grammar Tool (2008). The input Mandarin data are taken from the online project Mandarin Syllable Frequency Counts for Chinese Characters (Tsai 2000, http://technology.chtsai.org/syllable/). Each token has an overall weight value which refers to the sum of its violated constraints' weights. The syllables which violate 0 constraints receive an overall weight of 0. The syllables violating one or more constraints receive a weight which is the sum of its violated constraints. For instance, the syllable 'dui' violates two constraints (*HH and Agree[back]: [w]V) whose weights are 2.46 and 4.19. Therefore, the overall weight of the nonword 'dui' is 2.46 + 4.19 = 6.65, which falls within the weight range of the stimuli set E (stimuli sets are identified in column 4) in Table (37). This represents how the gradient score of each syllable was calculated in accordance with the assumptions of the gradient grammar. If the gradient grammar can account for speakers' phonotactic knowledge, then the overall weight of the violated constraints should be inversely related to the acceptability of a Mandarin syllable.

In column 2, the gradient weight of each constraint is not taken into consideration, instead the number of violated constraints is counted. The constraints are manually constructed based on the Mandarin data and phonological generalizations, instead of automatically generated by the Hayes and Wilson Phonotactic Learner (HWPL), as illustrated in §3. This is not completely categorical since there is a gradience of the number of violated constraints, which is consistent with a cumulative categorical grammar. If a cumulative categorical grammar influences speakers' acceptability judgments, the prediction is that the greater the number of violated constraints, the lower the acceptability of the nonword.

In column 5, the overall number of stimuli tokens in each group (A-H) is displayed. From every group, one syllable type is used in the Practice trial (1*8 tokens), as shown in column 6. The 7th column indicates that 10 stimuli types were randomly chosen from each group, except Group C with less than 10 stimuli types, due to the fact that only 4 attested syllables with constraint violations exist in Mandarin. The attested syllable tokens include both real syllables without constraint violations and real syllables with constraint violations. Meanwhile, the nonword stimuli include both accidental gaps and systematic gaps. Finally, column 8 indicates the total number of rated syllables from each group.

4.2.3 Procedures

The participants listened to the 81 syllables recorded by a native speaker of Mandarin Chinese. The 81 tokens were presented in a random order. Similar to Gong & Zhang (2019, 2021)'s study, the rating task was self-paced, and the participants had no time limit. How the real words and non-words differ in terms of the number of violated constraints and their cumulative weight range is displayed in Table 1 above. The nonword stimuli were all recorded with the fourth tone, whereas the attested syllables were recorded with the fourth tone if its fourth tone is attested; otherwise, they were recorded with their attested tone. Although most syllables have more than one tone, Duanmu's (2007: 253) study on the tonal frequency of Chinese syllables reports the fourth tone has the highest frequency ((38) is cited from Duanmu 2007: 253).

(38) Frequency of tones

Tone:	First	Second	Third	Fourth	TOTAL
Number of syllables:	337	255	316	347	1,255

The syllables were presented one by one, with the written text of guidance in Mandarin "请听这个音节, 根据你的汉语普通话知识，判断它是否可能为汉语普通话。" (Listen to the following token. Based on your knowledge of Mandarin Chinese, decide if it can be a Mandarin word.) After listening to each recorded token, the participants rated the acceptability of each syllable on a scale from 1 to 7. The end points of the rating scale were defined as follows: 1 (完全不可能 "No, impossible") to 7 (完全可能 "Yes, definitely possible"). To mitigate item order effects, all the items were randomized, and every participant saw a different random order of items. As displayed in Table (39), before being presented with 73 stimuli syllables, each participant was given a practice trial which included 8 syllables presented in a random order: one from each stimuli set (with 0 to 5 violated constraints). The eight syllables included in the practice trial are [pa, lui, iung, ria, pyəŋ, ky, gyau, hyiu]. The acceptability ratings of each participant in the practice trial were collected to check the validity of participants' responses. If the participants' responses for the attested syllables with no constraint violation (from set A) were lower than or the same as all other unattested and ungrammatical stimuli syllables, his or her responses were not included in the data analyses. I carried out the experiment using Qualtrics survey software.

(39) Eight tokens used in the Practice Trial

experiment stimuli set	syllable	No. of Violated constraints	Attested in corpus
A	pa	0	√
B	luəi		*
C	juŋ	1	√
D	ʐia		*
E	ky	2	*
F	pyəŋ	3	*
G	gyɑu	4	*
H	hyiu	5	*

4.2.4 Data Analyses

The mean duration of the stimuli tokens is 545 ms (SD =31). Among the data collected from 51 participants, the acceptability responses of one participant on tokens in the practice trial have shown that the rating for the attested syllable [pa] with no constraint violation is 1. It is lower than or equal to ratings of the other unattested and ungrammatical stimuli syllables [lui, ria, ky, gyau, hyiu]. Therefore, the responses from the participant in question were eliminated from the following data analyses.

Overall, 4012 acceptability ratings of Mandarin syllables collected from 50 participants were included in the analysis in this section. Each participant was presented with 81 syllables, which are labeled as four types: systematic gaps (SG), accidental gaps (AG), attested syllables with no constraint violations (attested), and attested syllables with one constraint violation (attested_V). The examples of the data used for the statistical analyses are illustrated and displayed in Table (40) below. Five factors are included in the correlational analysis with acceptability ratings: (i) syllable type ("type" or word type) , (ii) number of constraint violations generated from Hayes and Wilson Phonotactic Learner ("#violations (HWPL)"), (iii) penalty score of each syllable calculated in Hayes and Wilson Phonotactic Learner ("penalty (HWPL)"), (iv) number of constraint violations based on phonological generalizations illustrated in §2 ("#violations (PG)"), (v) weight of constraint violations based on phonological generalizations illustrated in §3 ("weight (PG)").

With respect to the acceptability judgment, the raw ratings (scale: 1 to 7) from 50 participants were transformed into z scores, as shown in the column "zrating" in Table (40). This step aims to minimize the influence from different rating standards used by each participant (Gong & Zhang, 2019, 2021; Bates, Mächler, Bolker, & Walker, 2014; Cowart, 1997). Following Gong & Zhang (2019, 2021)'s method, the z score transformations are used to address the problem of variations among participants in the present study.

(40) Examples in the dataframe of Mandarin Nonword acceptability rating experiment

participant	token	zrating	wordtype	syllable type	#violations (PG)	Weight (PG)	#violations (HWPL)	Penalty (HWPL)
p49	an	2.015331	real_word	attested	0	0	1	1.033
p43	biu	-0.6848	non word	SG	1	2.46	23	16.822
p33	ɕiuŋ	2.174668	real_word	attested_V	1	2.46	5	1.793
p16	dia	0.524623	non word	AG	0	0	4	5.445

(41) Summary of Mandarin Nonword acceptability rating experimental data

token	#violations (PG)	weight(PG)	#violations (HWPL)	Penalty (HWPL)
Length:4012	Min.: 0.00	Min.: 0.00	Min.: 0.0	Min.: 0.000
	1st Qu.: 0.00	1st Qu.: 0.00	1st Qu.: 5.0	1st Qu.: 4.698
	Median: 2.00	Median: 9.39	Median: 9.0	Median: 12.496
	Mean: 2.09	Mean: 11.11	Mean: 15.7	Mean: 15.219
	3rd Qu.: 4.00	3rd Qu.: 18.95	3rd Qu.: 22.0	3rd Qu.: 22.798
	Max.: 5.00	Max.: 27.36	Max.: 53.0	Max.: 43.013

The summary of Mandarin Nonword acceptability rating experimental data is displayed in Table (41) above. The correlational analyses (see section 4.3 below) were conducted between the z scores of acceptability ratings and syllable type, number of constraint violations in Hayes and Wilson Phonotactic Learner (HWPL) (labeled as “#violations (HWPL)”), penalty in Hayes and Wilson Phonotactic Learner (HWPL), number of constraint violations and weight based on phonological generalizations (labeled as “#violations (PG)”and “weight (PG)”). The value of the factor “#violations (PG)” ranges from 0 to 5 and the value of weight (PG) ranges from 0 to 27.36 with an average of 11. Meanwhile, the value of “#violations (HWPL)” ranges from 0 to 53 (mean = 15.7) and the penalty score (HWPL) are between 0 and 43.013 (mean = 15.2). Meanwhile, the acceptability ratings (z values) of syllables from Mandarin speakers ranges between -1.98 and 2.51 with the mean being close to 0.

4.3 Correlations between different factors

The following correlation tests indicate that the four factors, i.e. number of violated constraints (PG), weight (PG), number of violated constraints in Hayes and Wilson Phonotactic Learner (HWPL) and penalty score (HWPL), are strongly correlated with each other (all cor values > 0.80), as illustrated in Table (42) and Figure 1. The fact that the predictors correlate with each other may lead to the limitations and overlaps of statistical analysis. Therefore, the multivariate linear regression is carried out for each factor respectively in §5.

(42) Correlation tests among factors

correlation test	t	df	p-value	95 percent confidence interval (Cor)
weight (PG) & #violations (PG)	377.75	4010	< 2.2e-16	0.9862384
penalty (HWPL) & #violations (HWPL)	196.97	4010	< 2.2e-16	0.9520113
#violations (PG) & #violations (HWPL)	86.096	4010	< 2.2e-16	0.805566
penalty (HWPL) & weight (PG)	112.75	4010	< 2.2e-16	0.8719049

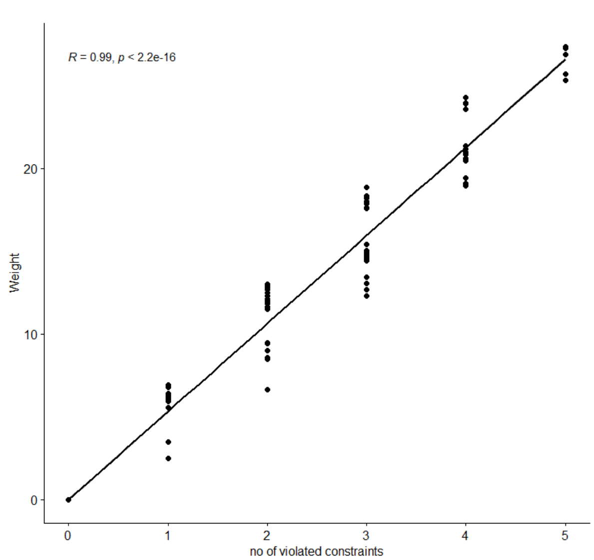


Figure 1a. Correlation between #violations (PG) and weight (PG)

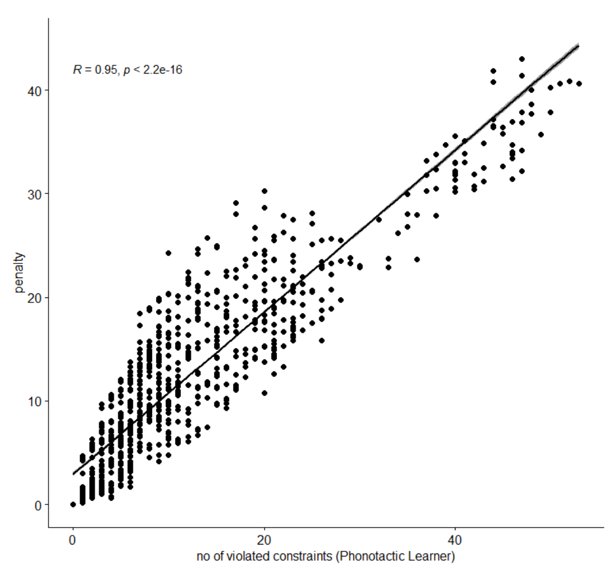


Figure 1b. Correlation between #violations (HWPL) and penalty (HWPL)

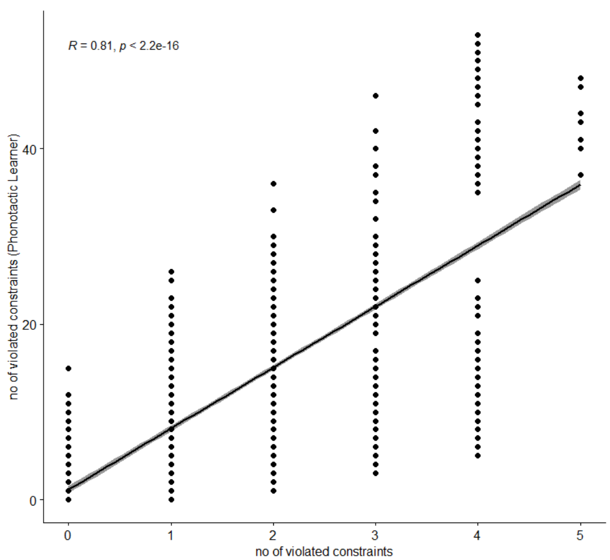


Figure 1c. Correlation between #violations (PG) and #violations (HWPL)

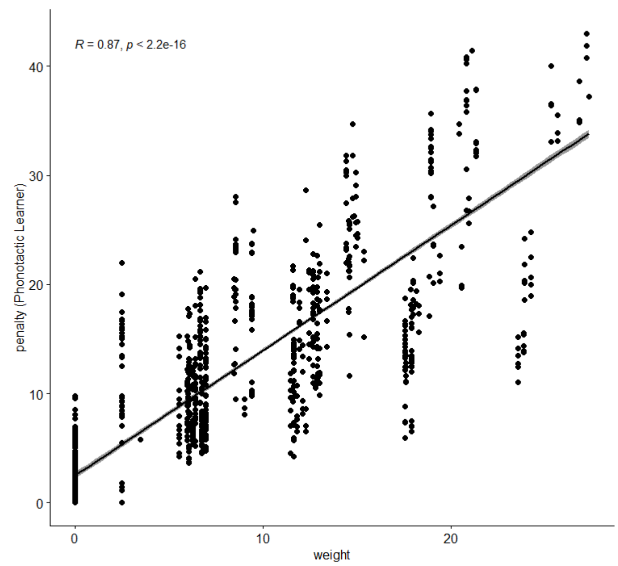


Figure 1d. Correlation between weight (PG) and penalty (HWPL)

4.4 Syllable type and acceptability ratings

Based on Gong & Zhang (2019: 8, 2021: 256)'s study, there exists significant variation in acceptability ratings of different syllable types, as illustrated in Figure 2 below (Gong & Zhang 2019: 8, 2021:256) which is directly cited from Gong & Zhang's paper. The five distinct syllable types in Gong & Zhang (2019, 2021)'s paper include: real word, accidental gap, systematic gap, allophonic gap and tonal gap. The result of his study shows that real words are significantly more acceptable than nonwords and the systematic gaps receive the lowest acceptability ratings. Meanwhile, tonal gaps and allophonic gaps are more acceptable than accidental gaps. This shows the gradiences of acceptability among different syllable types. Mandarin speakers are sensitive not only to systematic phonemic constraints, but also allophonic and tonal principles (Gong & Zhang, 2019, 2021).

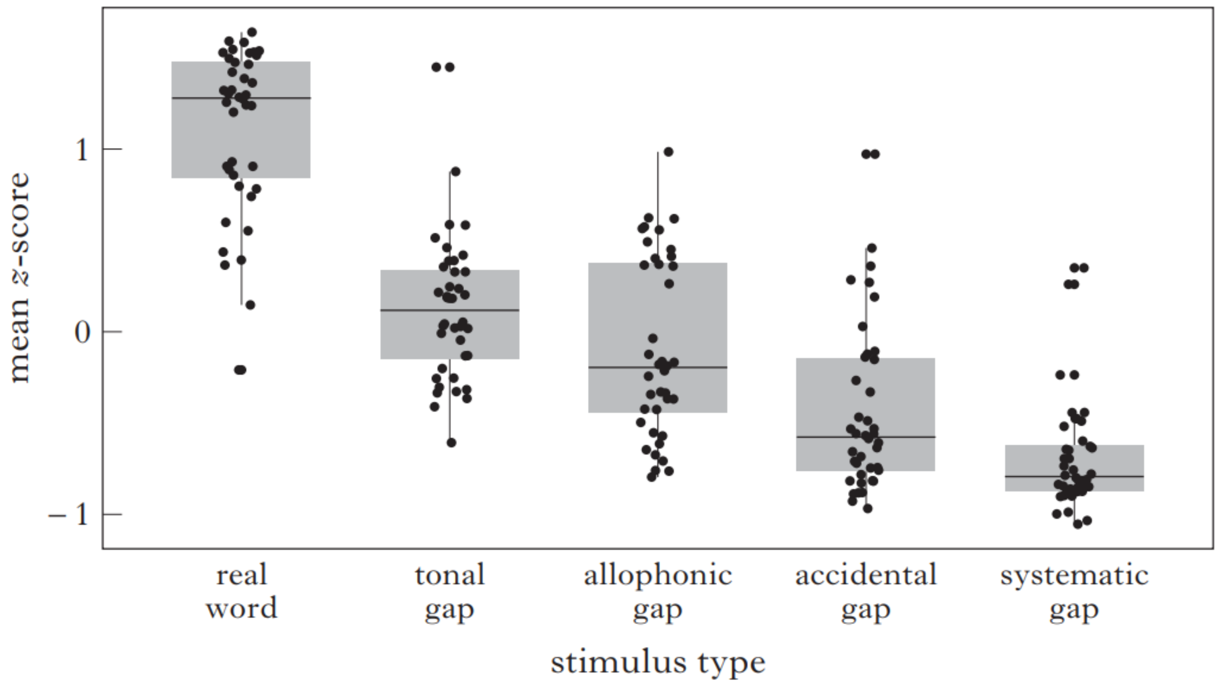


Figure 2 Mean z-scores of well-formedness ratings by Stimulus types

(cited from Gong & Zhang2021:256, see also Gong & Zhang 2019: 8)

In the current study, syllable types investigated include attested syllables with 0 constraint violations, attested syllables with one constraint violation, unattested syllables with 0 constraint violations (accidental gaps) and unattested syllables with one or more constraint violations (systematic gaps). This is different from Gong & Zhang (2019, 2021)'s study in that real words are further divided into the ones without constraint violations and the ones with one constraint violation (*HH (GV)). Besides, allophonic gaps and tonal gaps are not the focus of this present paper, so the stimuli tokens do not include them.

As shown in Figure 3 below, the acceptability ratings of real words and nonwords collected from Mandarin speakers in the present experiment are illustrated. This replicates Gong & Zhang (2019, 2021)'s experimental results: real words (attested syllables with or without constraint violation) are significantly more acceptable than unattested syllables (accidental gaps or systematic gaps).

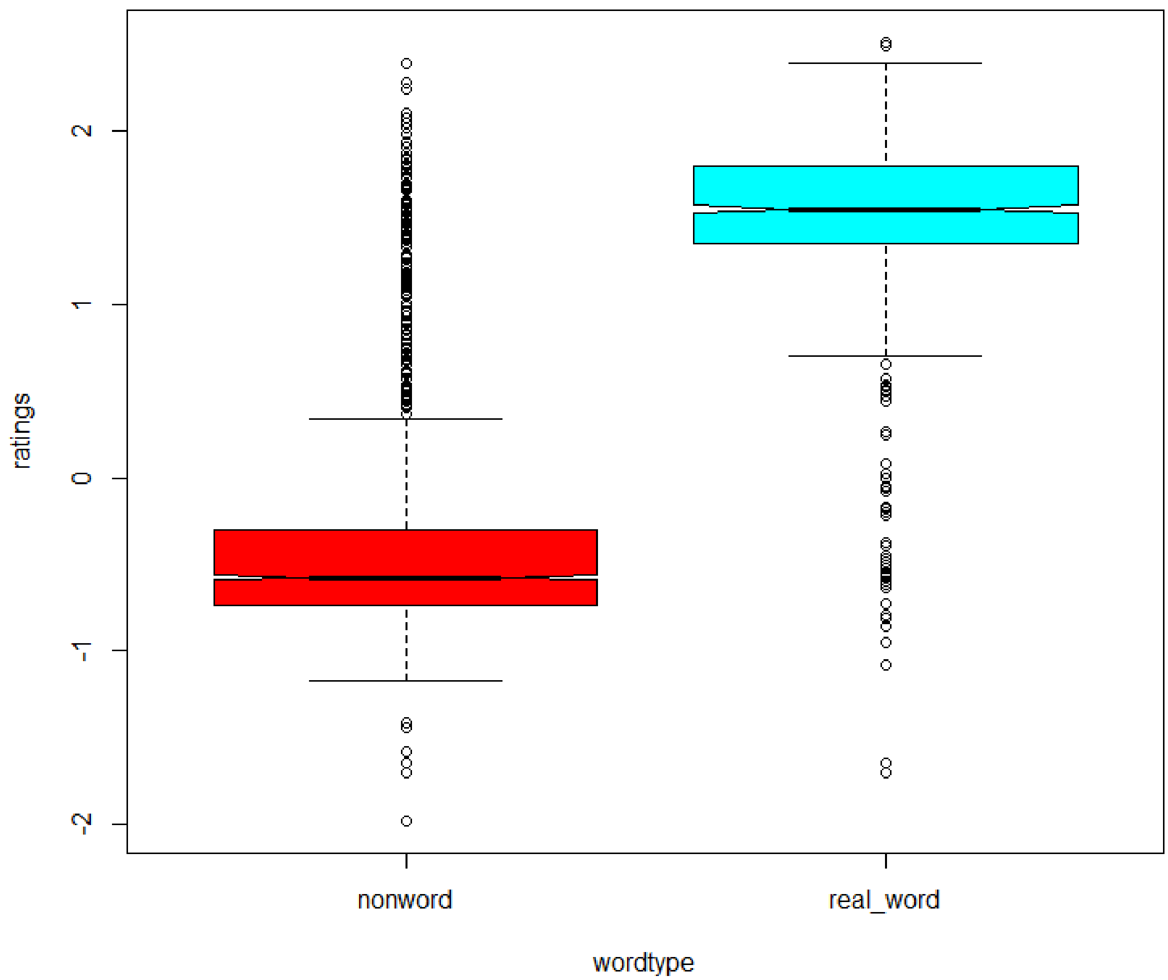


Figure 3. Effect of word types on acceptability ratings

The correlations between the 4-level syllable types and acceptability ratings are illustrated in Figure 4. In the current acceptability judgment study, the acceptability ratings of attested syllables are significantly higher than unattested ones. Furthermore, the accidental gaps tend to be rated higher than systematic gaps, but lower than attested syllables. However, distinctions between real words with 1 violation and those without constraint violation are not significant. The rating scale used by the participants is from 1 to 7. As displayed in (43), the raw scale is transformed to z-scores for the normalization of each individuals' scaling variations. The average rating of attested syllables with 0 violation is 1.41 (SD=0.69), which is close to that of attested syllables with 1 violation: 1.52 (SD=0.59). In comparison, the average values of accidental gaps (AG) and systematic gaps (SG) are rather different: the AG Mean acceptability (AG) is around positive 0.03 (SD = 0.86) while the SG Mean (SG) is about negative 0.4 (SD=0.69).

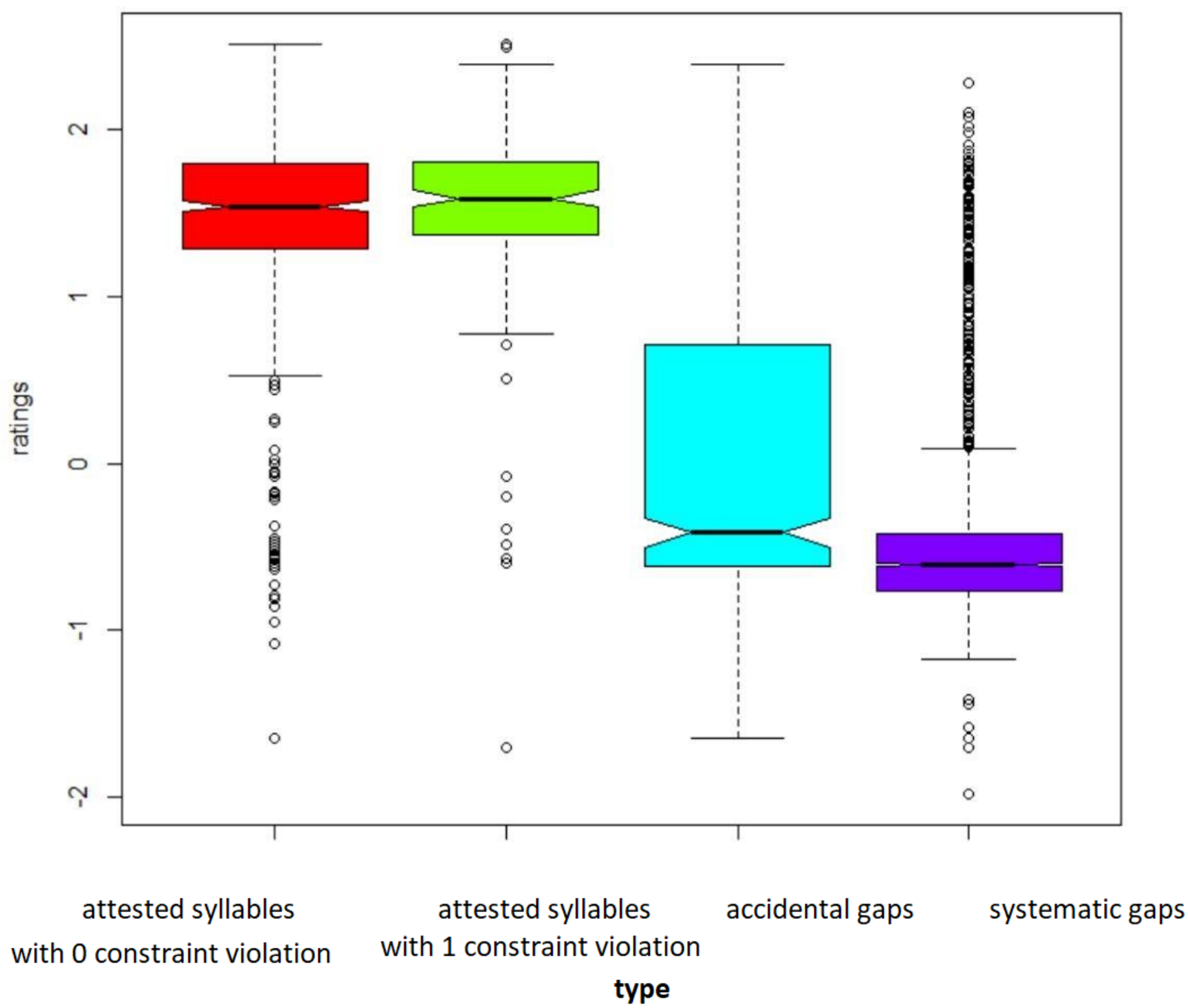


Figure 4. Effect of syllable types on acceptability ratings

(43) Acceptability ratings of different syllable types

Syllable type	zrating (mean)	zrating (median)	zrating (sd)
attested	1.40494925	1.540732	0.6855149
attested_V	1.5272722	1.5867632	0.5944262
AG	0.03462986	-0.4146903	0.8604114
SG	-0.39509098	-0.6059654	0.6900566

The fact that accidental gaps with 0 constraint violation are rated much lower ($p < 0.001$) than attested syllables with one violation (attetsed_V) shows that Mandarin speakers' acceptability ratings are affected by and more sensitive to syllable types than constraint violations. Therefore, syllable types play a bigger role in acceptability rating than the number of constraint violations. For instance, among the four Mandarin syllables: [du] (attested syllable with 0 violation), [juŋ] (attested syllable with 1 violation: *GV-HH), [dən] (accidental gap which does not violate any constraint), as well as [bjuŋ] (systematic gap which violates one constraint *GV-HH), the overall tendency is that both attested syllable (without violation [du] and with one violation [juŋ]) are significantly more acceptable than the accidental gap [dən] and the systematic gap [bjuŋ]. In order to check whether the effect of syllable type on acceptability ratings is significant, a mixed-effects linear regression is carried out as displayed in Table (44).

(44) Mixed-effects linear regression analyses of syllable types on acceptability ratings

```
> m1 <- lmer(zrating~ type +(1 | token), data=dt)
> emmeans(m1, list(pairwise ~ type), adjust = "tukey")
```

`emmeans of type`					
type	emmean	SE	df	lower.CL	upper.CL
attested	1.3958	0.0383	1413	1.3207	1.471
attested_V	1.5268	0.2105	319	1.1127	1.941
AG	0.0562	0.0579	512	-0.0576	0.17
SG	-0.3112	0.0215	913	-0.3534	-0.269
Degrees-of-freedom method: kenward-roger Confidence level used: 0.95					
`pairwise differences of type`					
	estimate	SE	df	t.ratio	p.value
attested-attested_V	-0.131	0.2139	330	-0.612	0.9282
attested-AG	1.34	0.0694	671	19.297	<.0001
attested-SG	1.707	0.0439	1266	38.868	<.0001
attested_V-AG	1.471	0.2183	329	6.737	<.0001
attested_V-SG	1.838	0.2116	322	8.687	<.0001
AG-SG	0.367	0.0618	545	5.947	<.0001

The analyses show that the acceptability ratings of syllables of each syllable type, except attested syllable with 0 violations and attested syllables with one violation, are significantly different from any other syllable types (all p-values < 0.0001). As shown in Figure 4, accidental gaps are significantly more acceptable than systematic gaps (p-value <0.0001). Meanwhile, ratings of attested syllables with 1 violation are not significantly different from attested syllables with 0 violation (p-value = 0.93), but significantly different from accidental gaps (p-value <0.0001) and systematic gaps (p-value <0.0001).

4.5 Interactions between syllable types and other factors

To explore the interactional effects between the 4-level syllable types and other factors including the number of constraint violations based on phonological generalizations ("#violations (PG)"), the overall weight of violated constraints ("weight (PG)"), the number of constraint violations in Hayes and Wilson Phonotactic Learner (HWPL) ("#violations (HWPL)") and the penalty score generated by Hayes and Wilson Phonotactic Learner ("Penalty (HWPL)"), I carried out a statistical analysis with a linear regression model. The mixed-effects linear regression model was implemented to gauge the effects of the five factors on the acceptability ratings obtained in the syllable acceptability judgment experiment. The z scores were converged from the raw acceptability ratings of the participants. Following Gong & Zhang (2018, 2019, 2021)'s method, analyses in this section were carried out using lme4 and lmerTest packages in R (see also Bates et al., 2014). "Token" is included as a random effect, while participant was not used as a random effect because the transformation of raw ratings to z scores contributes to the normalization of individuals' scaling variations.

As illustrated in the model output provided in (45) below, among syllable types, attested syllables with no constraint violations were used as the baseline. The default contrasting method used in (45) is Treatment Contrasts (contr.treatment()) (Schad, et al. 2020). The interaction between syllable type and the number of constraint violations (labeled as the factor "#violations (PG)"), as well as the interaction between syllable type and the summed weight based on phonological generalizations (labeled as the factor “weight (PG)”) were dropped in the Table (42) due to non-convergence and lack of significant yielded results. The interactions between syllable types and the two factors ("#violations (PG)" and “weight (PG)”) are discussed in more detail in Section 5.

As shown in the rows 2-5 in Table (45), the effect of the three syllable types is significant, more specifically, attested syllables with no constraint violations ($t=23.99$; $p<.001$), accidental gaps ($t=-6.31$; $p<.001$) and systematic gaps ($t=-17.14$; $p<.001$). In comparison, the ratings of attested syllables with one constraint violation are not significantly different from ratings of attested syllables without constraint violations (p-value = 0.5957). This shows that Mandarin speakers do not make further distinctions among attested syllables. Real words are regarded as equally acceptable with or without constraint violations. Therefore, cumulative categorical grammar is not predictive among attested words.

The factor ‘syllable type’ plays a major role in the acceptability of a Mandarin syllable, as illustrated in Table (45). This is in line with previous studies (Gong & Zhang 2018, 2019, 2021). To illustrate, Gong & Zhang (2018, 2019: 9, 2021: 255) carried out a series of statistical analyses on factors including syllable type, syllable duration and neighborhood density and explored their effect on acceptability ratings of Mandarin attested and unattested syllables. In Gong & Zhang’s analysis, syllable type was the major factor which significantly influenced the acceptability ratings of Mandarin attested and unattested syllables (p-value < 0.001).

(45) The lmer model for subjects' acceptability ratings

>lmer(zrating~type+#violations(HWPL)+penalty+no+weight+type:penalty+type:weight+type:no +type:#violations (HWPL)+(1 | token), data=dt)

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	1.376	0.057	1389	23.985	< 0.001 ***
typeattested_V	0.210	0.395	273	0.531	0.5957
typeAG	-1.118	0.177	517.8	-6.308	6.08e-10 ***
typeSG	-1.31	0.076	1140	-17.139	< 0.001 ***
#violations (HWPL)	0.0194	0.042	1390	0.461	0.6449
penalty	-0.011	0.065	1287	-0.169	0.8659
#violations (PG)	-0.179	0.080	835.8	-2.241	0.0253 *
weight (PG)	0.013	0.013	828	1.014	0.3111
typeattested_V:penalty	0.131	0.675	264	0.194	0.8465
typeAG:penalty	-0.048	0.079	892.6	-0.607	0.5441
typeSG:penalty	-0.002	0.065	1279	-0.036	0.971
typeattested_V:#violations (HWPL)	-0.039	0.255	269.9	-0.151	0.880
typeAG:#violations (HWPL)	-0.004	0.050	968.2	-0.084	0.933
typeSG:#violations (HWPL)	-0.014	0.042	1378	-0.323	0.7469

Given all the factors in the linear regression model, apart from syllable types, the number of violated constraints based on phonological generalizations displays a relatively bigger impact on acceptability ratings (p-value = 0.025) than other factors. However, the effect of the summed weight of all violated constraints is not significant on the acceptability of Mandarin syllables (p value = 0.3111). Likewise, the overall effect of the number of violations in Hayes and Wilson Phonotactic Learner (HWPL) (p value = 0.6449) and penalty score (HWPL) were not significant either (p value = 0.8659). The interactions between penalty score (HWPL)and syllable types, between the number of violations in Hayes and Wilson Phonotactic Learner (HWPL) ("#violations (HWPL)") and syllable types are not significant, as displayed in Table (45) (all p values >0.1). The following subsections 4.5.1 to 4.5.4 further analyze the interactions between syllable types and other factors on acceptability ratings of syllables.

4.5.1 Interaction between the number of violations (Phonological Generalizations) and syllable type

The interaction between the number of violations (based on phonological generalizations) and the 4-level syllable types (attested syllables with 0 violation, attested syllables with 1 violation, accidental gaps, systematic gaps) with respect to their effect on acceptability ratings is illustrated in Figure 5 below. Most of the attested syllables have 0 constraint violation and elicit a positive rating. Likewise, the attested syllables with one constraint violation also elicit a positive z score. On the contrary, the accidental gaps with 0 violation receive both positive and negative ratings, while the systematic gaps elicit mostly negative ratings.

With the increasing number of constraint violations, there is a slight decreasing tendency in the syllable rating value. Meanwhile, the number of violations does not seem to influence the ratings of attested words at all, because the ratings do not decrease as the number of constraint violations increases among real words. The number of the violated constraints among real words is either 0 or 1. For example, the syllable [pa] is an attested word with 0 constraint; while the syllable [juŋ] is attested with one constraint violation: *GV-HH.

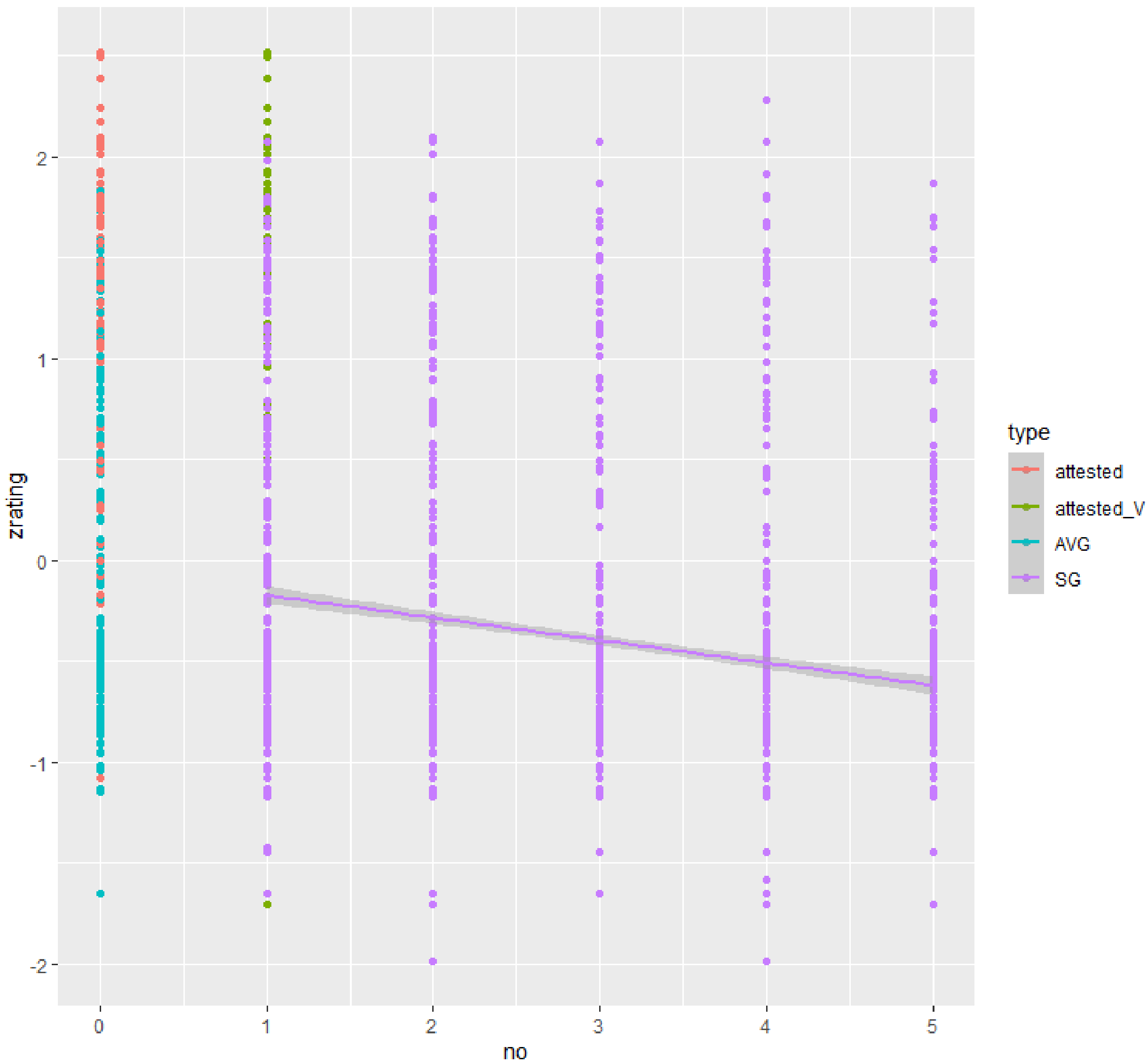


Figure 5. Effect of number of constraint violations and syllable type on acceptability ratings

The interaction between the number of violations based on phonological generalizations and different syllable types with respect to their effect on acceptability ratings does not converge in the linear regression model in Table (45). This is due to the fact that the syllable type and the number of violated constraints are highly correlated. For instance, the attested forms with 0 violations, the attested forms with one violation, accidental gaps with 0 violations and the systematic gaps with 1 or more violations.

4.5.2 Interaction between weight (generalizations) and syllable type

The interaction between the summed weight of violated constraints (based on phonological generalizations, x-axis) and different syllable types (represented with different colors) with respect to their effect on acceptability ratings (y-axis) is illustrated in the Figure 6 below. Most of the attested syllables with a weight of 0-10 received a positive (z-score) rating. Accidental gaps with a weight of 0 are assigned with both positive and negative ratings. It is observed that the effect of the summed weight on acceptability ratings is relatively weak among the less "grammatical" syllables (weight >10). Meanwhile, the value of weight does not seem to influence the ratings of attested words. In other words, the gradience in listeners' evaluations applies only to relatively well-formed unattested syllables, while the "bad" forms are rated as bad without significant gradience in ratings.

Based on the linear regression model reported in Table (45), the weight of constraint violations does not exert significant influence on the acceptability ratings of Mandarin monosyllables. The interaction between the summed weight based on phonological generalizations and different syllable types with respect to their effect on acceptability ratings does not converge. This is due to the partially one-to-one correspondence and high correlation between the syllable types and the weight (attested with a weight of 0 and accidental gap with a weight of 0).

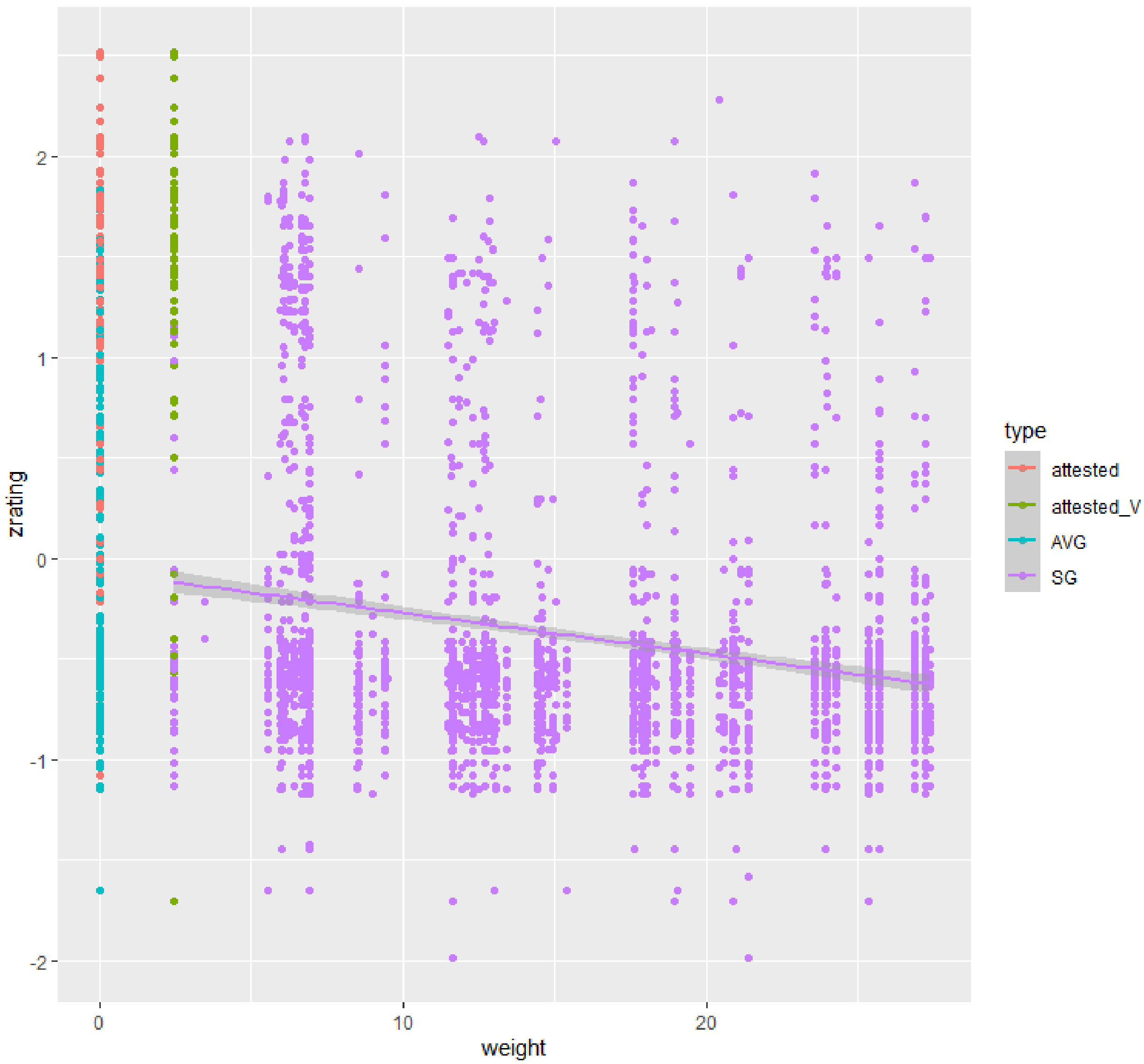


Figure 6. Effect of weight and syllable type on acceptability ratings

4.5.3 Interaction between the number of violations (Hayes and Wilson Phonotactic Learner (HWPL)) and syllable type

As shown in Figure 7, the interaction between the number of violations (Hayes and Wilson Phonotactic Learner (HWPL)) and different syllable types with respect to their effect on acceptability ratings is fitted in the linear regression model.

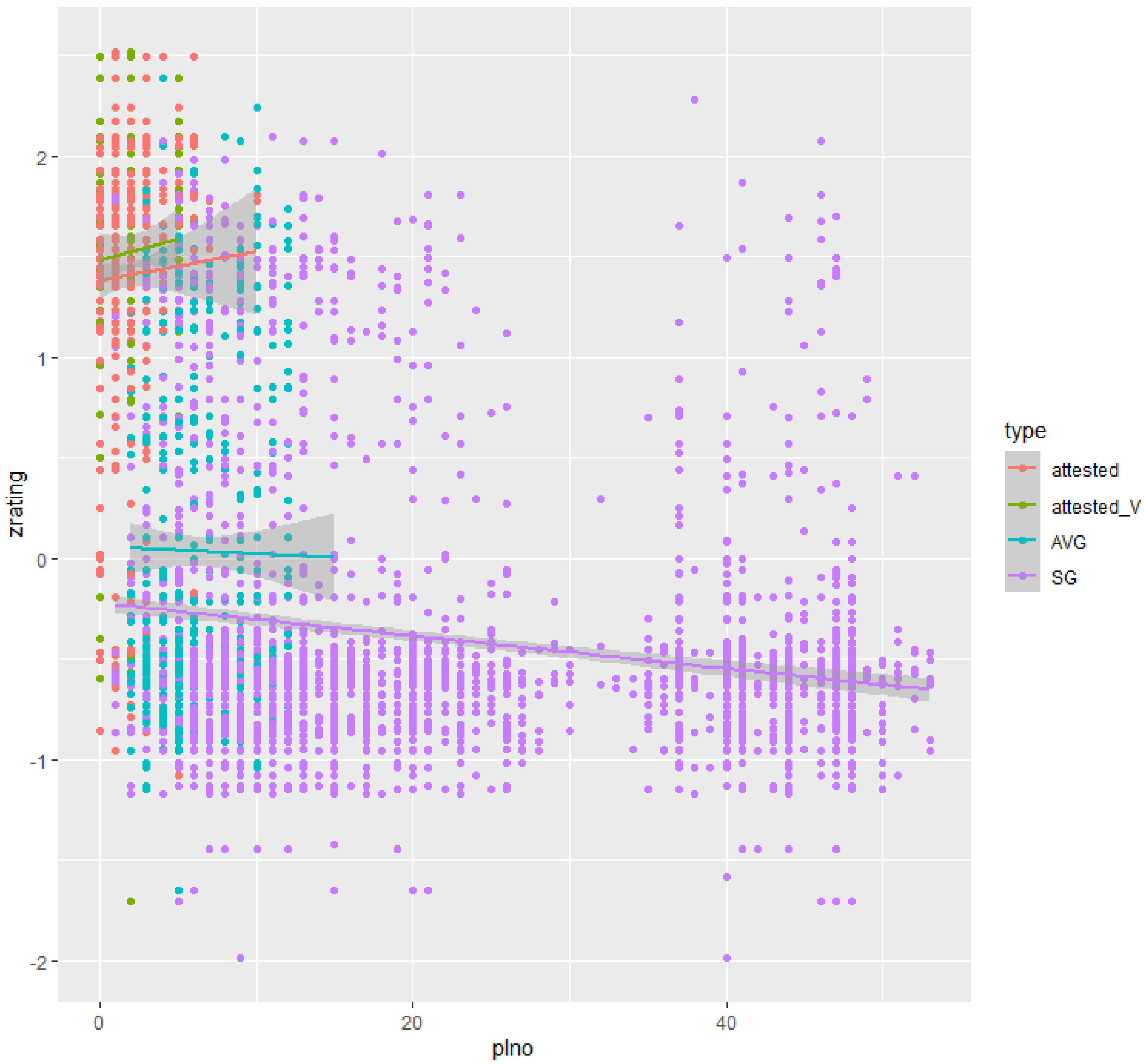


Figure 7. Effect of number of constraint violations in Hayes and Wilson Phonotactic Learner (HWPL) and syllable type on acceptability ratings

By means of the algorithm of the Hayes and Wilson Phonotactic Learner (HWPL), none of the Mandarin attested syllables violate more than 20 Hayes and Wilson Phonotactic Learner (HWPL) constraints. Not surprisingly, attested syllables are rated high in terms of their acceptability. In contrast, accidental gaps and systematic gaps display a slight negative correlation between the number of constraint violations in Hayes and Wilson Phonotactic Learner (HWPL) and the acceptability ratings. Meanwhile, it does not seem to influence the ratings of attested words.

Based on the statistical Analysis of Variance reported in Table (45), in general, only syllable type, but not the number of constraint violations in Hayes and Wilson Phonotactic Learner (HWPL) (p-value = 0.90), shows a significant influence on the acceptability ratings of Mandarin monosyllables. The interaction between the number of violations and different syllable types with respect to their effect on acceptability ratings is not significant (p-value = 0.77) either.

4.5.4 Interaction between penalty score (HWPL) and syllable type

As shown in Figure 8, the interaction between the penalty score (HWPL) (Hayes and Wilson Phonotactic Learner (HWPL)) and different syllable types with respect to their effect on acceptability ratings is fitted in the linear regression model. The penalty score (HWPL) of each Mandarin attested syllable is less than 10, and they are mostly assigned with a positive z score. In contrast, as for the accidental gaps and systematic gaps, we see a slight negative correlation between the penalty score (HWPL) in Hayes and Wilson Phonotactic Learner (HWPL) and the acceptability ratings. Meanwhile, the penalty score (HWPL) does not seem to influence the ratings of attested words at all because the ratings do not decrease as the penalty increases for real words.

Based on the statistical Analysis of Variance in the table (45), only syllable type, but not the penalty score (HWPL) in Hayes and Wilson Phonotactic Learner (HWPL) (p-value = 0.92), shows a significant influence on the acceptability ratings of Mandarin monosyllables. The interaction between the penalty and different syllable types with respect to their effect on acceptability ratings is not significant (p-value = 0.72) either.

Figure 8. Effect of penalty and syllable type on acceptability ratings

4.6 Summary

In summary, the effects of syllable types (attested syllables, accidental gaps, systematic gaps) on acceptability ratings is significant (all p values < 0.0001). The acceptability of attested syllables with 0 violations is no different from the attested syllables with one constraint violation (p value > 0.1). Meanwhile, accidental gaps overall tend to receive significantly higher acceptability than systematic gaps. These observed results are consistent with previous studies, where real words are more acceptable than accidental gaps and systematic gaps. The new observation from the current study is that within the scope of real words, 0 constraint violation and 1 constraint violation does not influence the acceptability judgment of Mandarin speakers significantly. Raters are forgiving of constraint violations as long as they get an attested form; With respect to the unattested ones, the phonotactically well-formed ones (accidental gaps) on average received a higher acceptability than the ill-formed ones (systematic gaps).

Chapter 5

Grammars and Speakers' Acceptability Ratings

The experimental data from §4 are analyzed in this section to compare 3 different models of grammars (a-c, see (46-47) below) using 2 approaches to defining constraints, manual and data-driven. The goal of the present analysis is to identify the approach which can best reflect speakers' phonotactic knowledge by best predicting the syllable acceptability ratings from native speakers. Thus 6 grammars (3 models * 2 approaches) and their effect on acceptability ratings of Mandarin syllables are discussed in this section:

(46) Models of grammars

(a) <u>categorical grammars</u>: forms are assessed on whether they have a constraint violation or not:

(i) based on manually constructed constraints and principles (illustrated in §3);

(ii) based on the constraints generated from the computer algorithm Hayes and Wilson Phonotactic Learner (HWPL).

(b) <u>cumulative categorical grammar</u>: forms are assessed based on the number of violated constraints:

(i) based on manually constructed constraints and principles;

(ii) based on the constraints generated from Hayes and Wilson Phonotactic Learner (HWPL).

(c) gradient grammar: forms are assessed based on:

(i) the weight of violated constraints as determined by Hayes & Wilson's Maxent Grammar Tool (2008);

(ii) penalty score (HWPL)generated by the Hayes and Wilson Phonotactic Learner (HWPL) (Hayes & Wilson 2008).

The two categorical grammars in (46a) are represented by the categorical grammaticality ("1" or "0") of the relevant syllables. The two cumulative categorical grammars in (46b) are represented by the number of violated constraints. The two gradient grammars in (46c) are represented by the penalty score of the relevant syllable via Hayes and Wilson Phonotactic Learner (HWPL) and the overall weight of violated constraints (manually constructed phonological generalizations). The Hayes and Wilson Phonotactic Learner (HWPL) (Hayes & Wilson 2008) is a computer algorithm, in which the frequency of real words as the learning data and a list of syllables as testing data are the language input. The output of the Hayes and Wilson Phonotactic Learner (HWPL) includes the automatically generated constraints and penalty scores (HWPL)of each syllable in the testing data.

I carried out the statistical analyses between the values of the grammars described above and the native speakers' acceptability ratings in this section to study: (i) Is a gradient grammar more or less predictive of speakers' grammaticality judgements than a categorical grammar? (ii) Is a gradient grammar derived from the Hayes and Wilson Phonotactic Learner (HWPL) ("data driven") more or less predictive of speakers' grammaticality judgements than one derived manually ("phonologically driven")? The findings then give us an answer to the question which grammar can best account for native speakers' grammaticality judgment of Mandarin attested monosyllabic words and nonwords.

(47) Statistical analyses on 6 grammars and acceptability ratings of syllables

approach	**grammar**	**independent variable**	**dependent variable**	**random effects**
manual	categorical	with or without violation (PG)	rating	token type, participant (z scores)
data-driven		with or without violation (HWPL)	rating	token type, participant (z scores)
manual	cumulative categorical	number of violated constraints (PG)	rating	token type, participant (z scores)
data-driven		number of violated constraints (HWPL)	rating	token type, participant (z scores)
manual	gradient	weight (PG)	rating	token type, participant (z scores)
data-driven		penalty score (HWPL)	rating	token type, participant (z scores)

§5.1 below illustrates the correlational analysis between each of the six grammars in Table (46) above and the acceptability ratings of Mandarin syllables. §5.2 carries out a comparative analysis between the manually constructed categorical grammar and the data-driven categorical grammar from Hayes and Wilson Phonotactic Learner (HWPL). Following that, §5.3 analyzes the effect of the number of constraint violations based on phonotactic principles (cumulative categorical grammar). §5.4 discusses the weight of violated constraints and acceptability ratings. Subsequently, in the §5.5 and 5.6, the data-driven number of constraint violations via Hayes and

Wilson Phonotactic Learner (HWPL) (cumulative categorical grammar) and penalty score (HWPL)(gradient grammar) are discussed with respect to their correlations with acceptability ratings of Mandarin syllables.

5.1 Correlational Analyses on 6 grammars

To answer the question about which factor has the strongest correlations with native speakers' acceptability ratings, this section carries out the correlational analyses between each of the six grammars and the acceptability ratings of Mandarin syllables. Results from Kendall's rank correlation tau indicate a relatively stronger correlation between the acceptability ratings of Mandarin syllables and each of the following factors: the number of violated constraints (phonological generalizations) (tau = -0.35); the number of violated constraints (Hayes and Wilson Phonotactic Learner (HWPL)) (tau = -0.35); the summed weight based on phonological generalizations (tau = -0.35); the penalty score (HWPL)(tau = -0.38). However, the categorical grammars show a medium and weak association with acceptability ratings: the binary categorical grammaticality based on phonological generalizations (tau = -0.34); the binary categorical grammaticality via Hayes and Wilson Phonotactic Learner (HWPL)) (tau = -0.19).

As displayed in the table (48), the data-driven categorical grammar (labeled as "cat (HWPL)") is weakly associated with the native speakers' acceptability ratings on syllables, while the manually constructed categorical grammar (labeled as "cat(PG)") has a comparatively stronger correlation with acceptability ratings. With respect to the cumulative categorical grammars, both the manually constructed and machine-driven number of violated constraints are correlated with syllable ratings. Meanwhile, the machine-driven gradient grammar represented by the penalty score (HWPL)of syllables has the strongest correlation with syllable ratings (rau = -0.38), while the manually constructed weight has a relatively lower association with syllable ratings (rau = -0.345).

(48) Correlational Analyses on 6 grammars and acceptability ratings

Kendall's rank correlation tau			
data	z	p value	tau
zrating and #violations (PG)	-30.822	< 0.0001	-0.3534842
zrating and #violations (HWPL)	-32.789	< 0.0001	-0.3531222
zrating and weight (PG)	-31.495	< 0.0001	-0.3451263
zrating and penalty (HWPL)	-36.181	< 0.0001	-0.3833895
zrating and cat (PG)	-26.277	< 0.0001	-0.3401302
zrating and cat (HWPL)	-14.888	< 0.0001	-0.1927193

In a nutshell, the gradient grammars and cumulative categorical grammars are overall more correlated with syllable acceptability ratings ($|\tau b| > 0.34$). However, the categorical grammars, especially the machine driven one, has a relatively lower association with syllable acceptability ratings.

5.2 Comparison of the manually constructed and data-driven categorical grammars

Two categorical grammars, the categorical grammar based on the phonological principles and the one generated by the Hayes and Wilson Phonotactic Learner (HWPL), are compared in the current study. A total of 91 out of 1255 syllables have 0 constraint violations by the Hayes and Wilson Phonotactic Learner (HWPL); while 462 out of 1255 syllables are labeled with 0 constraint violations based on the phonological constraints and principles summarized in the previous §3.

(49) Categorical grammaticality of Mandarin syllables

Categorical Grammaticality	1255 syllables	
	grammatical	ungrammatical
Phonological generalizations (§3)	462	793
Hayes and Wilson Phonotactic Learner (HWPL)	91	1164

5.2.1 Manually constructed categorical grammar

In the nonword acceptability experiment, out of 4012 tokens, 1079 were labeled as grammatical based on the phonological generalizations illustrated in §3.

As displayed in Figure 9 and Table (51), as for the 1079 grammatical syllables, the mean of the ratings (z scores) from Mandarin speakers is positive: 0.72 (SD = 1.04). In comparison, the mean of the ratings (z scores) on ungrammatical ones is negative: -0.26 (SD = 0.84). Therefore, the collected data indicate that the phonological constraints and generalizations in §3 correctly classify the Mandarin syllables binarily with respect to the well-formedness (grammatical or ungrammatical) of the stimuli.

(50) Categorical grammaticality (PG) of Mandarin syllable tokens

Categorical Grammaticality	4012 experimental tokens/ratings	
	grammatical	ungrammatical
Phonological generalizations	1079	2933

(51) Ratings of syllable groups based on categorical grammar (Phonological generalizations)

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	SD
ungrammatical	-1.9826	-0.7626	-0.5937	-0.2647	-0.2165	2.5146	0.83757
grammatical	-1.648	-0.4703	1.1679	0.7166	1.5868	2.5146	1.03685

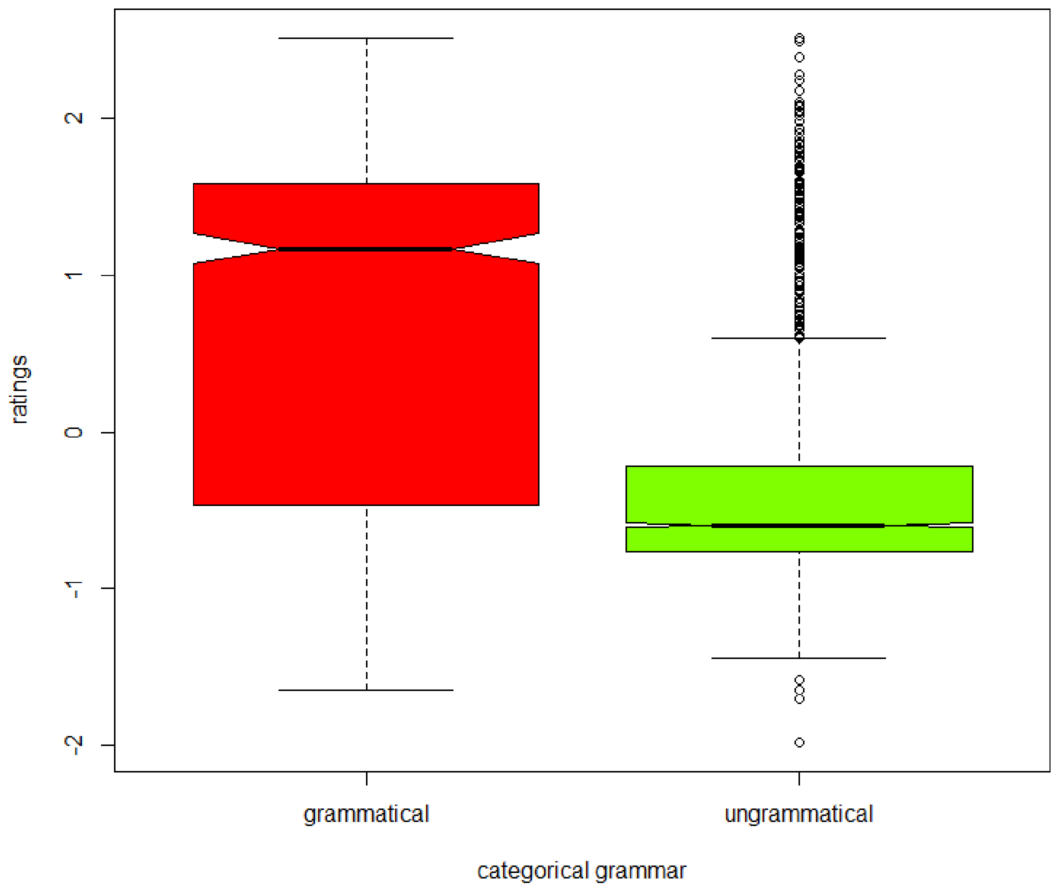


Figure 9. Ratings of syllable groups based on categorical grammar (Phonological generalizations)

5.2.2 Data-driven categorical grammar

Calculated from the Hayes and Wilson Phonotactic Learner (HWPL), 102 out of 4012 experimental tokens are grammatical. For these grammatical syllables, the mean of z scores from Mandarin speakers is positive: 1.41 (SD = 0.63). In comparison, the mean of the ratings on ungrammatical ones is negative: -0.05 (SD = 0.97). Therefore, the collected data show that the Hayes and Wilson Phonotactic Learner (HWPL) correctly classifies the Mandarin syllables binarily with respect to the well-formedness (grammatical or ungrammatical) of the stimuli.

(52) Categorical grammaticality (HWPL) of Mandarin syllable types and tokens

Categorical Grammaticality	4012 tokens/ratings from participants	
	grammatical	ungrammatical
Hayes and Wilson Phonotactic Learner (HWPL)	102	3910

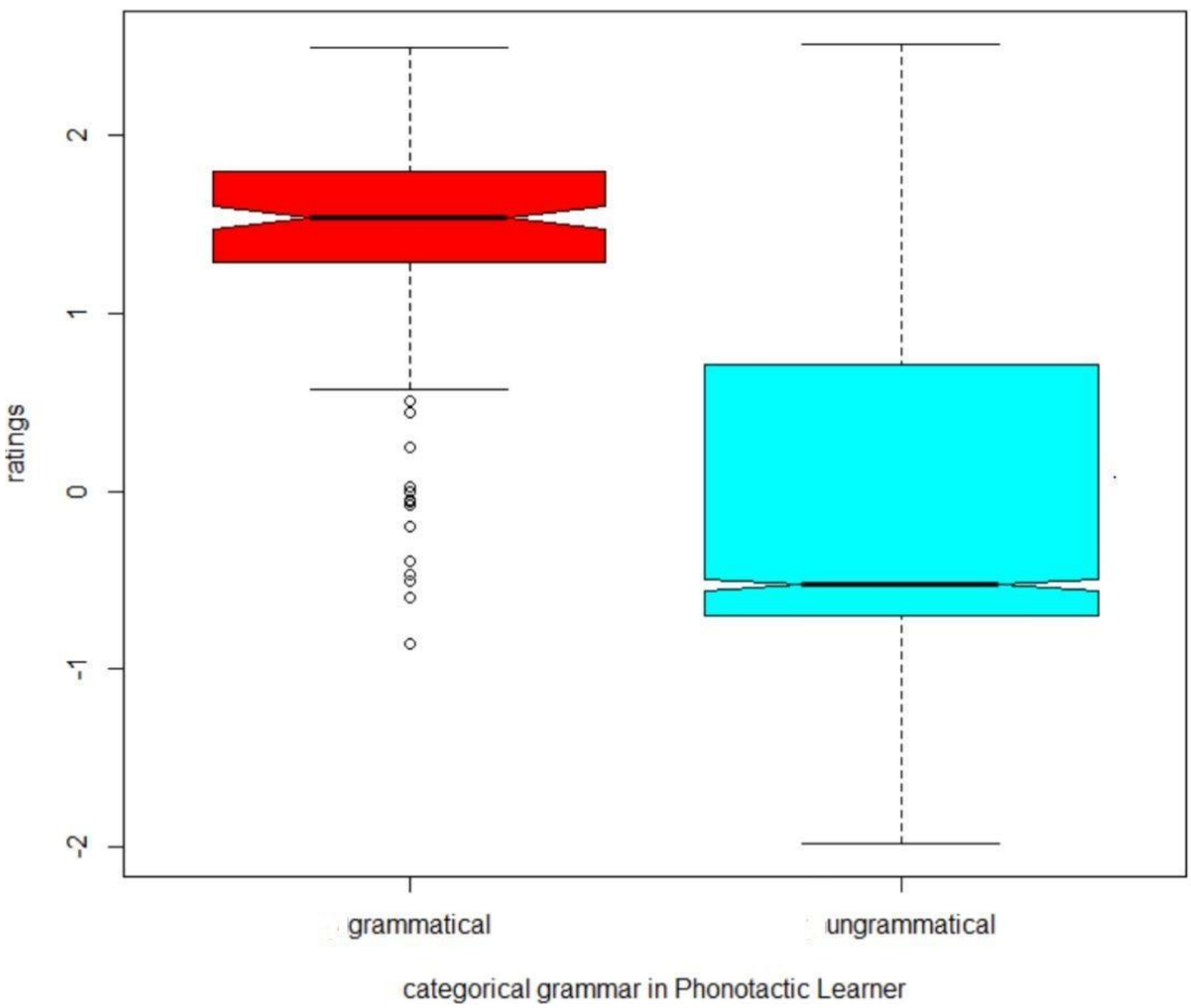


Figure 10. Ratings of syllable groups based on categorical grammar (Hayes & Wilson Phonotactic Learner (HWPL))

(53) Ratings of syllable groups based on categorical grammar (Hayes and Wilson Phonotactic Learner (HWPL))

Phonotactic Learner	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	SD
ungrammatical	-1.98256	-0.696	-0.5271	-0.0531	0.70902	2.51458	0.96793
grammatical	-0.8535	1.2839	1.5353	1.4056	1.7952	2.4916	0.63092

5.2.3 Linear Regression on categorical grammars

As displayed in Table (54) and Table (55), the mixed-effects linear regression analyses show that the z-ratings of the grammatical syllables and ungrammatical ones (based on phonological generalizations) are not significantly different (p-value = 0.54). Likewise, the zratings of the grammatical syllables and ungrammatical ones in Hayes and Wilson Phonotactic Learner (HWPL) are not significantly different (p-value =0.948) either. However, the acceptability ratings of grammatical words based on manually constructed categorical grammar are significantly different from ratings of nonwords (p-value < 0.001). Besides, the acceptability ratings of grammatical words based on data-driven categorical grammar are significantly different from ratings of nonwords (p-value < 0.001).

(54) Linear regression on binary categorical grammar (phonological generalizations)

```
> summary(lmer(zrating~ cat + wordtype + cat:wordtype +(1 | token), data=dt))
```

	Estimate	Std.Error	df	t value	Pr(>|t|)
(Intercept)	1.39585	0.03829	1331.94216	36.454	<2e-16 ***
ungrammatical	0.13098	0.21394	304.28514	0.612	0.5409
wordtype nonword	-1.33969	0.06942	622.32558	-19.298	<2e-16 ***
ungrammatical: wordtype nonword	-0.49836	0.22268	315.14987	-2.238	0.0259 *

(55) Linear regression on binary categorical grammar (Hayes & Wilson Phonotactic Learner)

> summary(lmer(zrating~ catpl + wordtype + catpl:wordtype +(1 | token), data=dt))

	Estimate	Std.Error	df	t value	Pr(>|t|)
(Intercept)	1.39495	0.08589	1330.19680	16.241	<2e-16 ***
ungrammatical	0.00623	0.09606	1318.26140	0.065	0.948
wordtype nonword	-1.66716	0.04773	1168.9402	-34.930	<2e-16 ***

Therefore, neither the manually constructed categorical grammar, nor the data-driven categorical grammar, can predict the binary differences of acceptability ratings among Mandarin syllables. Compared to the manually constructed categorical grammar, the data-driven categorical grammar from the Hayes and Wilson Phonotactic Learner (HWPL) identifies a much more narrow range of grammatical syllables. Only 91 out of 1255 stimuli types are labeled as grammatical with zero constraint violation via HWPL. Therefore, the categorical grammar based on phonological generalizations is overall more similar to the division between real words (N = 400) and nonwords (N=855) than the data-driven one.

5.3 Cumulative number of constraint violations (Phonological Generalizations) and acceptability ratings

The number of constraint violations based on phonological generalizations represents the cumulative categorical grammar which is neither as binary as categorical grammaticality, nor as gradient as weight or penalty score (HWPL). In the syllable acceptability judgment experiment, the number of violated constraints was introduced as a factor and analyzed with respect to its statistical significance. For example, the syllables [bu] is attested with zero violations, the syllable [bjuŋ] is a systematic gap which violates one constraint *GV-HH, and the syllable [gyɑu] is a systematic gap that violates four constraints. Consistent with predictions based on a cumulative categorical grammar, [bu] should be more acceptable than [bjuŋ], and [bjuŋ] should be more acceptable than [gyɑu]. These predictions were critically evaluated using Mandarin speakers' acceptability judgments.

5.3.1 Linear Regression model on #violations (PG)

In order to check whether the effect of the number of violated constraints on acceptability ratings is significant, a mixed-effects linear regression is carried out as displayed in Table (56). The overall effect of the number of violations is not significant in predicting the syllable acceptability (t value = 0.658, p = 0.511).

(56) Linear Regression on Number of violated constraints (cumulative categorical grammar)

> summary(lmer(zrating~ no + wordtype + no:wordtype +(1 | token), data=dt))

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	1.39574	0.03703	1353.72892	37.694	<2e-16 ***
no	0.13108	0.19912	269.01082	0.658	0.511
wordtype nonword	-1.34984	0.05099	947.10724	-26.471	<2e-16 ***
no: wordtype nonword	-0.28500	0.19964	269.76637	-1.428	0.155

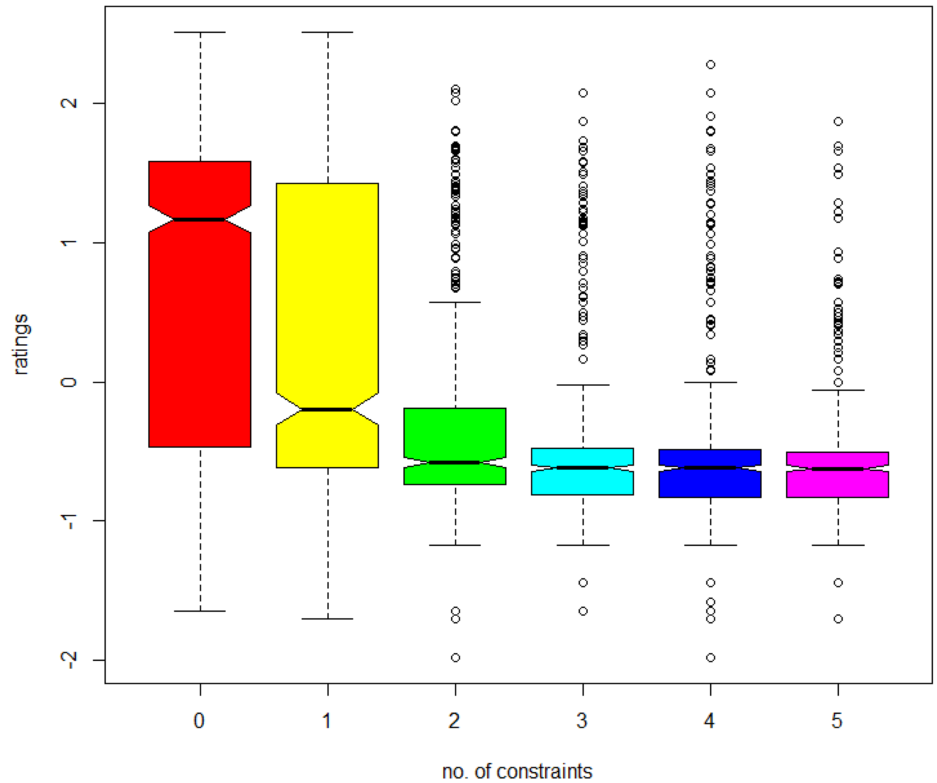


Figure 11. Number of violated constraints (phonological generalizations) and ratings

Figure 11 illustrates the correlations between the number of violated constraints (based on phonological generalizations) and the acceptability ratings. In the current acceptability judgment study, when the number of constraint violations is less than 2, the general trend is that the number of violated constraints and the ratings are negatively correlated. That is to say, the more constraints a Mandarin syllable violates, the less acceptable it is for native speakers. However, if the number of violated constraints is greater than 2, such a correlation is no longer present. This result supports the view that Mandarin speakers are more sensitive to the differences in acceptability associated with syllables featuring no or few distinctions, and they are less sensitive when judging the relative acceptability of the syllables featuring three or more constraint violations (no>2).

5.3.2 Pairwise comparison and data binning

As displayed in the table (57), the mean of the ratings for syllables with 0 constraint violations (mean=1.037) is the highest among all other syllable groups with 1 to 5 constraint violations. Furthermore, the mean of the ratings for syllables with one and two constraint violations are -0.03 and -0.30 respectively. Furthermore, the mean ratings of syllables with three, four and five constraint violations are displayed as -0.50, -0.52 and -0.58.

Further analyses show that the acceptability ratings of syllables with 0 violations are significantly different from syllables with 1 or more violations (all p-values < 0.001). Likewise, the acceptability ratings of syllables with 1 violation are significantly different from any other groups (all p-values < 0.01). In comparison, the ratings of syllables with 2, 3, 4, 5 violations are not significantly different from each other (all p-values > 0.05), as shown in Table below.

The observed results indicate that Mandarin speakers are more sensitive to the distinctions between syllables with fewer violated constraints (phonological generalizations). Based on this preliminary examination, a lower number of violated constraints (no<2) is inversely related to the acceptability of Mandarin syllables, while syllables with 2 or more violated constraints are not differentiated in that their acceptability are equally poor.

(57) Mixed-effects linear regression analyses of no on acceptability ratings
> m1 <- lmer(zrating~ nogp +(1 | token), data=dtt)
> emmeans(m1, list(pairwise ~ nogp), adjust = "tukey")

`emmeans of no`					
No of violation	emmean	SE	df	lower.CL	upper.CL
0	1.037	0.038	1035	0.963	1.1115
1	-0.0346	0.0436	1221	-0.12	0.0509
2	-0.293	0.0492	1026	-0.389	-0.1965
3	-0.5006	0.0549	848	-0.608	-0.3929
4	-0.518	0.0702	643	-0.656	-0.38
5	-0.5766	0.1497	474	-0.871	-0.2824
Degrees-of-freedom method: kenward-roger Confidence level used: 0.95					

`pairwise differences of nogp`					
	estimate	SE	df	t.ratio	p.value
0-1	1.0716	0.0578	1135	18.547	<.0001
0-2	1.33	0.0621	1029	21.415	<.0001
0-3	1.5376	0.0667	903	23.042	<.0001
0-4	1.555	0.0798	711	19.476	<.0001
0-5	1.6136	0.1544	494	10.448	<.0001
1-2	0.2584	0.0657	1106	3.935	0.0012
1-3	0.466	0.0701	971	6.652	<.0001
1-4	0.4834	0.0826	757	5.849	<.0001
1-5	0.542	0.1559	505	3.476	0.0073
2-3	0.2076	0.0737	922	2.818	0.0555
2-4	0.225	0.0857	743	2.625	0.0926

2-5	0.2836	0.1576	506	1.8	0.4668
3-4	0.0174	0.0891	711	0.195	1
3-5	0.0759	0.1594	504	0.476	0.997
4-5	0.0586	0.1654	500	0.354	0.9993

As displayed in Figure 12, some syllables with 0 violations were assigned with low acceptability ratings (1-3). This is caused by the lower ratings of accidental gaps, which are not attested but don't violate any constraint. Native speakers tend to rate accidental gaps lower than real words, as displayed in Figure 2 (see also Gong & Zhang 2019, 2021). Meanwhile, the syllables with 1 violation include both real words and systematic gaps, which leads to higher acceptability ratings (5-7). Real words with 1 violation are viewed as more acceptable than unattested syllables with 1 violation, as shown in Figure 2. Therefore, in order to analyze the data more accurately, the number of violated constraints itself is not explanatory enough to cover the complexity of acceptability ratings. Other factors like syllable type should also be taken into consideration.

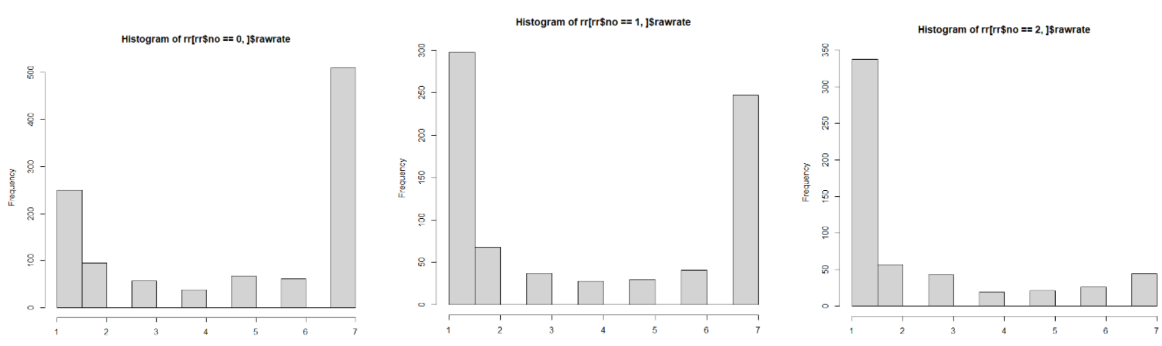


Figure 12. Acceptability ratings' distribution of syllables with 0 - 2 constraint violations

In §4, it is displayed that the maximum number of constraint violations in the Mandarin stimuli syllables is 5, while the minimum number of constraint violations is 0. The differences between syllables with 0-2 constraint violations and those with 3 -5 constraint violations are worth exploring. In Figure 12, it is illustrated that the effect of the number of constraint violations (no = 0-2) on acceptability stands out: speakers tend to assign higher acceptability ratings to syllables with fewer violations. The syllables with 0 violations are associated with the highest acceptability ratings, which is followed by syllables with 1 violation and syllables with 2 violations.

In contrast, the effect of the higher number of constraint violations (no > 2) on acceptability is not significant: the frequencies of low acceptability ratings are similar and hardly distinguishable among syllables with 2 or more violations. The frequency of low ratings (1-3) is much higher than high ratings (5-7) in Figure 13.

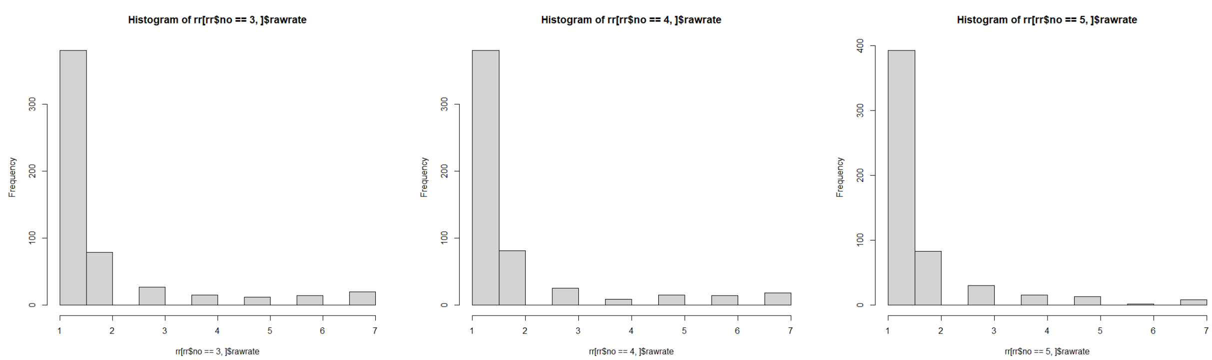


Figure 13. Acceptability ratings' distribution of syllables with 3 - 5 constraint violations

The maximum number of violated constraints in attested Mandarin syllables is 1 and the maximum number of violated constraints in unattested Mandarin syllables is 5. In summary, the effect of the number of constraint violations on acceptability ratings is not consistent between stimuli with 0-1 violations and stimuli with 2 or more violations. The effect is much weaker for syllables with 2 or more violations. These observed results are consistent with Gong & Zhang

(2018:20)'s analyses, where they maintains that highly ungrammatical forms are quickly rejected by Mandarin speakers.

5.3.3 Differences of ratings within nonwords

To check the effect of the number of violated constraints and to avoid possible influences from memory of real words, this section analyses the rating variance among nonwords only. As listed in the Table (58) below, the average rating of nonwords with 0 violations is positive: 0.035 (SD=0.86). Following that, the average ratings of nonwords with 1 and 2 violations are negative: -0.12 (SD=0.87) and -0.30 (SD=0.77) respectively. Comparatively, the average ratings of syllables with 3, 4 and 5 violations are -0.49, -0.51 and -0.57 respectively.

Within nonwords, statistical analyses show that the ratings of syllables with 0 violations (accidental gaps) are significantly different from syllables with 1 violation (p-value = 0.00393), and from syllables with 2 violations (p-value = 3.37e-11). Furthermore, ratings of syllables with 1 violation are significantly different from syllables with 2 violations (p-value = 0.000303). However, the ratings of syllables with two or more violations are not significantly different from each other (all p-values > 0.01).

(58) Ratings of nonwords with different number of constraint violations (Phonological Generalizations)

no	Zrating mean	Zrating median	Zrating sd
0	0.03462986	-0.4146903	0.8604114
1	-0.1166122	-0.5270864	0.8686936
2	-0.29573969	-0.5763949	0.7656009
3	-0.49105855	-0.6209326	0.5796221
4	-0.50563321	-0.6209326	0.6021809
5	-0.56939037	-0.6233796	0.4494051

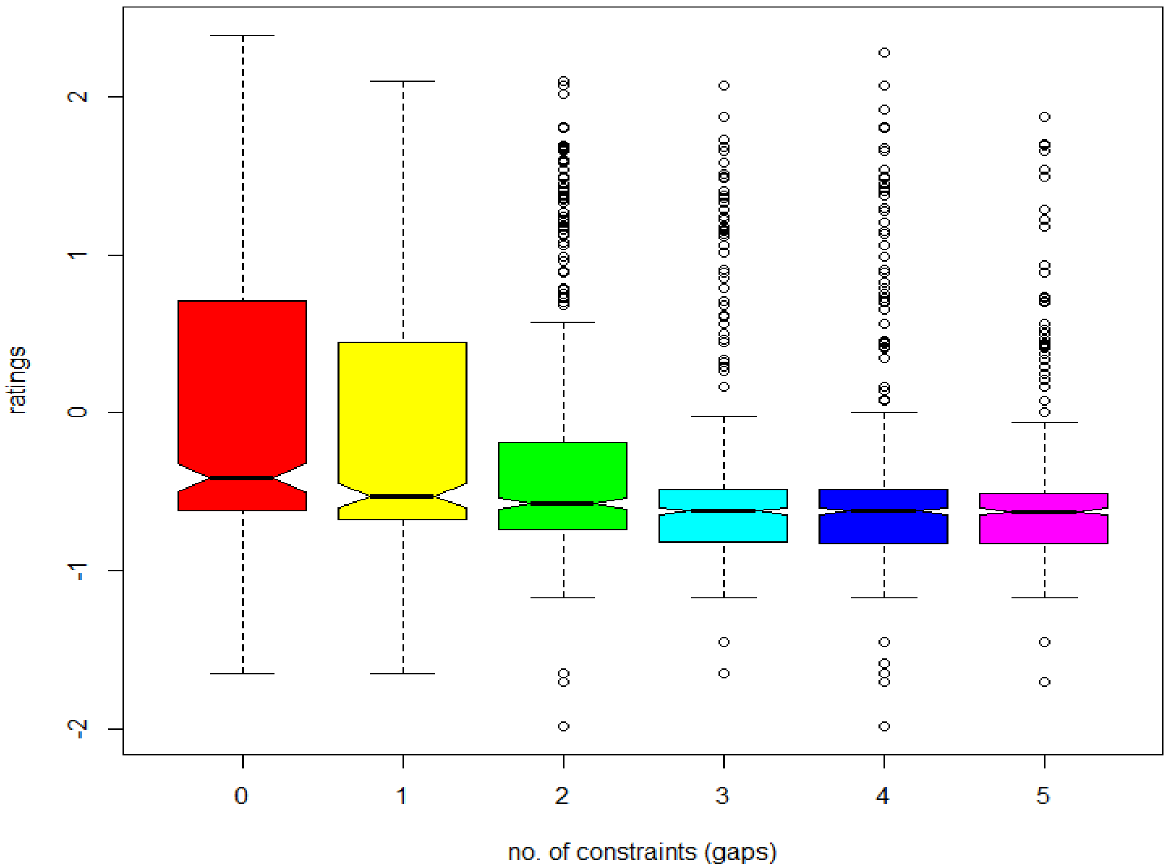


Figure 14. Ratings of nonwords with different number of constraint violations (Phonological Generalizations)

Therefore, even within nonwords only, the effect of the number of violations on acceptability ratings still persists, as seen from the average ratings associated with more "grammatical" (phonologically plausible) nonwords with less than 2 violations, but not with nonwords with 2 or more violations. In other words, for a Mandarin nonword featuring between 2 and 5 violations, its acceptability rating does not seem to reflect the number of the violated constraints.

5.4 Weight of constraints (based on Phonological Generalizations) and acceptability ratings

In this section, the summed weight of the violated constraints based on phonological generalizations represents gradient grammar of Mandarin syllables. In the syllable acceptability judgment experiment, the cumulative weight was introduced as a factor and analyzed with respect to its statistical significance. The following examples illustrate the range of cumulative weight values in some representative acceptability judgment task stimuli. The cumulative weight of the violated constraints in the attested syllable [tuo] is 0 (no violated constraints); the cumulative weight associated with the syllable [bjuŋ] is 2.46 since it violates one constraint *GV-HH (MaxEnt weight = 2.46); furthermore, the cumulative weight of the violated constraints associated with the syllable [gyɑu] is 23.9 in that it violates four constraints. If listeners are sensitive to the measure expressing the cumulative weight of the violated constraints, we predict that we would obtain higher acceptability ratings for well-formed syllables like [bu], with the lowest weight, and lower acceptability ratings for syllables like [gyɑu], associated with greater cumulative weight of the violated constraints. The following analyses explore the effect of weight on syllables' acceptability.

In Figure 15, the correlations between the summed weight of violated constraints (based on phonological generalizations, x-axis) and average acceptability ratings (y-axis) are illustrated. Fig. 15 shows that the weight of the violated constraints is overall negatively correlated with the obtained acceptability ratings; this is particularly apparent when the cumulative constraint weight of a syllable is less than 10. That is to say, the greater the weight measure is, the less acceptable the relevant Mandarin syllable is for native speakers. However, the relationship ceases to exist for constraint weight greater than 10. This reveals that Mandarin speakers are more sensitive to the distinctions between syllables with lower cumulative constraint weight (weight ≤ 10), but show less sensitivity towards syllables with higher weights (weight > 10). Therefore, the effects of the summed weight on acceptability ratings are not consistent between stimuli with a weight of 0-10 and stimuli with a weight greater than 10. This result is consistent with the findings in the previous section where the number of violated constraints is examined. This is because of a strong positive correlation between the weight measure (tested in this section) and the number of violated constraints.

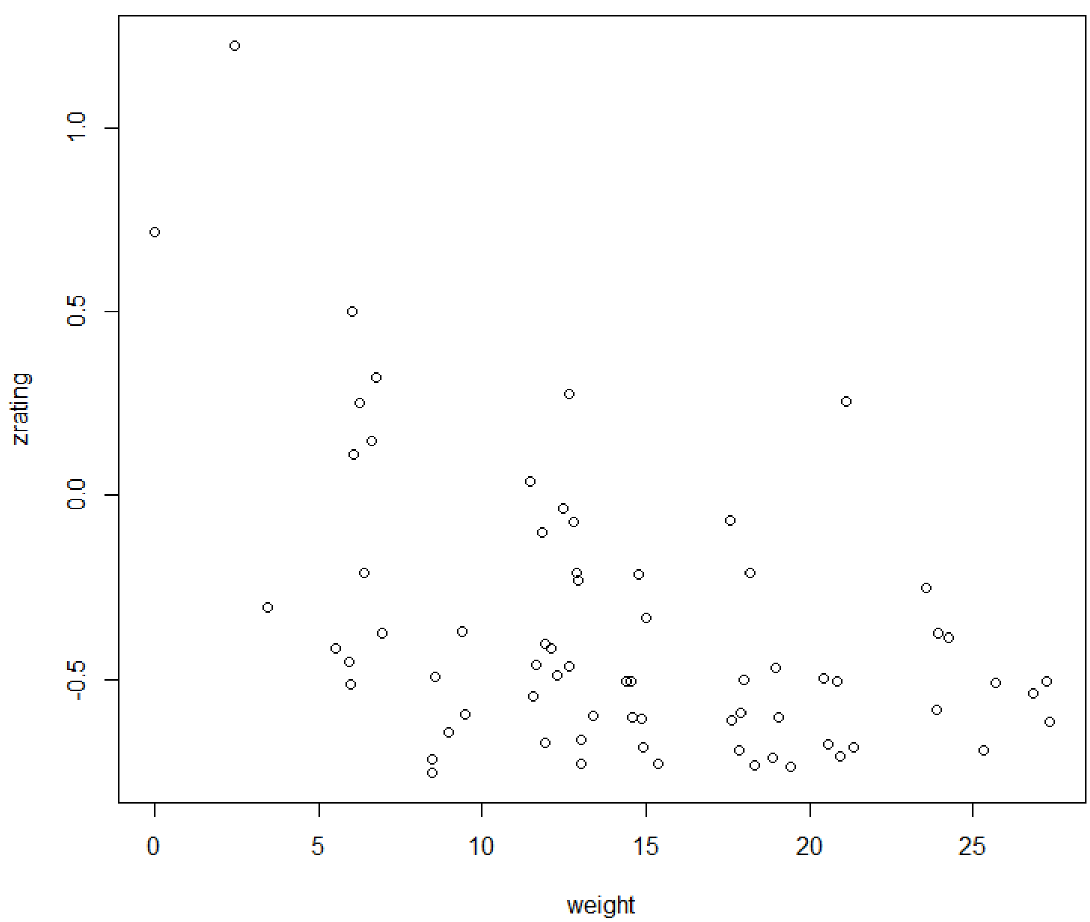


Figure 15. Mean of ratings of syllables with different weights

5.4.1 Linear Regression model on weight (PG)

In this section, to check whether the effect of the weight of violated constraints (phonological generalizations) on acceptability ratings is significant, I carried out the mixed-effects linear regression as displayed in Table 59 below. The overall effect of the weight of violations is not significant on the acceptability ratings of Mandarin syllables (t value = 0.652, p = 0.515).

(59) Statistical analyses of effect of weight on acceptability ratings (gradient grammar)

> summary(lmer(zrating~ weight + wordtype + weight:wordtype +(1 | token), data=dt))

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	1.39575	0.03719	1342.98943	37.531	<2e-16 ***
weight	0.05328	0.08166	270.60488	0.652	0.515
wordtype nonword	-1.36572	0.05138	937.16741	-26.579	<2e-16 ***
Weight: wordtype nonword	-0.08036	0.08171	270.76419	-0.983	0.326

5.4.2 Pairwise comparison and data binning

As shown in the summary of the weight of Mandarin syllables in Table (60) below, the weight range falls between 0 and 30. To study the effect of summed weight of a syllable in greater detail and finer accuracy, the syllables are divided into 6 groups ("WA" – "WF") in (61) based on their weight calculated by Maximum Entropy.

(60) Summary of weight

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	0.00	9.39	11.11	18.95	27.36

(61) Weight groups

weight range	weightgp
weight<5	WA
5<weight<10	WB
10<weight<15	WC
15<weight<20	WD
20<weight<25	WE
25<weight	WF

The acceptability ratings of the six groups of stimuli syllables based on their weight ("WA" - "WF") are displayed in Figure 16 and Table (62) below. The weights of syllables in the group "WA" are the lowest: between 0 to 5, while the syllables in the group "WF" have the heaviest weight: over 25. The mean acceptability rating for syllables with a weight of 0-5 is around 0.96, which is the highest average rating obtained in the present study. The mean ratings obtained for syllables with a weight of 5-10 and 10-15 are about -0.13 and -0.38, respectively. Following that, the average ratings of syllables with a weight of 15-20, 20-25, and 25-30 are -0.48, -0.51 and

-0.58 respectively.

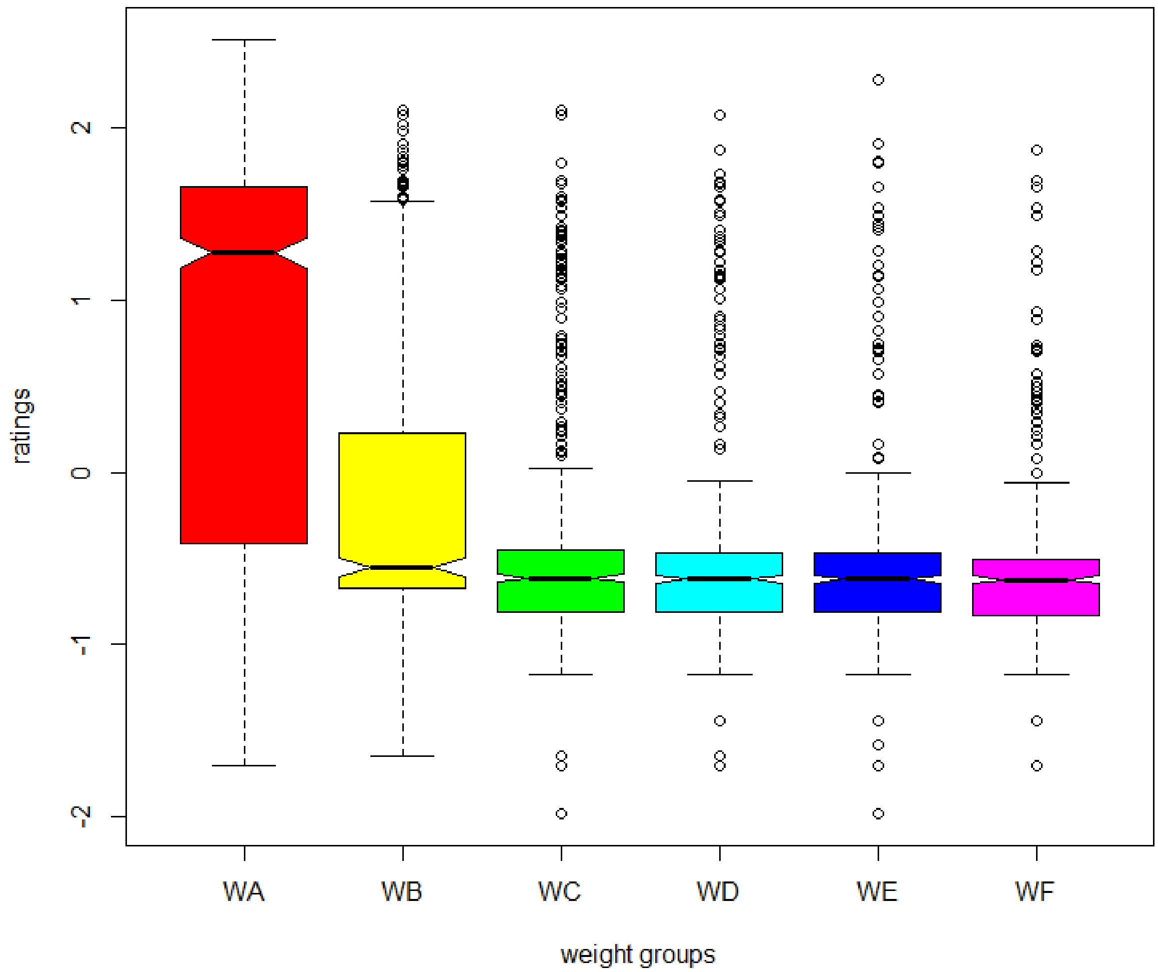


Figure 16. Ratings of syllables with different weights

Furthermore, the implemented statistical analyses show that the acceptability ratings of syllables with a weight lower than 5 (i.e. Group WA) are significantly different from any other groups of syllables (all p-values < 0.0001), as listed in Table 62. Likewise, the acceptability ratings of syllables in Group WB (5 < weight <= 10) are significantly different from any other groups (all p-values =<0.05), as shown in Table (62).

(62) Mixed-effects linear regression analyses of weight on acceptability ratings

> m1 <- lmer(zrating~ weightgp +(1 | token), data=dt)

> emmeans(m1, list(pairwise ~ weightgp), adjust = "tukey")

`emmeans of weightgp`					
weightgp	emmean	SE	df	lower.CL	upper.CL
WA	0.96	0.0367	1043	0.888	1.0315
WB	-0.129	0.0404	1180	-0.208	-0.0498
WC	-0.377	0.0504	957	-0.476	-0.2783
WD	-0.478	0.0654	779	-0.607	-0.3499
WE	-0.51	0.0847	641	-0.676	-0.3437
WF	-0.577	0.1516	477	-0.874	-0.2787
Degrees-of-freedom method: kenward-roger Confidence level used: 0.95					
`pairwise differences of weightgp`					
	estimate	SE	df	t.ratio	p.value
WA-WB	1.0885	0.0545	1115	19.957	<.0001
WA-WC	1.3368	0.0623	985	21.445	<.0001
WA-WD	1.4379	0.075	833	19.171	<.0001
WA-WE	1.4695	0.0923	688	15.927	<.0001
WA-WF	1.5361	0.1559	495	9.85	<.0001
WB-WC	0.2483	0.0646	1036	3.845	0.0018

WB-WD	0.3494	0.0769	868	4.545	0.0001
WB-WE	0.381	0.0938	710	4.062	0.0008
WB-WF	0.4476	0.1569	502	2.854	0.051
WC-WD	0.1011	0.0826	839	1.224	0.8252
WC-WE	0.1327	0.0985	708	1.347	0.7585
WC-WF	0.1993	0.1597	507	1.248	0.813
WD-WE	0.0316	0.107	688	0.296	0.9997
WD-WF	0.0982	0.1651	512	0.595	0.9914
WE-WF	0.0666	0.1736	510	0.383	0.9989

In comparison, the ratings of syllables in Group WC (10 < weight <= 15), Group WD (15 < weight <= 20), Group WE (20 < weight <= 25) and Group WF (25 < weight) are not significantly different from each other (all p-values > 0.5). Therefore, as for Mandarin syllables with a weight above 10, their acceptability ratings are not significantly distinguishable from each other. In other words, native speakers are less sensitive to the differences among syllables with higher weight (weight threshold > 10).

In summary, the effects of the summed weight of constraint violations on acceptability ratings are not consistent between syllables with lower weight (<10) and syllables with higher weight (>10). The effect is much weaker for syllables with higher weights. These observed results are consistent with the analyses in §5.3 and with results reported in Gong & Zhang (2018: 20), who reported that highly ungrammatical forms are readily rejected by Mandarin speakers.

5.4.3 Acceptability ratings of nonwords

To check the effect of the summed weight and mitigate possible effects of the attested words on acceptability judgements, this section analyses the rating of nonword stimuli only. As listed in the Table below, the average ratings of nonwords with a weight of 0-5 is positive: 0.018 (SD=0.86). Furthermore, the average ratings of nonwords with a weight of 5-10 and 10-15 are negative: -0.15 (SD=0.85) and -0.41 (SD=0.66), respectively. As summarized in Table 63, average ratings of the syllables with a weight of 10-15, 15-20, 20-25 and >25 are -0.41, -0.48, -0.49 and -0.57 respectively.

(63) Ratings of nonwords in different weight groups

weight range	weightgp	Zrating mean	Zrating median	Zrating sd
weight<5	WA	0.01807662	-0.4242641	0.8605005
5<weight<10	WB	-0.15094739	-0.5501529	0.8544748
10<weight<15	WC	-0.41391155	-0.6153034	0.6638666
15<weight<20	WD	-0.47567844	-0.6209326	0.616256
20<weight<25	WE	-0.49166113	-0.6209326	0.6123218
25<weight	WF	-0.56939037	-0.6233796	0.4494051

The statistical analyses conducted using acceptability judgements of the non-word stimuli show that the ratings of syllables with a weight smaller than 5 are significantly higher than syllables with a weight of 5-10 (p-value < 0.001), and from syllables with a weight of 10-15 (p-value < 0.001). Furthermore, ratings of syllables with a weight of 5-10 are significantly different from syllables with a weight of 10-15 (p-value < 0.001). However, the ratings of syllables with a weight above 10 are not significantly different from each other, as shown in Table (63).

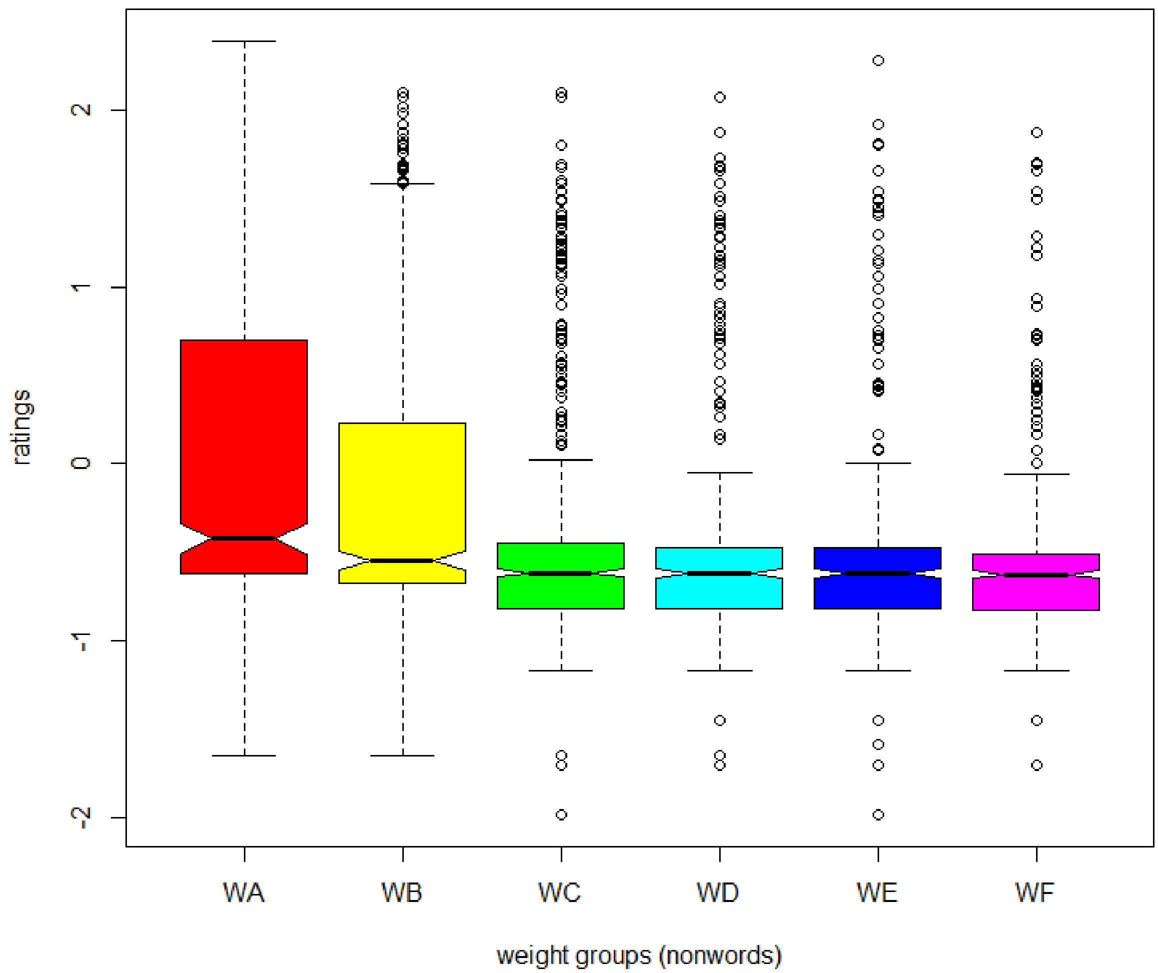


Figure 17. Ratings of nonwords in different weight groups

Therefore, among nonwords, the effect of the summed weight on acceptability ratings is present for more “grammatical” nonwords with a weight of 0-10, but not for nonwords with a weight greater than 10. In other words, within the weight range 0-10, the higher the cumulative constraint violation weight of a Mandarin nonword is, the lower its acceptability rating.

However, for a Mandarin nonword with a weight over 10 (the summed weight of violated constraints), its acceptability rating does not co-vary with the summed weight.

5.5 Number of constraint violations (Hayes and Wilson Phonotactic Learner (HWPL)) and its effect on acceptability ratings

The number of violations of constraints generated by the Hayes and Wilson Phonotactic Learner (HWPL) is analyzed in this section. This represents what I call the "cumulative categorical grammar" because it's neither binary as categorical grammaticality ("grammatical" or "ungrammatical"), nor gradient, as weight or a penalty score (HWPL). In the syllable acceptability judgment experiment, the number of violated constraints in Hayes and Wilson Phonotactic Learner (HWPL) was introduced as a factor and analyzed with respect to its statistical significance. To illustrate the relationship between the number of the violated constraints and the elicited acceptability ratings, consider the following examples. The attested syllable [ha] violates two constraints generated by Hayes and Wilson Phonotactic Learner (HWPL), the syllable [gyau] is a systematic gap which violates nine constraints, and the systematic gap [hiui] violates forty-four constraints. If the number of constraint violations of a syllable generated by Hayes and Wilson Phonotactic Learner (HWPL) and its acceptability are correlated, the three syllables will receive significantly different acceptability ratings from Mandarin speakers.

5.5.1 Linear Regression model on #violations (HWPL)

In order to check whether the effect of the number of violated constraints (Hayes and Wilson Phonotactic Learner (HWPL)) on acceptability ratings is significant, the mixed-effects linear regression is carried out as displayed in Table 64. The overall effect of the number of violations (HWPL) is not significant in predicting the syllable acceptability (p = 0.527).

(64) Statistical analyses of the effect of #violations (HWPL, cumulative categorical grammar)

> summary(lmer(zrating~ #violations (HWPL) + wordtype + #violations (HWPL):wordtype +(1 | token), data=dt))

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	1.37495	0.05457	1321.63075	25.196	<2e-16 ***
#violations (HWPL)	0.01471	0.02324	1293.044	0.633	0.527
wordtype nonword	-1.46163	0.06294	1138.481	-23.222	<2e-16 ***
#violations(HWPL): wordtype nonword	-0.02718	0.0233	1286.58863	-1.167	0.243

Figure 18 illustrates the correlations between the number of violated constraints (Hayes and Wilson Phonotactic Learner (HWPL)) of syllables and their acceptability ratings. The z-scores are mostly negative for syllables with 10 or more constraints which indicates that they are rated with a relatively low acceptability. The range of the relevant z scores are similar and hardly distinguishable from the z-scores reported for syllables with 10 or more violations. In the current acceptability judgment study, when the number of constraint violations is less than 10, the general trend is that the number of violated constraints and the ratings are negatively correlated. That is to say, the more constraints in Hayes and Wilson Phonotactic Learner (HWPL) a Mandarin syllable violates, the less acceptable the form is for native Mandarin speakers. However, if the number of the violated constraints is 10 or greater than 10, there is no obvious variation among the syllables' acceptability degree, and it does not correlate with syllable acceptability. Therefore, Mandarin speakers are more sensitive to the distinctions between syllables with fewer constraint violations (no < 10), and they show less capacity in differentiating syllables with more constraint violations (no>10 or no=10).

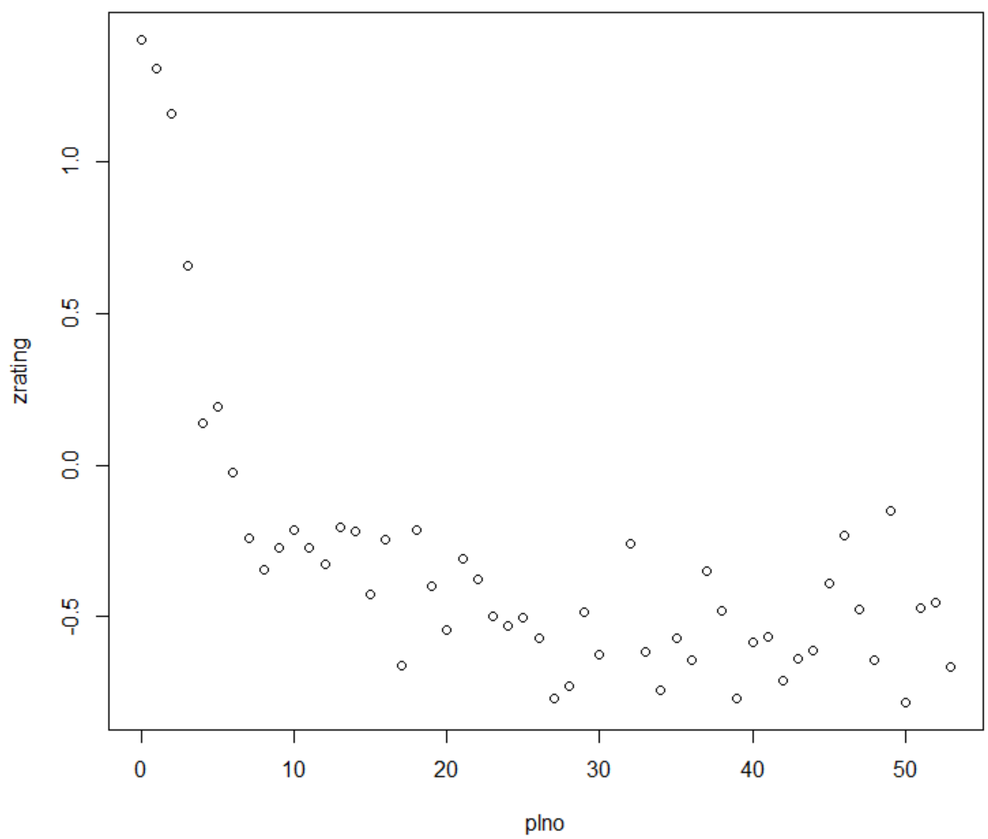


Figure 18. Average ratings of syllables with different #violations (HWPL)

The observed results indicate that Mandarin speakers are more sensitive to the distinctions among syllables with fewer violated constraints (Hayes and Wilson Phonotactic Learner (HWPL)). The lower number of violated constraints is better correlated with the acceptability of Mandarin syllables, compared to the larger number of violated constraints.

5.5.2 Pairwise comparison and data binning (what point the grammar breaks down)

As shown in Table (65), the number of the violated constraints (#violations (HWPL)) ranges between 0 and 53. To examine the effect of number of constraint violations more closely, the syllables are divided into 6 groups (“PNA” - “PNF”) in (66) based on the calculations by Hayes and Wilson Phonotactic Learner (HWPL):

(65) Summary of #violations in Hayes and Wilson Phonotactic Learner (HWPL)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0	5.0	9.0	15.7	22.0	53.0

(66) Different #violations (HWPL) group

#violations (HWPL) range	#violations (HWPL) group
#violations (HWPL) <= 10	PNA
10 < #violations (HWPL) <= 20	PNB
20 < #violations (HWPL) <= 30	PNC
30 < #violations (HWPL) <= 40	PND
40 < #violations (HWPL) <= 50	PNE
50 < #violations (HWPL)	PNF

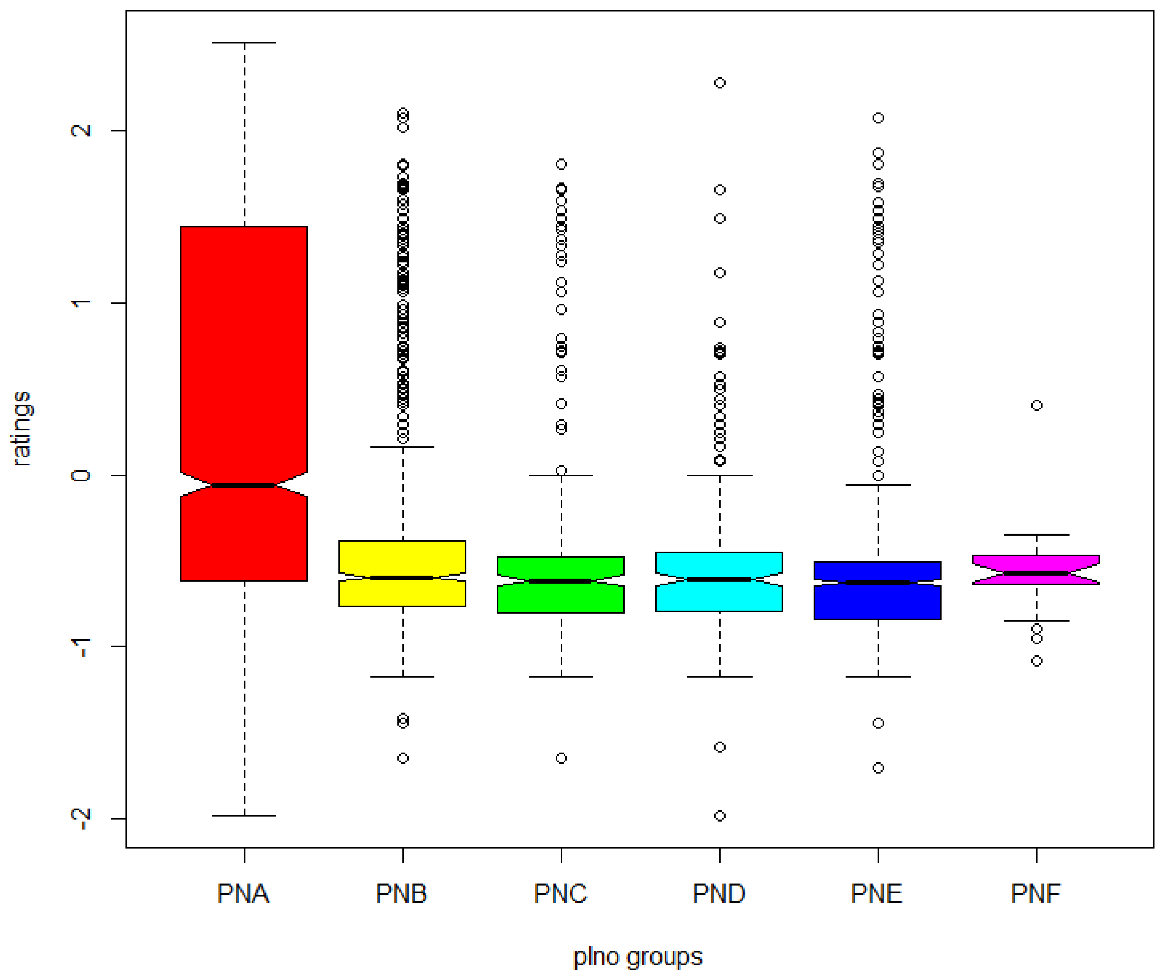


Figure 19. Ratings of different #violations (HWPL) groups

The acceptability ratings of the six groups of stimuli syllables based on their weight ("PNA" - "PNF") are displayed in Table (67). The number of constraints violated by syllables in the group "PNA" is the lowest: between 0 to 10, while that of the syllables in the group "PNF" is the highest: over 50. The mean of ratings for syllables with 0-10 violated constraints is around 0.43, which is the highest average rating. Following that, the means of ratings for syllables with 10-20 and 20-30 violations are about -0.29 and -0.46 respectively. The average ratings of syllables with 30-40, 40-50 and over 50 violations are -0.57, -0.49 and -0.53 respectively.

Further statistical analyses show that the acceptability ratings of (Group PNA) syllables with 0-10 violations (Hayes and Wilson Phonotactic Learner (HWPL)) are significantly different from Group PNB-PNE (p-values < 0.001).

(67) statistical analyses between syllable groups with different number of violations (PL)

> m1 <- lmer(zrating~ #violations (HWPL)gp +(1 | token), data=dt)

> emmeans(m1, list(pairwise ~ #violations (HWPL)gp), adjust = "tukey")

`emmeans of #violations (HWPL)gp`					
#violations (HWPL)gp	emmean	SE	df	lower.CL	upper.CL
PNA	0.433	0.0325	1047	0.369	0.497
PNB	-0.289	0.0583	986	-0.404	-0.175
PNC	-0.462	0.0907	976	-0.639	-0.284
PND	-0.571	0.1554	805	-0.876	-0.266
PNE	-0.492	0.1291	705	-0.745	-0.238
PNF	-0.53	0.4393	707	-1.392	0.333
Degrees-of-freedom method: kenward-roger Confidence level used: 0.95					
`pairwise differences of #violations (HWPL)gp`					
	estimate	SE	df	t.ratio	p.value

PNA-PNB	0.7223	0.0668	1000	10.817	<.0001
PNA-PNC	0.8945	0.0963	983	9.289	<.0001
PNA-PND	1.0044	0.1588	814	6.326	<.0001
PNA-PNE	0.9247	0.1331	720	6.945	<.0001
PNA-PNF	0.9628	0.4405	709	2.186	0.2457
PNB-PNC	0.1722	0.1078	979	1.598	0.6002
PNB-PND	0.2821	0.166	825	1.699	0.5325
PNB-PNE	0.2025	0.1417	744	1.429	0.7093
PNB-PNF	0.2405	0.4432	711	0.543	0.9944
PNC-PND	0.1099	0.1799	844	0.611	0.9903
PNC-PNE	0.0302	0.1578	781	0.192	1
PNC-PNF	0.0683	0.4486	716	0.152	1
PND-PNE	-0.0796	0.2021	762	-0.394	0.9988
PND-PNF	-0.0416	0.466	717	-0.089	1
PNE-PNF	0.0381	0.4579	707	0.083	1

However, the acceptability ratings of syllables with more than 10 violations in Group PNB, PNC, PND, PNE and PNF are not significantly different from any other groups (all p-values > 0.5). Therefore, Mandarin syllables with a number of violated constraints (Hayes and Wilson Phonotactic Learner (HWPL)) above 10 are not significantly distinguishable from each other. In other words, native speakers are not sensitive to the differences among syllables with higher number of violations in Hayes and Wilson Phonotactic Learner (HWPL).

In summary, the effects of the number of constraint violations in Hayes and Wilson Phonotactic Learner (HWPL) on acceptability ratings are not consistent across stimuli with fewer violations (<10) and stimuli with more violations (>10). The effect is much weaker for syllables with more violations. These observed results are consistent with the previous sections where the number of violated constraints and weight are tested. The data show the ratings of the highly ungrammatical items are not distinguishable from each other. Therefore, analyses tapping into effects of syllable well-formedness reveal that highly ungrammatical forms yield comparably low ratings regardless of the grammar model which is used to quantify the extent of syllable ungrammaticality and phonotactic ill-formedness.

5.5.3 Differences of ratings within nonwords

To check the effect of the number of violated constraints in Hayes and Wilson Phonotactic Learner (HWPL) and to avoid possible influences from memory of real words, this section analyses the rating variance among nonwords only. As listed in the Table (68) below, the average ratings of nonwords with 0-10 violations is -0.16 (SD=0.83). Following that, the average ratings of nonwords with 10-20 and 20-30 violations are negative: -0.35 (SD=0.73) and -0.48 (SD=0.60) respectively. Comparatively, in the last four rows of the table, the average ratings of syllables with 20-30, 30-40, 40-50 and over 50 violations are -0.48, -0.52, -0.56 and -0.53 respectively.

(68) Ratings of nonwords in syllable groups with different number of violations in Hayes and Wilson Phonotactic Learner (HWPL)

#violations (HWPL) range	dts#violations (HWPL)gp	zrating mean	zrating median	zrating sd
#violations (HWPL) <= 10	PNA	-0.1607767	-0.5270864	0.8281995
10 < #violations (HWPL) <= 20	PNB	-0.3482791	-0.5936696	0.7321947
20 < #violations (HWPL) <= 30	PNC	-0.4786391	-0.6153034	0.5984511
30 < #violations (HWPL) <= 40	PND	-0.5187842	-0.6059654	0.4915715
40 < #violations (HWPL) <= 50	PNE	-0.555487	-0.6258243	0.5214963
50 < #violations (HWPL)	PNF	-0.5273202	-0.5713348	0.3520181

Using acceptability ratings of nonwords only, statistical analyses show that the ratings of syllables with fewer than 11 violations are significantly different from syllables with 11-20 violations (p-value < 0.001), and from syllables with 21-30 violations (p-value < 0.001). However, the ratings of syllables with 10-53 violations are not significantly different from each other, as shown in the previous Table (67).

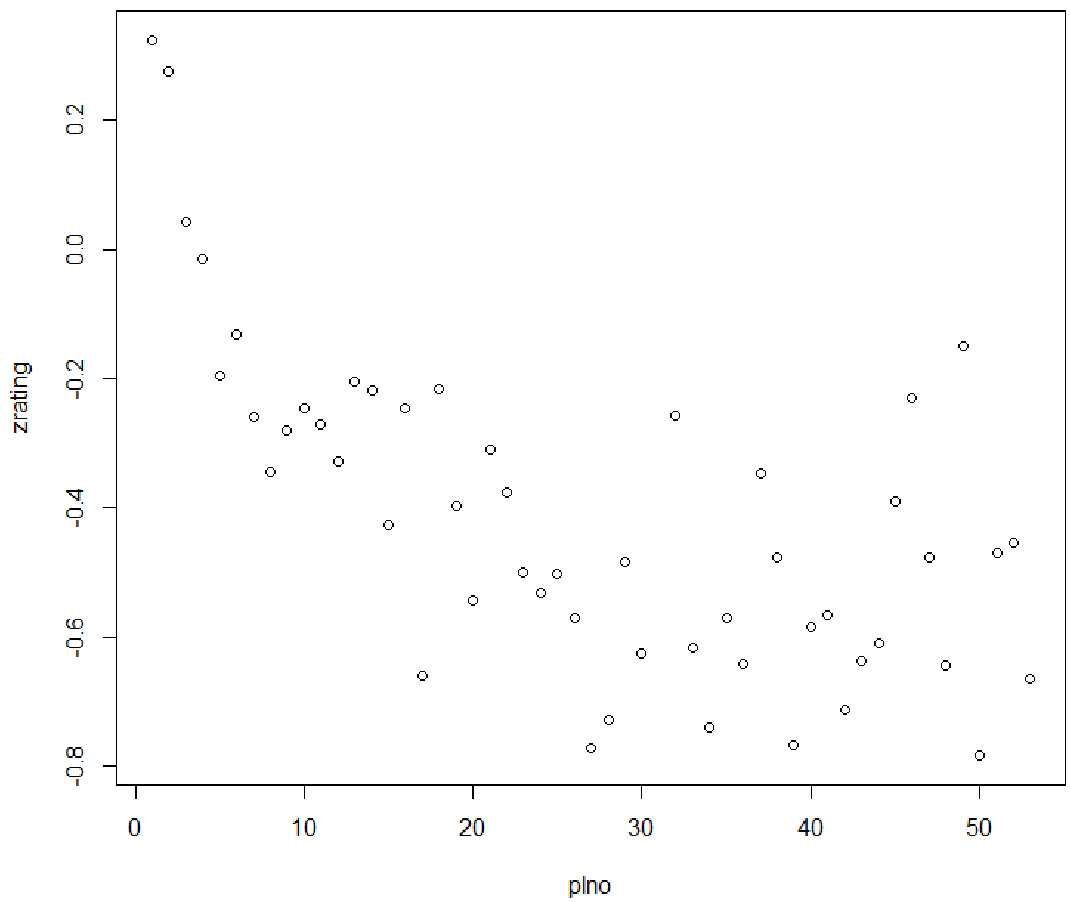


Figure 20. Average ratings of nonwords with different numbers of constraint violations in Hayes and Wilson Phonotactic Learner (HWPL)

Therefore, among nonwords, the effect of the number of constraint violations (Hayes and Wilson Phonotactic Learner (HWPL)) on acceptability ratings is more pronounced for more "grammatical" nonwords with 0-10 violations, but not for nonwords with over 10 violations. In other words, for a Mandarin nonword with more than 10 violations, its acceptability rating does not vary with the number of constraints.

5.6 Penalty score (HWPL) (Hayes and Wilson Phonotactic Learner (HWPL)) and acceptability ratings

The penalty score (HWPL) is generated by Hayes and Wilson Phonotactic Learner (HWPL) which represents the gradient grammar. The effect of penalty score (HWPL)on acceptability ratings of syllables is illustrated in this section. For example, the syllable [juŋ] is attested with a penalty score (HWPL) 0.296, the syllable [fai] is an accidental gap with a penalty score (HWPL) 3.403, while the systematic gap [hjui] has a generated penalty of 41.81 in Hayes and Wilson Phonotactic Learner (HWPL). If there exists a correlation between penalty scores (HWPL) of Mandarin syllables and their acceptability, the ratings of the three syllables would display significant differences from each other.

5.6.1 Linear Regression model on penalty (HWPL)

In order to check whether the effect of the penalty scores (HWPL) calculated via Hayes and Wilson Phonotactic Learner (HWPL) on acceptability ratings is significant, a mixed-effects linear regression is carried out as displayed in Table 69. The overall effect of the penalty score (HWPL)is not significant in predicting the syllable acceptability (t value = 0.40, p value = 0.69).

In Figure 21 below, the correlations between the penalty score (HWPL) in Hayes and Wilson Phonotactic Learner (HWPL) and acceptability ratings from native speakers are illustrated. From the data in this Figure, we see that when the penalty is lower than 10, the penalty score (HWPL)and the ratings are negatively correlated. That is to say, the higher the penalty score (HWPL) of a Mandarin syllable, the less acceptable it is for native speakers. However, if the penalty score (HWPL) is higher than 10, there is no obvious variation among the syllables' acceptability degree. Therefore, the Mandarin speakers are more sensitive to the distinctions between syllables with low penalty (penalty < 10), while speakers show less capacity in differentiating syllables with high penalty (penalty >10).

(69) statistical analyses on the effect of(Hayes and Wilson Phonotactic Learner-HWPL) penalty (gradient grammar)

> summary(lmer(zrating~ penalty + wordtype + penalty:wordtype +(1 | token), data=dt))

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	1.38294	0.05676	1325.77750	24.366	<2e-16 ***
penalty	0.01425	0.03571	1283.962	0.399	0.690
wordtype nonword	-1.33973	0.06808	1067.78	-19.679	<2e-16 ***
penalty: wordtype nonword	-0.03491	0.03577	1279.30458	-0.976	0.329

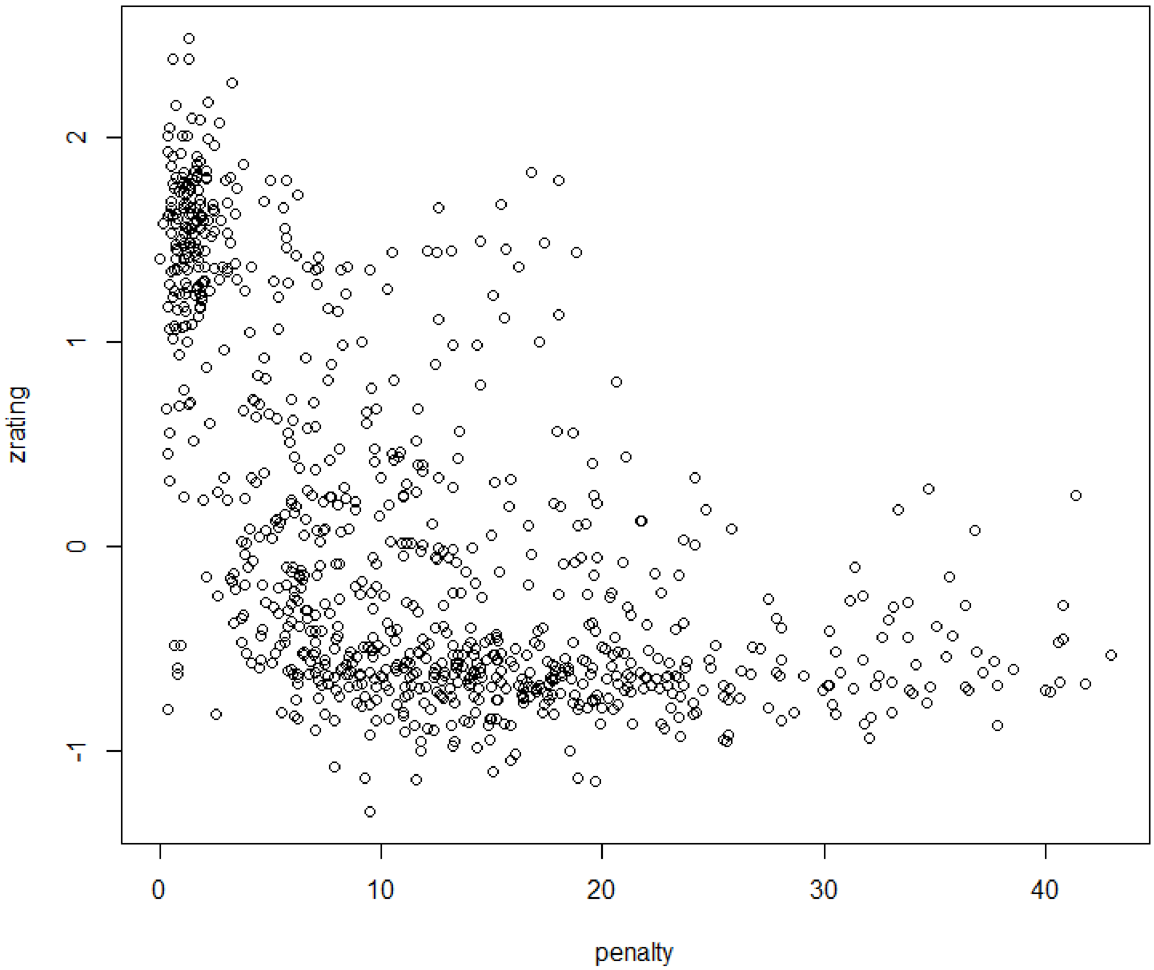


Figure 21. Average ratings of syllables with different penalty score (HWPL)

5.6.2 Pairwise comparison and data binning

In Table (70) below, the summary of the penalty score (HWPL)in Hayes and Wilson Phonotactic Learner (HWPL) of Mandarin syllables shows that the penalty range falls between 0.00 and 43.013. To study the effect of penalty score (HWPL)in greater detail and finer accuracy, the syllables are divided into 5 groups (PTA-PTE) in Table (71), based on the calculations by Hayes and Wilson Phonotactic Learner (HWPL):

(70) Summary of penalty score (HWPL)

```
---------------------------------------------------------------
   Min.  1st Qu.  Median    Mean   3rd Qu.    Max.
  0.000   4.698  12.496   15.219  22.798   43.013
---------------------------------------------------------------
```

(71) Penalty groups

penalty range	penaltygp
penalty <= 10	PTA
10 < penalty <= 20	PTB
20 < penalty <= 30	PTC
30 < penalty <= 40	PTD
40 < penalty	PTE

The acceptability ratings of the five groups of stimuli syllables based on their penalty score (HWPL)("PTA" - "PTE") are displayed in Figure 22 and Tables (71)-(72). The penalty score (HWPL)of syllables in the group "PTA" is the lowest: between 0 to 10, while those of the syllables in the group "PTE" are the highest: over 40. The mean of ratings for syllables with a penalty of 0-10 is around 0.64, which is the highest average rating. Following that, the means of ratings for syllables with a penalty of 10-20 and 20-30 are about -0.33 and -0.53 respectively.

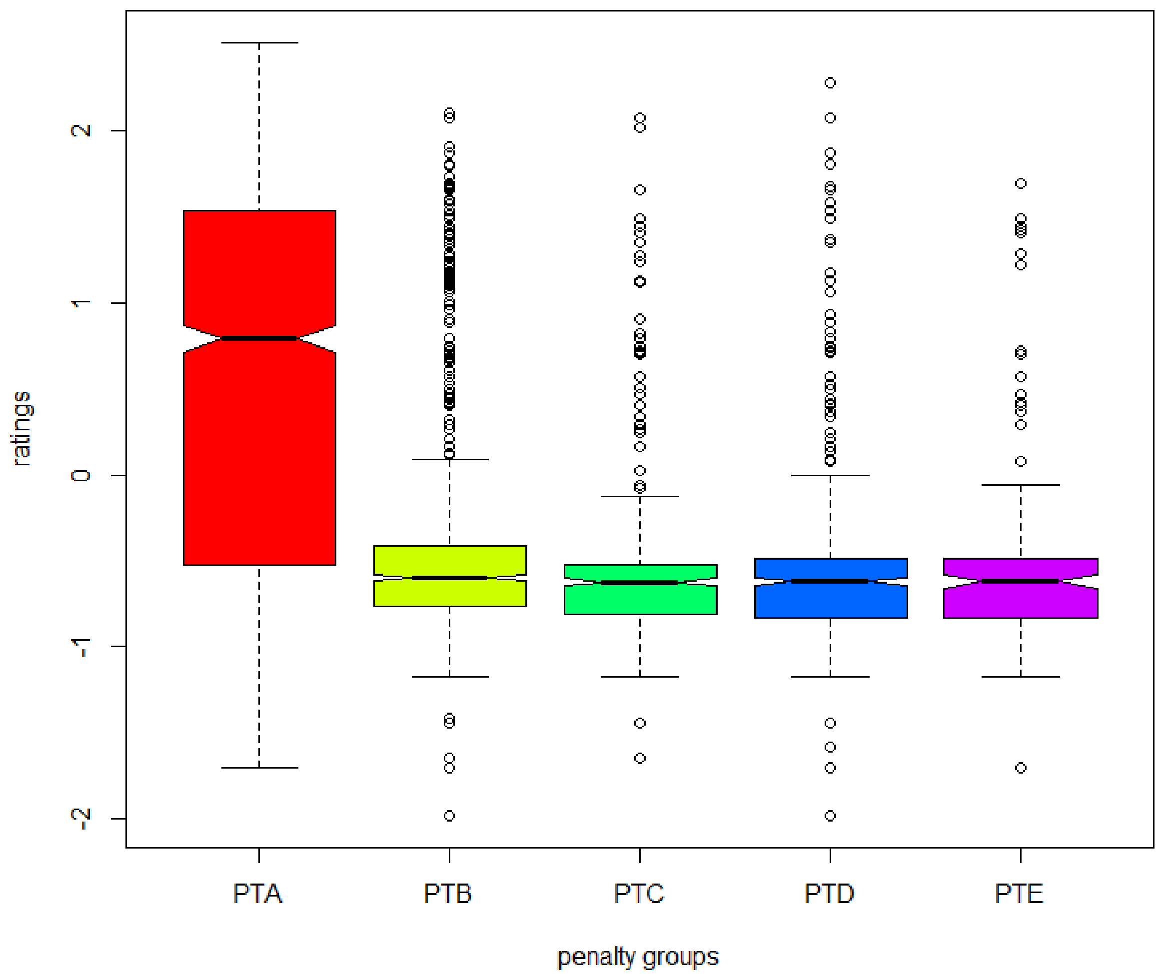


Figure 22. Ratings of syllables in different penalty groups

The average ratings of syllables with 30-40 and over 40 are -0.53 and -0.47 respectively. The observed results show that Mandarin speakers are more sensitive to the distinctions among syllables with a lower penalty score (HWPL)in Hayes and Wilson Phonotactic Learner (HWPL). The lower penalty scores (HWPL) are better correlated with the acceptability of Mandarin syllables, compared to the higher ones.

(72) Statistical analyses between syllable groups with different penalty (HWPL)

```
> m1 <- lmer(zrating~ penaltygp +(1 | token), data=dt)
> emmeans(m1, list(pairwise ~ penaltygp), adjust = "tukey")
```

`emmeans of penaltygp`					
penaltygp	emmean	SE	df	lower.CL	upper.CL
PTA	0.637	0.0328	1074	0.573	0.7017
PTB	-0.328	0.042	1001	-0.411	-0.2459
PTC	-0.53	0.0736	851	-0.675	-0.3856
PTD	-0.532	0.101	680	-0.73	-0.3337
PTE	-0.471	0.2196	596	-0.902	-0.0398
Degrees-of-freedom method: kenward-roger Confidence level used: 0.95					
`pairwise differences of penaltygp`					
	estimate	SE	df	t.ratio	p.value
PTA-PTB	0.96568	0.0533	1028	18.114	<.0001
PTA-PTC	1.16747	0.0806	883	14.485	<.0001
PTA-PTD	1.16929	0.1061	708	11.016	<.0001
PTA-PTE	1.10847	0.2221	603	4.992	<.0001
PTB-PTC	0.2018	0.0848	885	2.381	0.1215
PTB-PTD	0.20362	0.1093	717	1.862	0.3389
PTB-PTE	0.14279	0.2236	606	0.639	0.9687
PTC-PTD	0.00182	0.1249	733	0.015	1
PTC-PTE	-0.059	0.2316	616	-0.255	0.9991
PTD-PTE	-0.06082	0.2417	609	-0.252	0.9991

Further statistical analyses indicate that the acceptability ratings of syllables with a penalty score (HWPL) lower than 10 in Group PTA are significantly different from any other groups of syllables (all p-values < 0.0001). However, the acceptability ratings of syllables with a penalty score (HWPL) over 10 in Group PTB, PTC, PTD and PTE are not significantly different from each other (all p-values > 0.1). Therefore, Mandarin speakers are less sensitive to the differences among highly “ungrammatical” syllables (high penalty score (HWPL)).

In summary, the effects of the penalty score (HWPL)in Hayes and Wilson Phonotactic Learner (HWPL) on acceptability ratings are not consistent between stimuli with a penalty below 10 and stimuli with a penalty above 10. The effect is much weaker for syllables with higher penalty scores (HWPL). These observed results are consistent with the analyses in previous sections where highly ungrammatical tokens are not distinguishable for Mandarin speakers.

5.6.3 Acceptability ratings of nonwords

To check the effect of the penalty score (HWPL)in Hayes and Wilson Phonotactic Learner (HWPL) and to avoid possible influences from memory of real words, this section analyses the rating variance among nonwords only. As listed in the Table (73) below, the average ratings of nonwords with a penalty of 0-10 is negative: -0.03 (SD=0.86), which indicates that they are rated with relatively low acceptability. Following that, the average ratings of nonwords with a penalty of 10-20 and 20-30 are negative: -0.37 (SD=0.73) and -0.53 (SD=0.53) respectively. Comparatively, in the last three rows of the table, the average ratings of syllables with a penalty of 30-40 and >40 are -0.55 and -0.53.

(73) Ratings of nonwords in different penalty groups

penalty range	penaltygp (nonwords)	zrating mean	zrating median	zrating sd
penalty <= 10	PTA	-0.0264753	-0.455846	0.8597977
10 < penalty <= 20	PTB	-0.3679237	-0.5961604	0.7280207
20 < penalty <= 30	PTC	-0.5294551	-0.6233796	0.5336176
30 < penalty <= 40	PTD	-0.5491982	-0.6209326	0.5096566
40 < penalty	PTE	-0.5258986	-0.6209326	0.5376396

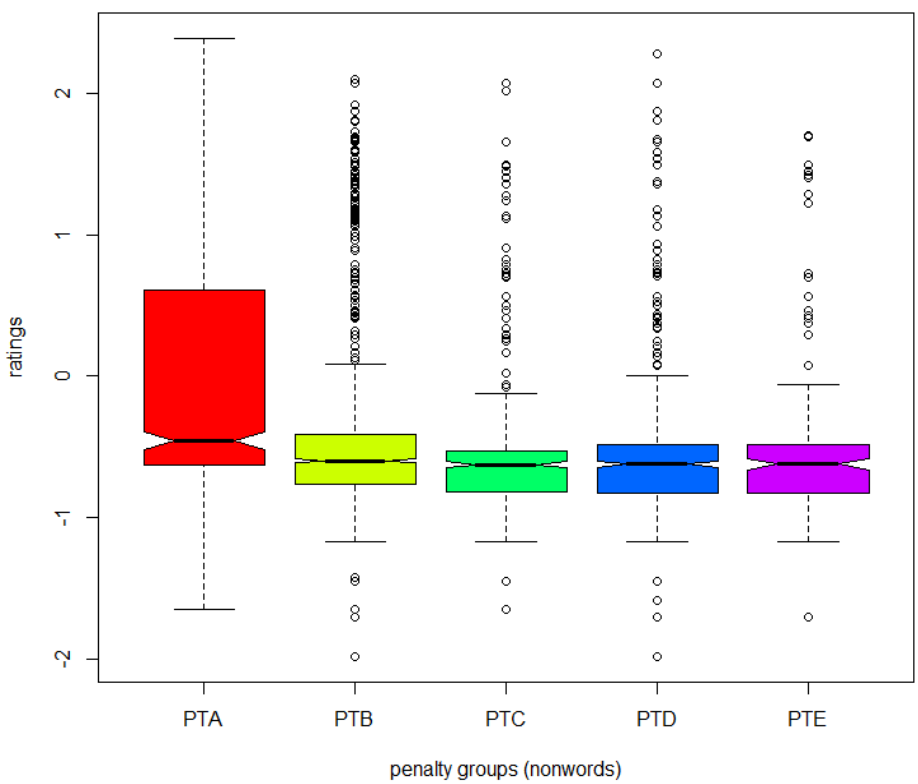


Figure 23. Ratings of nonwords in different penalty groups

Within nonwords, statistical analyses show that the ratings of syllables with a penalty of 0-10 are significantly different from syllables with a penalty of 11-20 (p-value < 0.001), and from syllables with a penalty of 21-30 (p-value < 0.001). However, the ratings of syllables with a penalty above 10 are not significantly different from each other, as shown in the previous section.

Therefore, even among nonwords only, the effect of the penalty score (HWPL)on acceptability ratings still exists for more “grammatical” nonwords with a penalty of 0-10, but not nonwords with a penalty greater than 10. In other words, within the penalty range 0-10, the higher the penalty of a Mandarin nonword is, the less its acceptability rating. However, for a Mandarin nonword with a penalty over 20, its acceptability rating does not vary with the penalty.

5.7 Discussion

The previous experiments on Mandarin nonword acceptability judgments were carried out with binary (categorical) nonword acceptability judgments, without distinctions among systematic gaps, or using a relatively smaller range of Mandarin syllable tokens as their stimuli, and not including all the (C)(G)V(X) combinations. Gong & Zhang (2019, 2021) carried out a Mandarin nonword judgment experiment and found that systematic gaps received lower acceptability ratings than accidental gaps, allophonic gaps and tonal gaps (see also Myers & Tsay, 2005; Myers, 2002). These results show gradience in acceptability among various syllable types: allophonic gaps, tonal gaps, real words and systematic gaps. Gong & Zhang’s experiment (2019, 2021) didn’t further divide the Mandarin systematic gaps and does not differentiate the grammaticality within systematic gaps. Gong & Zhang (2021: 241) argued that both lexical statistics and grammatical constraints play a role in Mandarin speakers’ phonotactic judgment. In Gong & Zhang (2021: 265)'s modeling procedures, they checked the phonetic naturalness bias, the allophony bias and the suprasegmental bias. Gong & Zhang (2021: 260-265) adjusted the weights of three types of constraints: systematic, allophonic and suprasegmental. They then found that adjusting the three learning biases leads to the improvement of the model's predictions. Based on their experimental evidence, Gong & Zhang (2021: 271) suggests that extralexical factors and biases are included in native speakers' phonotactic knowledge. In the present study, the allophonic constraints and suprasegmental restrictions are not included, which can be explored in future experiments and research to further compare categorical and gradient grammars.

The ratings of the accidental gaps display higher variation, as displayed in Figure 4 and (43). It also shows a bimodal distribution with some accidental gaps rated as high as real words and others rated as low as nonwords. The present syllable acceptability judgment experiment does not further divide stimuli within accidental gaps. Therefore, the experimental analyses collapse those distinctions within accidental gaps. The distinctions among ratings of accidental gaps could be caused by factors including sequence frequency, dialectal differences and phonetic overlap, among many others. Future studies can further differentiate accidental gaps and check the effect of these factors.

With respect to the constraints, the Phonotactic Learner overall generated 400 data-driven constraints, which is far more than the 13 manually constructed ones in (35). The manually constructed backness agreement principle is found in the data-driven constraints: *[-back][+continuant,+back][+word_boundary], *[+sonorant,-coronal][-back][-labial,+back] and *[+back][-back], . The OCP *HH principle corresponds to the data-driven constraints: *[-distributed][-labial,+high][+high], *[+consonantal,+sonorant,-lateral][+high][+high],

*[+voice][-labial,+high][+high], *[+aspiration][-labial,+high][+high], as well as *[-coronal,+high] [+high,-back]. The OCP *Lab Lab principle corresponds to the data-driven constraints: *[-continuant,-coronal][+labial,+phonemic][+labial] and *[+labial][+labial] [-word_boundary]. Future research can further explore the comparison between the cumulative categorical grammar with manually selected constraints and the cumulative categorical grammar with data driven constraints (the pure HWPL model).

5.8 Summary

In summary, the data driven categorical grammar is weakly associated with the native speakers' acceptability ratings on syllables (tau = -0.19), while the manually constructed categorical grammar has a comparatively stronger correlation with acceptability ratings (tau = -0.34). With respect to the cumulative categorical grammars and gradient grammars, both the manually constructed and data-driven ones are more correlated with syllable ratings (all |tau| >= 0.35). The data-driven gradient grammar (penalty score (HWPL)) has the strongest association with syllable acceptability (tau = -0.38), while the data-driven categorical grammar has the lowest (tau = -0.19).

Compared to other factors, the effect of syllable types (attested, accidental gap, systematic gap) on acceptability ratings stands out ($p < 0.001$). However, neither the manually constructed categorical grammar, nor the data-driven categorical grammar, can predict the binary differences of acceptability ratings among Mandarin syllables.

Meanwhile, the cumulative categorical grammars ("#violations (PG)" and "#violations (HWPL)") and the gradient grammars ("weight (PG)" and "penalty (HWPL)") can partially predict the acceptability ratings of syllables. There appears to be a gradient decreasing tendency in the ratings of the “more grammatical” syllables (threshold: no<2, weight<10, penalty<10), as the number of violations, violation weight, and penalty increase. Yet the ratings of the highly ungrammatical nonwords (no>2, weight>10, penalty>10) are indistinguishable from each other.

Chapter 6

Phonetic Experiment on Mandarin glide [J]

Inside a Mandarin CGV(X) syllable where both a consonant and a glide occur before the nuclear vowel, the status of the glide seems to be ambiguous: either (1) the initial consonant and the following glide are regarded as the primary and secondary articulations respectively (C^G); or (2) the glide is segmental instead of a secondary articulation (CG). According to Wang (1999: 128, 129)'s analysis, the consonant and the glide are phonologically different segments, exemplified in (74a). Alternatively, some studies (Chao 1934: 42) argue CG is one sound because phonetically, the word-initial consonant and glide are produced almost simultaneously. Duanmu (2000: 16-28, 71-81, 2007) argues that initial consonant and glide both sit in the onset position and treats CG as a single sound which stands in the same timing slot. Duanmu (2000, 2007) pointed out phonetic evidence: unlike English [sw], Mandarin [sw] are produced at the same time.

In (74b), the initial consonant [n] is the primary articulation (coronal, nasal), whereas the following glide [^{j}] is the secondary articulation (dorsal, palatal). Besides, whether the coarticulation between the glide [J] and the following vowel could be affected by the vowel features is also discussed in the following analyses.

(74)

a. 捏 [nje] 'pinch' (H. Wang's transcription, 1999)

b. 捏 [n^{j}e] 'pinch' (Duanmu's transcription, 2000, 2007)

Kochetov (2016: 2-18) maintains that glides can trigger the preceding consonant to acquire secondary palatal/labial articulation or shift its primary articulation. For instance, in Kirundi, the word /ku-kubit-w-a/ is articulated as [gukubitk^{w}a], where a velar is produced with secondary labial articulation. In Kochetov's study (2016: 3, 27-28), the possible strategies to avoid hetero-morphemic or tauto-morphemic C + G[j, w] combinations include the following: (a) the primary articulation spreads its features (e.g., nasality) to the secondary articulation as in /am-ja/ > [am-ɲa]; (b) one of two articulations is deleted as in /pja/ > [pa]; (c) two articulations merge together as in /mja/ > [ɲa]; (d) two articulations switch positions as in /apja/ > [ajpa]. Different from Kirundi, in Mandarin C + G sequences, features of both the consonant and the glide are kept, but whether the glide is articulated secondarily or not is controversial.

The present phonetic experiment is intended to investigate Mandarin syllable structure by examining the features of the glide [J]. Is it a segmental glide (75a), or is it palatalization of the preceding consonant (75b)?

(75) Mandarin CJ sequences

(a) Cj segmental glide?

(b) C^j palatalization?

Based on Ladefoged & Maddieson (1996: 364)'s study, in Russian, the segmental glide [j] has a prolonged steady duration of F2 between the preceding stop release and the following vowel, unlike the palatalized [^{j}] at the secondary position. In Suh & Hwang (2016)'s phonetic study on Russian and Korean palatal glides, the results show that the Korean [J] greatly overlaps with the vowel, while the Russian palatalization greatly overlaps with the consonant. In their study, the tongue position of back vowels was more prone to change when followed by the front glide [J], which leads to a higher F2. However, they found that the Korean [J] is similar to both the Russian palatalization and the Russian segmental glide [j], depending on which of four factors is measured (Suh & Hwang, 2016): (i) similar to Russian palatalized [^{j}], Korean glide /J/ is relatively short without a "prolonged steady state"; (ii) like the Russian segmental glide, there is not a strong vowel-to-glide coarticulation for the Korean glide /J/; (iii) distinguished from both Russian palatalized [^{j}] and segmental [j], the Korean /J/ shows a strong glide-to-vowel coarticulation. Following Suh & Hwang's experimental approach (2016), the current experiment analyzes the Mandarin glide [J] in terms of four factors: the duration, the F2 slope change "throughout the vocoid (glide + vowel)", the vowel-to-glide coarticulatory effect, and the glide-to-vowel coarticulation.

The experiment is carried out to examine the features of Mandarin glide /j/ in terms of four phonetic factors. If the Mandarin glide /J/ shows a short and unsteady duration with a weak glide-to-vowel coarticulation, it behaves more like the Russian palatalization (75b), which suggests that the Mandarin glide [J] is palatalized as a secondary articulation after the consonant. On the contrary, if the Mandarin [J] shows a relatively long steady duration with a weak vowel-to-glide coarticulation, it behaves more like Russian segmental glide [j] (75a), which suggests that the Mandarin glide [J] is segmental instead of a palatalized secondary articulation. Otherwise, the Mandarin [J] might behave partially like Russian palatalization and partially like Russian segmental glide [j] based on the four measurements, which indicates that the status of the Mandarin glide [J] can have the features of both palatalization and segmental glide, as in the case of the Korean glide.

6.1 Methods

The current experiment follows the methodology adopted by Suh & Hwang (2016). Suh & Hwang (2016)'s phonetic study on Russian and Korean glides involved a reading production task in which participants were asked to read sentences which included the target tokens. Suh & Hwang (2016) analyzed the phonetic features of the Korean pre-vocalic glide /J/ and compared it with both the palatalized and segmental Russian glide /j/. In their experiment, a two-way contrast of Korean data— a voiceless labial stop ([p]), a labial stop followed by the glide [J] ([pJ]); as well as a three-way contrast of Russian data — a voiceless labial stop ([p]), a secondarily palatalized consonant ([p^j]) and a segmental glide [j] ([pj])— are used for the purpose of feature comparison (Suh & Hwang 2016, Kochetov 2006). Five male Russian speakers and five male Korean speakers, who have been living in America for two or more years, were recruited to record the language tokens. Suh & Hwang (2016) collected and analyzed four phonetic features of the Russian and Korean glides: the duration, the F2 slope change "throughout the vocoid (glide + vowel)", the vowel-to-glide coarticulatory effect and the glide-to-vowel coarticulation.

6.1.1 Participants

As in Suh & Hwang (2016)'s experiment, there were 5 male participants in this experiment, who are Mandarin speakers recruited at Stony Brook University, with an average age of 26. All the participants have lived in the USA for two or more years.

6.1.2 Stimuli

In Suh & Hwang (2016)'s experiment, the differences among CV, CjV (segmental glide) and C^jV (palatalization) can be observed in Russian syllables, while Korean syllables only have a two-way contrast: CV and CJV. The token-initial consonant /p/ is chosen because only the voiceless labial stop displays a three-way contrast in Russian (Suh & Hwang, 2016). The Russian nonword tokens in Suh & Hwang (2016)'s experiment follow the template: [p, pj, p^j] + [a, e, o, u] + [də], for instance, "pádə, pjádə, p^jádə"; while the Korean stimuli follow the template: [p, pJ] + [a e o ə u] + [də], for instance, "padə, pJadə". All the bi-syllabic nonword stimuli were presented in not only phonetic transcription but also orthography, in which the first syllable is stressed while the second one [də] is unstressed. The tokens were presented with a carrier sentence "X is not a Russian word" or "X is a nonexistent word" in Korean (Suh & Hwang 2016). They also included around 60-70% fillers mixed with targets in their stimuli.

(76) Attested first syllables of the stimuli

IPA	characters
pJan	篇 "page"
pan	攀 "climb"
pJe	撇 "cast"
pə	颇 "very"
pJɑu	飘 "fly"
pɑu	抛 "throw"
pJəŋ	乒 "table tennis"
pəŋ	烹 "cook"

Like Korean, Mandarin CG sequences do not allow a three-way contrast as in Russian. Following their study, I included Mandarin stimuli consisting of bi-syllabic nonwords (/pJ/ + 4 Mandarin phonemic vowels [a, e, ɑ, ə] + [də]). As displayed in Table (76), the first syllable of each stimuli token is attested in Mandarin. Furthermore, in Table (77), the first syllables of the stimuli are labeled with the first tone which is their attested tone in Mandarin syllables. The second syllable [də] is produced with the first tone as well. The format of the stimuli and fillers is "__不是中国话", which means "__is not a Chinese word."

(77) Stimuli of the phonetics experiment

Mandarin stimuli					
pJV	Orthography (pinyin)	piāndē	piēdē	piāodē	pīngdē
	Surface form (IPA)	pJandə	pJedə	pJɑudə	pJəŋdə
pV	Orthography (pinyin)	pāndē	pōdē	pāodē	pēngdē
	Surface form (IPA)	pandə	pədə	pɑudə	pəŋdə

6.1.3 Recording

The Mandarin stimuli were presented to the participants in both pinyin and phonetic transcriptions. Eight sentences (4 with pJV and 4 with pV) are produced 4 times by each of the 5 Mandarin speakers. Overall, there are 4 sentences * 2 types * 4 times * 5 people =160 Mandarin target tokens, which is the same number as that of the Korean tokens in Suh & Hwang (2016)'s experiment. About 50% of the stimuli are fillers (8 filler tokens * 4 times * 5 people) (see Appendix 2). The total number of tokens is 320 (160 targets + 160 fillers). The filler items are also nonwords which do not include the target sequences /pJ/. The sentences were recorded in a sound-treated room in the Phonetics Lab in the Linguistics Department at Stony Brook University with a Marantz 660 digital recorder at the sampling rate of 44.1 kHz.

6.1.4 Measurements

Following Suh & Hwang (2016)'s study, four values are measured and analyzed via Praat, which are the glide duration, F2 slope change, vowel-to-glide coarticulation and glide-to-vowel coarticulation.

As shown in Figure 24, the glide duration is between the release of the labial stop /p/ and the offset of the vowel. Based on Suh & Hwang's (2016) method, as cited and displayed in Figure 24 below, the release of /p/ (vocoid onset) refers to the "first pulse of the periodic waveform", while the offset of the vowel (vocoid offset) refers to the obvious drop in format energy and intensity. The ratio of GV (glide + vowel) to V (vowel) duration is calculated (Suh & Hwang, 2016). The higher the ratio, the longer the contribution of the glide duration.

F2 slope change is measured following Oh's (2008) and Suh & Hwang's (2016) method. Namely, the F2 values of the glide /J/ and the vowel are measured respectively at the steady point of their format duration. Besides, the midpoint value between them is also measured, which is the F2m value between F2j and F2v shown in Figure 24. Subsequently, the F2 slope change is calculated by subtracting the second slope from the first slope (Suh & Hwang 2016). In other words, the F2 slope change is calculated in the formula: F2 slope change = [(F2m - F2j)/(G-to-V time/2)] -

[(F2v - F2m)/(G-to-V time/2)]. A positive slope change suggests the F2 falling occurs in the early period of glide duration; while a negative one suggests it occurs in the latter part.

With F2j as the y-axis values and F2v as the x-axis values, the slope and the y-intercept are used to analyze the coarticulation effect. The Vowel-to-Glide coarticulation effect is calculated via a linear regression: the locus equation (Suh & Hwang 2016: 92; Lindblom 1963; Krull 1988, 1989; Sussman et al. 1991; Sussman 1994; Iskarous et al. 2010). The closer the slope is to 0 and the closer the y-intercept is to F2j, the smaller the coarticulation effect. In contrast, the closer the slope is to 1, the larger the coarticulation effect, which means the vowel F2 directly influences the glide F2 (Suh & Hwang 2016; Sussman et al. 1991: 1311). Locus equations are calculated from each speaker's recordings, which leads to overall 5 locus equations in the data analyses (5 Mandarin speakers × 1 token type /pJV/).

The Glide-to-Vowel coarticulation effect is represented by the "absolute vowel undershoot value" (Suh & Hwang 2016; Oh 2008: 364-365), which follows the formula: Absolute Vowel *Undershoot Value* = | *F2(vowel onset) – F2(vowel steady state)* |. F2(vowel onset) is measured via a "25ms window right-aligned with the end point of the vocoid duration" (Suh & Hwang 2016: 92). F2(vowel steady state) is the average F2 value of each vowel type. The higher the calculated value, the greater the influence from glide to vowel. Overall, 20 undershoot values are calculated from Mandarin speakers: 5 Mandarin speakers × 4 vowels × 1 token type /pJV/ = 20.

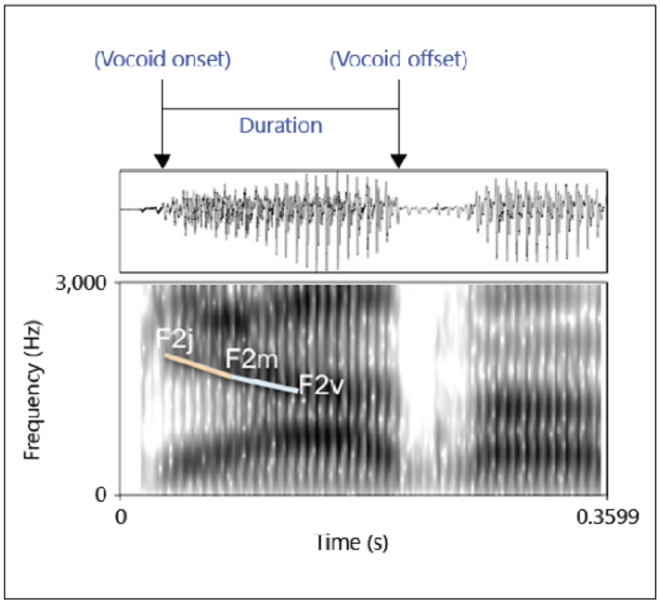


Figure 24. Spectrogram of Korean /pJadə/ (cited from Suh & Hwang, 2016: 91)

6.2 Results

6.2.1 Intrinsic Glide Duration

As shown in Table (78) below, the mean duration of glide [J] + vowel in Chinese tokens is about 152mm, and the mean vowel duration in Chinese tokens is around 141mm. The results from Suh & Hwang's experiment (2016) on Russian and Korean [pJV] indicate that the vocoid duration of syllables with glide was larger than the vocoid duration of [pV] syllables in Russian and Korean. The Russian and Korean data are cited from Suh & Hwang (2016) and displayed in Table (79) below.

(78) Mean vocoid duration and SD of Chinese pJV and pV

Chinese glide [j] duration (ms)				
Participant	pJv(mean)	pJv(SD)	pv(mean)	pv(SD)
p1	198.76	82.2	168.53	87.66
p2	108.1	27.75	103.23	30.06
p3	161.26	68.48	152.14	48.41
p4	126.64	40.59	104.91	44.53
p5	165.63	62.26	177.85	75.58
Mean	152.08		141.33	
No. of tokens	80		80	

(79) Comparison of Durational ratio of p[J]V to pV

language	Korean	Russian j	Russian ^{j}	Chinese
Durational ratio of p[J]V to pV of 5 speakers in each language	1.058709	1.540822	1.041845	1.179375
	1.394281	1.368157	1.010089	1.047079
	1.016583	1.350975	1.116964	1.059945
	1.073062	1.424451	1.007075	1.20713
	1.17054	1.258737	0.975752	0.93129

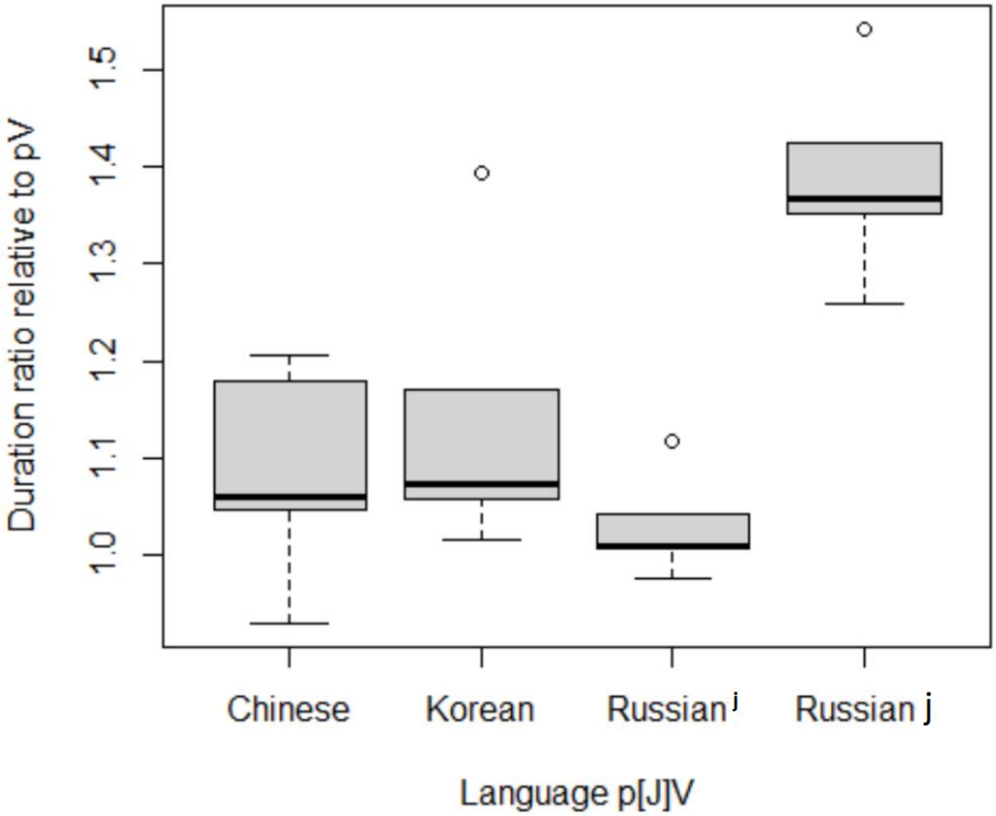


Figure 25. Durational ratio of pJV to pV

In Figure 25, the Russian and Korean data are cited from Suh and Hwang (2016), which shows that that the durational ratio of the JV sequence to the vowel for Russian [pjV] syllables are significantly higher than that of Russian [p^{j}V], Korean [pJV] and Chinese [pJV].

Further ANOVA analyses indicate that Chinese durational ratios are not significantly different from those of Korean tokens (F1,8 = 0.47, p = 0.512) and Russian [p^{j}V] tokens (F1,8 = 0.977, p = 0.352). In comparison, Chinese durational ratios are significantly different from those of Russian [pjV] tokens (F1,8 = 19.9, p = 0.00211). Therefore, the duration of the Chinese glide [J] is not as long as the Russian segmental glide, but similar to the Russian palatalized glide.

6.2.2 F2 Slope Change

The Chinese data of F2 slope change are illustrated in Table (80) and Figure 26. The average F2 slope change for Chinese tokens is 0.2, which is larger than the Russian [p^{j}V] tokens but much smaller than Russian [pjV] ones. Further ANOVA analyses indicate that the F2 slope change of Chinese tokens are significantly different from Russian [pjV] tokens (F1,8 = 31.53, p = 0.000501) and slightly different from Russian [p^{j}V] ones (F1,8 = 4.926, p = 0.0572). Meanwhile, the F2 slope change of Chinese tokens is not significantly different from Korean ones (F1,8=0.867, p =0.379).

(80) Mean F2 slope change and SD by individual speaker

Mean F2 slope change and SD by individual speaker		
CH	mean	SD
p1	-0.13	1.68
p2	0.78	6.37
p3	-1.19	2.09
p4	0.34	5.05
p5	1.2	4.8
mean	0.2	

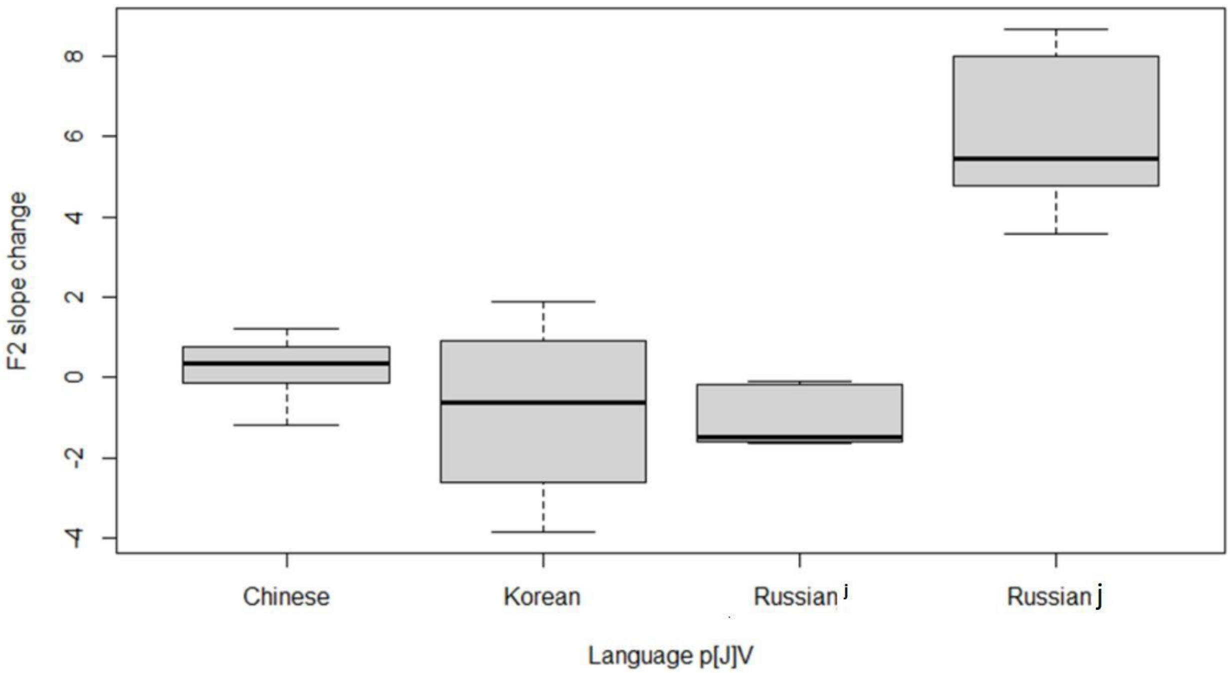


Figure 26. F2 slope change among different languages

6.2.3 Vowel-to-Glide Coarticulation

The locus equation scatterplots of pJV by a Chinese speaker is displayed in Figure 27. The slopes and y-intercepts of Chinese participants are included in Table (81). The results in Suh and Hwang (2016)'s paper show that Russian [p^{j}V] is significantly higher than Russian [pjV] in terms of the locus equation slope, but significantly lower than Russian [pjV] in terms of its y-intercept.

(81) Locus equation slopes and y-intercepts of Chinese participants

Locus equation slopes and y-intercepts of Chinese speakers		
CH pJV	slope	y-intercept
p1	0.078	2045.2
p2	0.59	1050.28
p3	0.51	1299
p4	0.34	1615.38
p5	0.31	1623.94
Mean	0.3656	1526.76

If the slope is close to 0, then there is less coarticulation; meanwhile, a higher slope means more coarticulation (Suh and Hwang 2016). The data in Figure 28 suggests that the locus equation slopes of Chinese tokens are similar to the Russian [p^{j}V] ones. That is to say, the vowel-to-glide coarticulation in Chinese GV sequences are stronger than the Korean ones and the Russian [pjV] ones, yet close to the Russian [p^{j}V] ones. With respect to the y-intercept, Chinese p[J]V are slightly different from both Russian [p^{j}V] and [pjV] tokens, which may suggest that the F2 frequencies of the glide vary among the three types of stimuli.

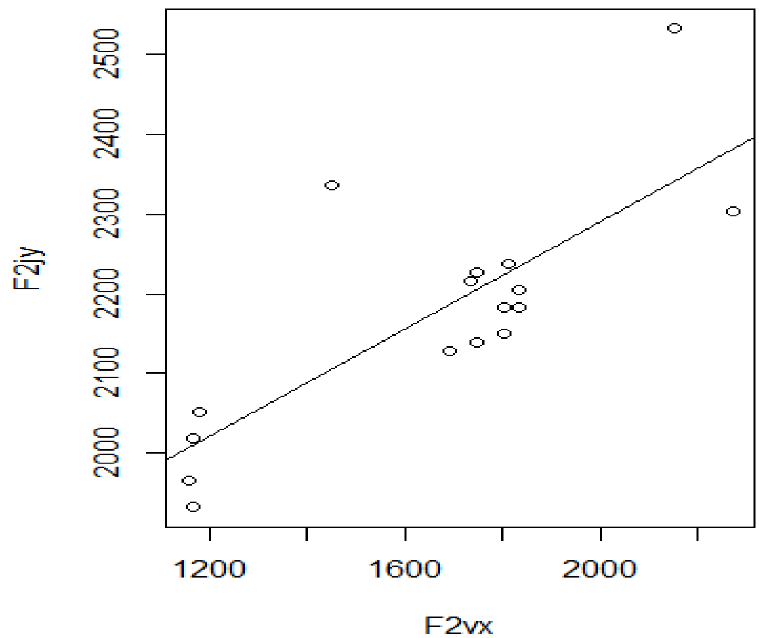


Figure 27. Locus equation scatterplots of pJV by a Chinese speaker

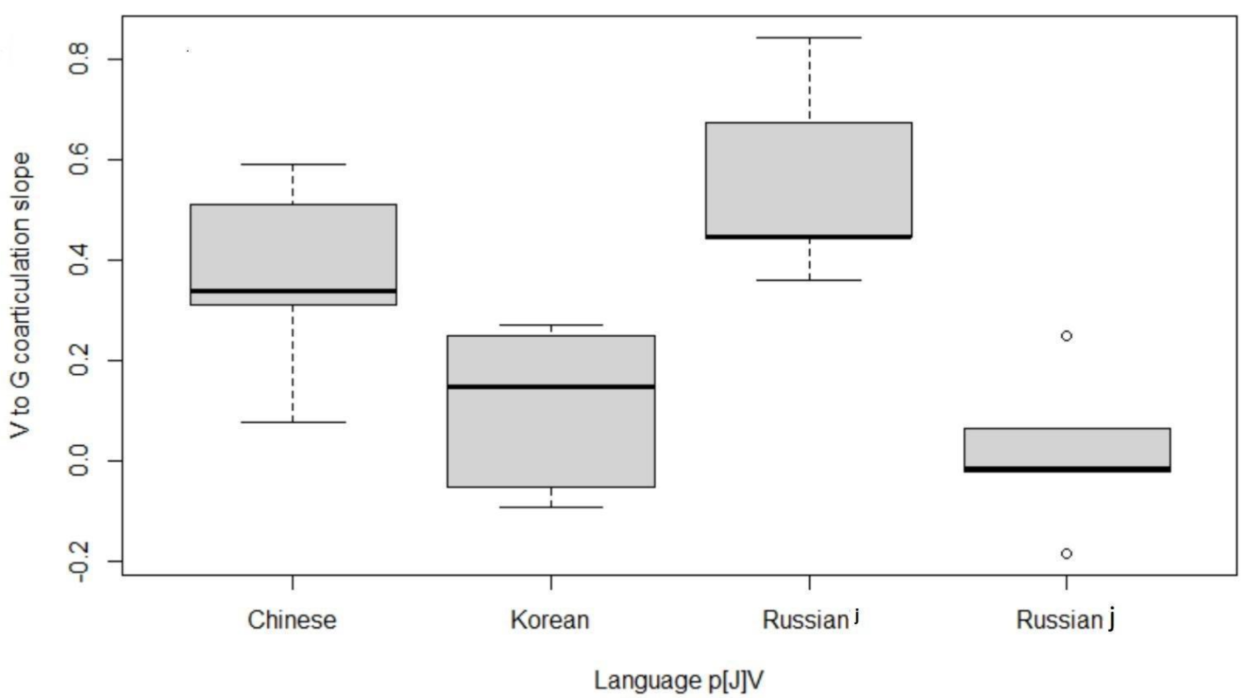


Figure 28. Locus equation slope among different languages

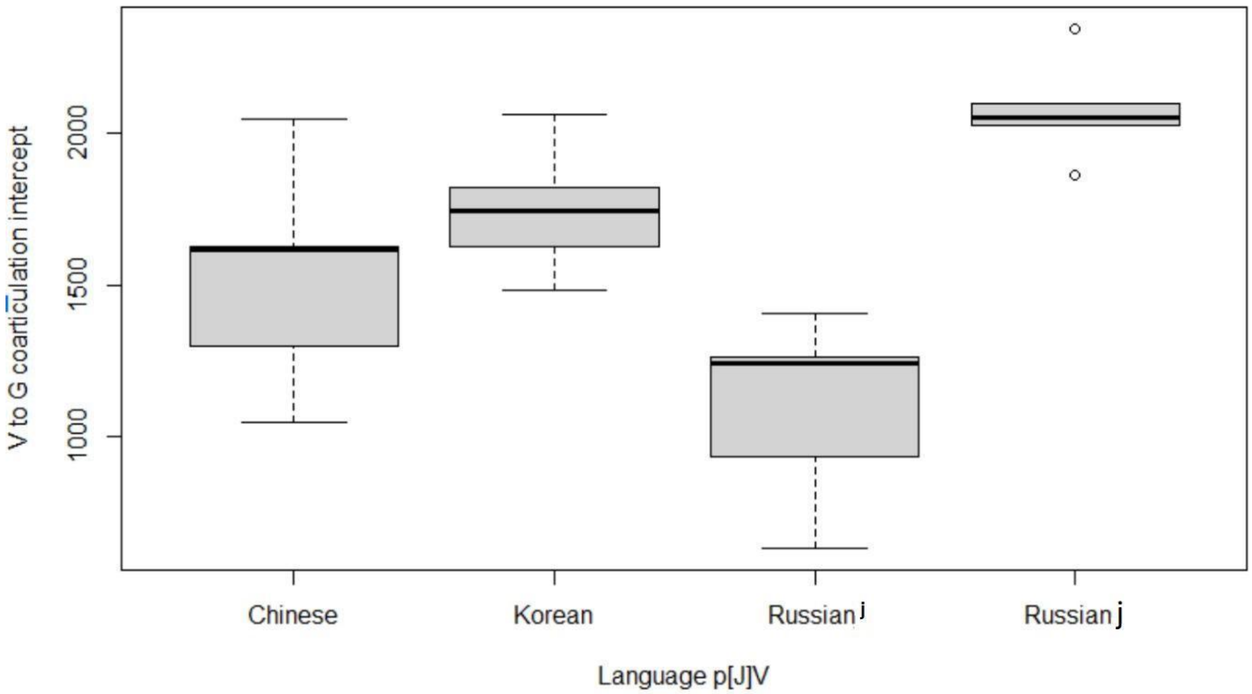


Figure 29. Locus equation intercept among different languages

Further ANOVA analyses show that the slope of Chinese tokens is not significantly different from Russian [pʲV] ones (F1,8 = 2.247, p = 0.172), but are slightly different from Russian [pjV] (F1,8 = 9.29, p = 0.0159) and Korean tokens (F1,8 = 4.978, p = 0.0562). With respect to the y-intercept values, Chinese tokens are not different from the Korean tokens (F1,8 = 1.282, p = 0.29) and slightly different from Russian [pʲV] (F1,8 = 3.916, p =0.0832) and [pjV] tokens (F1,8 =8.746, p =0.0182).

6.2.4 Glide-to-Vowel Coarticulation

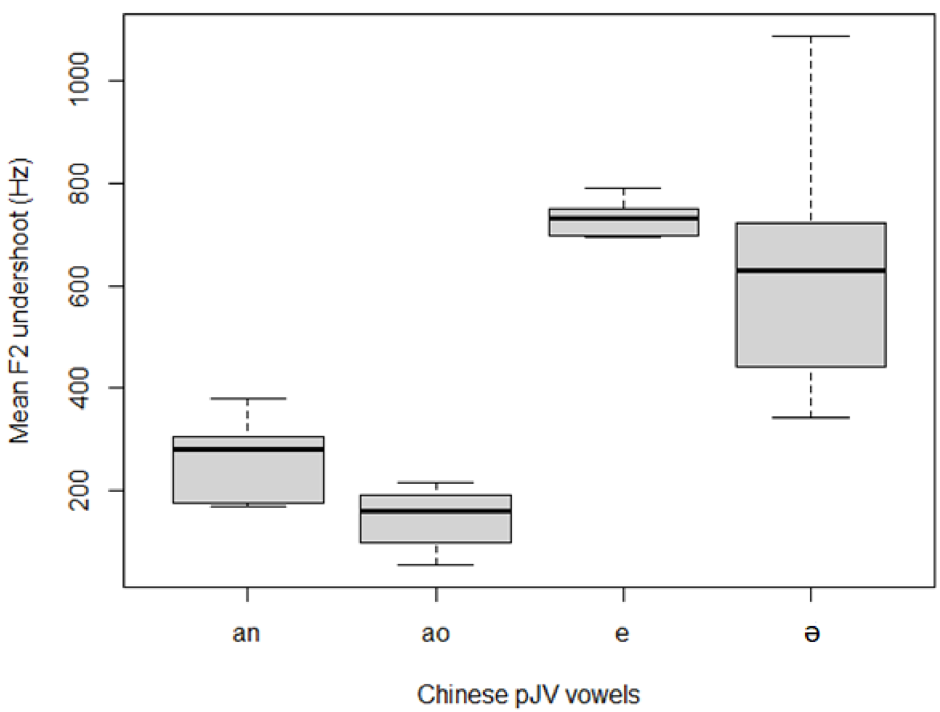
The data in Table (82) show that the mean absolute vowel undershoot value in Chinese stimuli is around 446 Hz. Based on Suh and Hwang's paper (2016), the Russian [pjV] and [pʲV] tokens are not significantly different in their glide-to-vowel coarticulation. That is to say, this parameter cannot be used to distinguish the palatalized [^{j}] and the segmental [j] in the Russian language. Therefore, the glide-to-vowel coarticulation will not be used as a parameter to gauge the structural status of the Mandarin glide [J] in the present study.

Sun and Hwang (2016) found that the effect of vowels on the glide-to-vowel coarticulation is significant in both the Korean and the Russian pJV structures. They (Sun & Hwang 2016: 96-97) also argued that compared to front vowels, the glide-to-vowel coarticulation on the back vowel after the palatal glide [J] were stronger. Similarly, the glide-to-vowel coarticulation in different Mandarin vowels are significantly different ($F(3, 16) = 16.81$, p-value < 0.001).

(82) Absolute vowel undershoot value in Chinese stimuli

F2 undershoot: Mean and SD (Hz)		
CH	mean	SD
p1	407.93	246.69
p2	380.54	242.77
p3	431.93	279.76
p4	469.93	332.38
p5	538.92	461.96
Mean	445.85	

Further pairwise analyses suggested that the glide-to-vowel coarticulation on the vowel [e] and the vowel [ə] are not significantly different from each other ($F(1,8) = 0.448$, p-value = 0.522). Meanwhile, the GV coarticulation on the vowel [a] are slightly different from the vowel [ɑ] ($F(1, 8) = 5.594$, p-value= 0.0456) and the vowel [ə] ($F(1, 8)= 7.977$, p-value= 0.0223). The absolute vowel undershoot value for the four Chinese vowels are illustrated in Figure 30, which indicates that the vowel [ə] is fronted and its F2 value is greatly increased after the glide or palatalized [j]. Meanwhile, the glide-to-vowel coarticulation on the low vowels [a] and [ɑ] is relatively weak. This indicates that the high and mid vowels may be more influenced by the preceding high glide [J], which have a closer articulation position with respect to their tongue height.



a ɑ e ə

Chinese pJV vowels

Figure 30. absolute vowel undershoot value for Chinese vowels

6.3 Discussion

In order to determine whether the Chinese glide [J] is palatalized [^{j}] or segmental [j], this section carried out a phonetic experiment to check the four parameters of Chinese glide [j]: the glide duration, the F2 slope change, vowel-to-glide coarticulation and glide-to-vowel coarticulation. The data collected from the present experiment suggests that the Chinese glide [J] is closer to the Russian palatalized [^{j}] rather than the Russian segmental [j]. Phonetically speaking, Suh & Hwang (2016) argued that compared to the Russian segmental [pj], there is a greater time overlapping between the lip (labial consonant) and the tongue movement (palatal glide) in the Russian palatalized [p^{j}] (see also Kochetov, 2006). In other words, the articulation timing of the initial labial consonant and the palatal glide [J] in Chinese overlap with each other more than the Russian segmental glide [j].

As for the glide duration measurement, the Chinese palatal glide [J] is not significantly different from the Russian palatalized [^{j}] in [p^{j}V] sequences ($p = 0.352$), but much shorter than the Russian segmental glide [j]. Furthermore, the F2 slope change of Chinese glide [J] is much smaller than the Russian segmental glide [j], but only slightly larger from the Russian palatalized [^{j}]. The third measurement is the vowel-to-glide coarticulation, which is indicated via the slope and y-intercept in the locus equations. The slope values relevant to Chinese glide [J] are not significantly different from Russian palatalized [^{j}] (p-value = 0.172) but are slightly lower than Russian segmental glide [j] ($p = 0.0159$). Therefore, the vowel-to-glide coarticulation for both Chinese [J] and Russian palatalized [^{j}] are smaller than Russian segmental [j]. In terms of the other value y-intercept in the locus equations, the Chinese glide [J] are slightly different from both Russian palatalized [^{j}] and Russian segmental [j]. The last measurement is the glide-to-vowel coarticulation based on the absolute vowel undershoot value. This measurement is significantly different among different Mandarin vowels (p-value < 0.001), which is similar to the significant effect of vowels in Russian and Korean (Suh & Hwang 2016). There is a relatively weak glide-to-vowel coarticulation on the low vowels [a] and [ɑ], while it shows that the higher vowels are more influenced by the preceding high glide [J]. This might be related to the fact that the high glide [J] is produced with tongue height closer to high and mid vowels.

6.4 Summary

In this section, I examined the phonetic character of [j] in Mandarin to check whether it is more or less like the palatalized [j] in Russian. In summary, the Chinese glide [J] is not significantly different from Russian palatalized [^{j}] in terms of the first three measurements: glide duration, the F2 slope change and vowel-to-glide coarticulation. The fourth measurement, glide-to-vowel coarticulation, is not employed for comparative analyses in the present study due to its indistinction between Russian [pjV] and [p^{j}V]. Based on the aforementioned comparisons of the Chinese glide [J] with the two Russian glide categories, we can conclude that the Chinese glide [J] shows more similarities to the Russian palatalized [^{j}] rather than the Russian segmental [j].Therefore, the status of the pre-vocalic glide [J] within the syllable is less likely to be segmental (Cj), instead, it shares more commonalities with the palatalized glide as a secondary articulation (C^{j}). While the phonetic character may be evidence for the phonological status of [j], it is not determinative.

Chapter 7

Conclusion

In this study, I concluded that all the systematic constraints of Mandarin syllables are guided by three principles: backness agreement, *HH and *LabLab, among which the backness agreement constraints are anti-OCP, while *HH and *LabLab abide by the OCP. I put forward five original constraints on CG backness agreement (Agree[back]: $C_{[-son, +cont]}$j, Agree[back]:$C_{[-son]}$ɥ and Agree[back]:DorG) as well as CV backness agreement (Agree[back]:$C_{[-son]}$y and Agree[back]: $C_{[hi]}V_{[hi]}$). I generalized both local and distant phonotactic constraints on Mandarin syllables. The local constraints on CG sequences include *CG-LabLab (OCP) and the anti-OCP backness agreement constraints: Agree[ba]:$C_{[-son, +cont]}$j, Agree[ba]: $C_{[-son]}$ɥ, Agree[ba]:DorG. Likewise, the CV sequences follow the OCP constraint *Lab[y] as well as backness agreement constraints: Agree[ba]:$C_{[-son]}$y and Agree[ba]-$C_{[+hi]}V_{[+hi]}$. Both GV and VX sequences follow the OCP constraint *HH. They also abide by the anti-OCP backness agreement: Agree[back]: [w]V, Agree[ba]-[ɥ]V and Agree[ba]: VX. Lastly, the G_X long- distance OCP constraints include *Lab_Lab and *$Hi_{[-ba]}$_$Hi_{[-ba]}$. For Mandarin phonology, the phonotactics are driven by two major principles: (1) the anti-OCP backness agreement; (2) the OCP constraints:*HH and *LabLab.

The acceptability ratings of syllable types including attested syllables with no violations, accidental gaps and systematic gaps are significantly different from each other (all p-values < 0.0001). However, ratings of attested syllables with 1 violation are not significantly different from attested syllables with 0 violation (p-value = 0.93). This shows that Mandarin speakers do not make further distinctions among attested syllables. Real words are regarded as equally acceptable with or without constraint violations. These observed results are consistent with previous studies, where real words are more acceptable than accidental gaps and systematic gaps. The new observation from the current study is that within the scope of real words, 0 constraint violation and 1 constraint violation does not influence the acceptability judgment of Mandarin speakers significantly (p > 0.1). Raters are forgiving of constraint violations as long as they get an attested form. With respect to the unattested words, the phonotactically well-formed ones (accidental gaps) on average received a higher acceptability than the ill-formed ones (systematic gaps) (p<0.01).

With respect to the interaction effects between the 4-level syllable types and other factors including “#violations (PG)”, “weight (PG)”, “#violations (HWPL)” and “Penalty (HWPL)”, a mixed-effects linear regression analysis shows that the interaction between syllable type and "#violations (PG)", as well as the interaction between syllable type and “weight (PG)” were dropped due to non-convergence and lack of significant yielded results. Meanwhile, interactions between syllable types and #violations (HWPL), between syllable types and penalty scores (HWPL) are not significant (all p-values > 0.1).

The present study includes 6 different grammars (3 models * 2 approaches). The two approaches are the manually constructed one based on phonological generalizations and the data-driven one via the Hayes & Wilson Phonotactic Learner. The three models of grammars are categorical grammars assessed based on whether they have a constraint violation or not, cumulative categorical grammars assessed based on the number of violated constraints, and gradient grammars assessed based on the summed weight (PG) and penalty (HWPL). In order to identify the grammar which can best reflect speakers' phonotactic knowledge by best predicting the syllable acceptability ratings from native speakers, I carried out the correlational analyses on the six grammars. Results from Kendall's rank correlation tau indicate that the data-driven gradient grammar (penalty-HWPL) has the strongest correlation with the acceptability rating of Mandarin syllables (tau = -0.38), followed by the two cumulative categorical grammars (#violations-HWPL, #violations-PG) and the manually constructed gradient grammar (weight-PG) with all |tau| values ≥ 0.35. In contrast, the two categorical grammars, especially the data-driven one (tau = -0.19), display a relatively weaker correlation with acceptability ratings of Mandarin syllables. Furthermore, the mixed-effects linear regression analyses on the two categorical grammars also indicate that their effect on acceptability ratings are not significant (all p-value>0.1). Therefore, neither the manually constructed categorical grammar, nor the data-driven categorical grammar, is sufficient to predict the acceptability ratings among Mandarin syllables.

The factors including number of violated constraints (PG), weight (PG), number of violated constraints in Hayes and Wilson Phonotactic Learner (HWPL) and penalty scores (HWPL) are highly correlated with each other (all cor values > 0.80). To explore which factor is the most predictive of the acceptability ratings of nonwords, a multivariate linear regression was carried out for each grammar represented by different factors respectively for comparative analyses.

Based on the linear regression analyses, the effect of all the four factors is not significant in predicting the syllable acceptability (all p-value > 0.1). None of the four grammars can predict the ratings for highly ungrammatical tokens. Now the big take-away is the point beyond which grammars cannot predict the degree of acceptability among ungrammatical syllables. Further pairwise comparison and data binning indicate that: (a) #violations (PG) is inversely related to the acceptability of Mandarin syllables when it's below 2 ($p<0.01$), while syllables with #violations above 2 are not differentiated in that their acceptability are equally poor; (b) the weight (PG) is overall negatively correlated with the obtained acceptability ratings when it is less than 10 ($p<0.01$), but not correlated when the weight is greater than 10; (c) Mandarin syllables with #violations (HWPL) below 10 are overall distinguishable from each other, but not above 10; (d) the acceptability ratings of syllables with a penalty score (HWPL) lower than 10 are significantly higher than other syllables (all p-values < 0.0001), yet undistinguishable for Mandarin speakers when the penalty is above 10 (p-values > 0.1). The observed results indicate that Mandarin speakers are more sensitive to the distinctions among "more grammatical" nonwords. The data show the ratings of the highly ungrammatical items are not distinguishable from each other. Therefore, analyses tapping into effects of syllable well-formedness reveal that highly ungrammatical forms yield comparably low ratings regardless of the grammar model which is used to quantify the extent of syllable ungrammaticality and phonotactic ill-formedness. There is a certain degree of phonological ungrammaticality speakers can tolerate in sound combinations.

Following the investigation on the phonotactic constraints of Mandarin syllables, in the phonetic experiment in §6, I studied the phonetics characteristics of CJV phonotactics. I examined four measurements relevant to the status of the Mandarin palatal glide [J], i.e., the glide duration, the

F2 slop change, the vowel-to-glide coarticulation and the glide-to-vowel coarticulation. The experimental results indicate that the features of Chinese glide [J] are more similar to the Russian palatalized [^{j}] in terms of the first three measurements than the Russian segmental [j]. That is to say, the status of the Mandarin glide [J] is more likely to be palatalized, instead of segmental. The Mandarin syllable template of the glide [J] is likely to be C^{J}VX when the initial consonant is present, instead of CJVX.

Bibliography

Algeo, John. 1978. What Consonant Clusters Are Possible?, Word, 29:3,206-224, DOI: 10.1080/00437956.1978.11435661

Armstrong, S. L., Gleitman, L. R., & Gleitman, H. (1983). What some concepts might not be. *Cognition, 13*(3), 263–308. https://doi.org/10.1016/0010-0277(83)90012-4

Beesley, Kenneth and Lauri Kartunnen. 2003. Finite state morphology. CSLI Publications, Stanford, CA.

Broselow, Ellen & Alice Niyondagara. 1990, Morphological Structure in Kirundi Palatalization: Implications for Feature Geometry, in F. Katamba (ed.), Studies in Inter-lacustrine Bantu Phonology, Afrikanistische Arbeitspapiere, Cologne.

Chandlee, Jane and Jeffrey Heinz. 2018. Strict locality and phonological maps. Linguistic Inquiry, 49(1):23–60.

Chandlee, Jane. 2014. Strictly Local Phonological Processes. Ph.D. thesis, University of Delaware.

Chandlee, Jane. 2017. Computational locality in morphological maps. Morphology, 27:599–641.

Chao, Yuen-Ren. 1934. The non-uniqueness of phonemic solutions of phonetic systems. Bulletin of the Institute of History and Philology, Academia Sinica 4(4): 363–398.

Chao, Yuen Ren. 1968. A Grammar of Spoken Chinese (2nd ed.). University of California Press. ISBN 978-0-520-00219-7.

Charles Douglas Johnson. 1972. Formal Aspects of Phonological Description. The Hague, Mouton.

Cheng, Chin-Chuan. 1973. A synchronic phonology of Mandarin Chinese. Monographs on linguistic analysis no. 4. The Hague: Mouton.

Cheng, Chin-Chuan. 1973. A synchronic phonology of Mandarin Chinese. Monographs on linguistic analysis no. 4. The Hague: Mouton.

Chomsky, Noam. 1956. Three models for the description of language. IRE Transactions on Information Theory, 2:113–124.

Clements, George N. 1985. The geometry of phonological features, Phonology 2, 225-252.

Coetzee, Andries W.. 2009. Grammar is both categorical and gradient. Retrieved from https://doi.org/doi:10.7282/T3D50JZQ

Dong, Shaowen. 1958. Yuyin changtan [Introduction to phonetics]. Beijing: Wenhua Jiaoyu Chubanshe.

Duanmu, S., & Yi, L. 2015. "Phonemes, Features, and Syllables: Converting Onset and Rime Inventories to Consonants and Vowels." Language and Linguistics 16(6), 819–842.

Duanmu, San. 1990. A formal study of syllable, tone, stress and domain in Chinese languages. Doctoral dissertation. Cambridge: MIT.

Duanmu, San. 2000. The Phonology of Standard Chinese. Oxford: Oxford University Press.

Duanmu, San. 2007. The Phonology of Standard Chinese. Oxford: Oxford University Press.

Duanmu, San. 2008. Syllable structure: The limits of variation. Oxford: Oxford University Press.

Durvasula, Karthik, 2020. O gradience, whence do you come? Talk at the Annual Meeting on Phonology (AMP).

Fu, Maoji. 1956. 'Beijing hua de yinwei he pinyin zimu' [Phonemes and Pinyin symbols in the Beijing Speech], Zhongguo Yuwen, 1956.5: 3-12.

Gerhard Jager and James Rogers. 2012. Formal language theory: Refining the chomsky hierarchy. Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1598):1956–1970.

Gong, S. (2018). Grammaticality and lexical statistics in Chinese unnatural phonotactics. UCL Working Papers in Linguistics, Vol. 30.

Gong, S. & Zhang, J. 2019. "The Gradient Acceptability in Mandarin Nonword Judgment." Proceedings of the 2019 Annual Meeting on Phonology. Stony Brook University, US. https://journals.linguisticsociety.org/proceedings/index.php/amphonology/article/view/4660

Gong, S. & Zhang, J. 2021. Modeling Mandarin speakers' phonotactic knowledge. Phonology, 38(2), 241-275.

Gorman, Kyle, 2013. Generative phonotactics. University of Pennsylvania dissertation.

Hayes, Bruce. & Colin Wilson. 2007. A Maximum Entropy Model of Phonotactics and Phonotactic Learning

Hayes, Bruce. 2011. Introductory Phonology. Hoboken, NJ: Wiley-Blackwell.

Hayes, Bruce. & Colin Wilson. 2008. A Maximum Entropy Model of Phonotactics and Phonotactic Learning. Linguistic Inquiry, 39: 3, 379-440.

Hopcroft, John E., Rajeev Motwani, and Jeffrey D. Ullman. 2006. Introduction to Automata Theory, Languages, and Computation (3rd Edition). Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

Hume, Elizabeth. 1990. 'Front Vowels, Palatal Consonants and the Rule of Umlaut inKorean', Proceedings of NELS 20, GLSA, Department of Linguistics, University of Massachusetts, Amherst, pp. 230-243.

Hume, Elizabeth. 1994. Vowels, Coronal Consonants and their Interaction in Nonlinear Phonology, Garland Publishing, New York.

Hume, Elizabeth. 1996. Coronal consonant, front vowel parallels in Maltese. Nat Lang Linguist Theory 14: 163-203. https://doi.org/10.1007/BF00133405

Kaplan, Ronald M. and Martin Kay. 1994. Regular models of phonological rule systems. Computational Linguistics, 20(3):331–378.

Karlgren, Bernhard. 1915-1926, 2003. 'Zhongguo yinyun xue yanjiu' [Etudes sur la Phonologie Chinoise], Beijing: Shangwu yinshuguan [高本漢, 2003,《中國音韻學研究》, 北京: 商務 印書館]

Kenstowicz, Michael. 1994. Phonology in generative grammar. Cambridge, Mass. & Oxford: Blackwell. Pp. ix + 704.

Kochetov, Alexei. 2016. Palatalization and glide strengthening as competing repair strategies: Evidence from Kirundi. Glossa: a journal of general linguistics 1(1): 14. 1–31, DOI:

http://dx.doi.org/10.5334/gjgl.32
Koskenniemi, Kimmo. 1983. Two-Level Morphology: A general computational model for word-form recognition and production. Ph.D. thesis, University of Helsinki.

Ladefoged P, Maddieson I (1996): The Sounds of the World's Languages. Cambridge, Blackwell.

Lawson, Mark V. 2003. Finite Automata. Taylor & Francis.

Lin, Yen-Hwei. 1989. Autosegmental treatment of segmental processes in Chinese phonology. Doctoral dissertation, University of Texas, Austin.

Luo, Changpei & Wang, Jun. 1981. Putong Yuyinxue Gangyao [Outline of general phonetics], new edn. Beijing Shangwu Yinshuguan. First pub. Beijing: Kexue Chubanshe, 1957.

McCarthy, John J. 1988. Feature geometry and dependency: A review. Phonetica. 38.Meeussen, Achille E. 1959. Essai de grammaire rundi. Tervuren: Musée Royal de l'Afrique Centrale.

Myers, J. 2002. An analogical approach to the Mandarin syllabary. Journal of Chinese Phonology, 11(Special Issue), 163–190.

Myers, J., & Tsay, J. 2005. The processing of phonological acceptability judgments. In Proceedings of symposium on 90-92 NSC projects (pp. 26–45).

Myers, J., & Tsay, J. 2015. Mandarin Wordlikeness Project [Raw data]. Retrieved from http://lngproc.ccu.edu.tw/MWP/index.html

Ntihirageza, Jeanine. 1993. Kirundi palatalization process and sibilant harmony: Implications for Feature Geometry. Carbondale, IL: The Southern Illinois University at Carbondale MA thesis.

Odden, David. 1991. Vowel Geometry. Phonology, 8(2), 261-289. Retrieved from http://www.jstor.org/stable/4420037

Oh E (2008): Coarticulation in non-native speakers of English and French: an acoustic study. J Phonet 36:361–384. Padgett J (2003): Contrast and post-velar fronting in Russian. Nat Lang Ling Theory 21:39–87.

Pulleyblank, Edwin G. 1983. Vowelless Chinese? in Chan, ed. 1984. pp. 568-619.

Riggle, Jason. 2011. Phonological Feature Chart – University of Chicago version 11.02. http://hum.uchicago.edu/~jriggle/

Roark, Brian and Sproat, Richard. 2007. Computational approaches to syntax and morphpology. Oxford University Press, Oxford.

Rogers, James, Jeffrey Heinz, Gil Bailey, Matt Edlefsen, Molly Visscher, David Wellcome, and Sean Wibel. 2010. On languages piecewise testable in the strict sense. In The Mathematics of Language, volume 6149 of Lecture Notes in Artifical Intelligence, pages 255–265. Springer.

Rogers, James, Jeffrey Heinz, Margaret Fero, Jeremy Hurst, Dakotah Lambert, and Sean Wibel. 2013. Cognitive and Sub-regular Complexity, chapter Formal Grammar. Springer.

Sagey, Elizabeth. 1986. The representation of features and relations in nonlinear phonology. PhD dissertation, MIT.

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. Journal of Memory and Language, 110. https://doi.org/10.1016/j.jml.2019.104038

Schutzenberger, Marcel-Paul. 1961. A remark on finite transducers. ¨ Information and Control, 4:185– 196.

Suh, Yunju & Hwang, Jiwon. 2016. The Korean Prevocalic Palatal Glide: A Comparison with the Russian Glide and Palatalization. Phonetica. 73. 85-100. 10.1159/000444189.

Tsai, C.-H. (2000). Mandarin Syllable Frequency Counts for Chinese Characters. Retrieved from http://www.gnu.org/copyleft/gpl.html

Wang, Hongjun 1999. Hanyu Fei Xianxing Yinxixue [Chinese Non-linear Phonology]. Beijing: Beijing Daxue Chubanshe.

Wang, Jenny Zhijie. 1993. The Geometry of Segmental Features in Beijing Mandarin. Doctoral dissertation, University of Delaware, Newark, DE.

Wang, Jenny Zhijie. 1999. Vowel Length in Cantonese. Paper presented at the 11th North American Conference on Chinese Linguistics, Harvard University.

Wang, Li. 1980. Hanyu Yinyun [Chinese Phonology], 2nd edn. Beijing: Zhonghua Shuju.

Xinhua Mandarin dictionary. https://zidian.aies.cn/pinyin_index.htm

Xu, Shirong. 1980. Putonghua yuyin zhishi [Phonology of Standard Chinese]. Beijing: Wenzi

Gaige Chubanshe.
Yang, C. (2005). On Productivity. Linguistic Variation Yearbook, 5(1):265–302

Yip, Moira. 2003. Casting doubt on the Onset–Rime distinction, Lingua, Volume 113, Issue 8, Pages 779-816, ISSN 0024-3841, https://doi.org/10.1016/S0024-3841(02)00130-4.

Appendices

Appendix (i) **Distinctive Feature Chart of Mandarin Consonants**

	p	p^h	m	f	t	t^h	n	l	ts	ts^h	s	tʂ	tʂ^h	ʂ	ʐ	tɕ	tɕ^h	ɕ	k	k^h	x	ŋ
sonorant	-	-	+	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	+
continuant	-	-	-	+	-	-	-	+	±	±	+	±	±	+	+	±	±	+	-	-	+	-
LABIAL	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CORONAL	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
anterior	0	0	0	0	+	+	+	+	+	+	+	-	-	-	-	+	+	+	0	0	0	0
distributed	0	0	0	0	-	-	-	-	-	-	-	-	-	-	-	+	+	+	0	0	0	0
DORSAL	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+
high	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+	+	+	+	+	+	+
low	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	-	-	-	-	-	-
front	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+	+	+	-	-	-	-
back	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	-	-	+	+	+	+

Appendix (ii) Distinctive Feature Chart of Mandarin Vowels

	a	e	o	ə	ɑ	i	u	y
LABIAL	-	-	+	-	-	-	+	+
round	0	0	+	0	0	0	+	+
CORONAL	-	-	-	-	-	-	-	-
DORSAL	+	+	+	+	+	+	+	+
high	-	-	-	-	-	+	+	+
low	+	-	-	-	+	-	-	-
back		-	+		+	-	+	-

Appendix (iii) Distinctive Feature Chart of Mandarin Glides

	j	w	ɥ
sonorant	+	+	+
continuant	+	+	+
LABIAL	-	+	+
round	-	+	+
CORONAL	-	-	-
anterior	0	0	0
distributed	0	0	0
DORSAL	+	+	+
high	+	+	+
low	-	-	-
front	+	-	+
back	-	+	-

Appendix (iv) Phonetic experiment Fillers: ___不是中国话。(“___ is not Chinese word.”)

Mandarin filler tokens				
Orthography (pinyin)	kuādē	kuāndē	kuīdē	kuāngdē
Surface form (IPA)	kwadə	kwandə	kwəidə	kwɑŋdə
Orthography (pinyin)	jüēdē	jüāndē	qüēdē	qüāndē
Surface form (IPA)	tɕɥedə	tɕɥendə	t^hɕɥedə	t^hɕɥendə