Grammaticality and Lexical Statistics in Chinese Unnatural Phonotactics[*]

Donald Shuxiao Gong

Abstract

Speakers possess phonotactic knowledge about the acceptability of non-words, yet the source of this knowledge is not clear. Statistical analyses on a Chinese non-word judgement experiment have shown that both lexical statistics (neighbourhood density and Phonotactic Learner) and grammaticality (phonetically principled phonotactic constraints) play a role in predicting speakers' behaviours. The results also indicate that a specific phonotactic constraint which bans Labial fricative to co-occur with Coronal glide (*[fj]) is well learned and noticed by speakers, despite its phonetic unnaturalness, suggesting that unnatural patterns are not necessarily dispreferred and underlearned.
Keywords: Phonotactics, Lexical Statistics, Naturalness, Neighbourhood Density, Chinese

1 Introduction

Segments do not randomly combine with each other to form a word. There are certain constraints which place restrictions on the ability of certain types of segments to occur next to one another, which is known as phonotactics. A range of studies has confirmed that language speakers possess such phonotactic knowledge, because they offer gradient acceptability to non-words (Albright, 2009; A. Coetzee, 2008; Frisch & Zawaydeh, 2001; Hay, Pierrehumbert, & Beckman, 2003; Hayes & White, 2013). For example, although all missing from the real lexicon, English speakers will most probably judge *blick* as a highly possible word, while *bnick* is impossible, and *lbick* as completely unacceptable. What is less clear is the source of this phonotactic knowledge. One interpretation would be that speakers' performance is guided by phonological grammar. To use the above example again, *lbick* is unacceptable because its onset cluster violates Sonority Sequencing Principle, a general preference of large rises in sonority from syllable edge to nucleus. Another theory contributes phonotactic knowledge to usage-based frequencies, i.e., lexical statistics. In this case, *bnick* and *lbick* are dispreferred because the onset sequence *bn* and *lb* are not well supported by the lexicon (in fact, the probability for *bn* and *lb* to occur word-initially is zero). Sometimes it is hard to disentangle these two, as ungrammatical forms may also lack lexical attestedness, and vice versa. A combinatory view would say that phonotactic knowledge is influenced by lexical statistics, while grammar still plays independent role (A. Coetzee, 2008). This paper uses data from Chinese to test these possibilities.

Chinese languages are often claimed to have a simple and restrictive syllable structure, and the most widely accepted model is CGVX (Duanmu, 2007). Despite the fact that Chinese is very strict about what sounds can occur with what, one puzzling question still remains: why there are so many theoretically possible syllables missing from the real inventory? According to Duanmu (2011)'s statistics, in Standard Chinese (SC), among all 1,900 possible syllables, only 404 are found; 79% of the syllable positions are missing. Many attempts have been made

to explain this large number of syllable gaps (Duanmu, 1990, 2003, 2007, 2008; Lin, 1989; Myers, 1995; Wiese, 1997), mainly based on feature co-occurrence or feature harmony. These attempts can be known as natural phonotactic constraints, as they all fall into some phonetic principles such as similarity avoidance (dissimilation). Syllables ruled out by such principled phonotactic constraints are classified as systematic gaps, whereas remaining missing forms are called accidental gaps (Halle, 1962). However, languages often present phonotactic constraints which will strike phonologists as being unnatural, or making no phonetic sense. For example, in SC, all labials can combine with the glide [j] except [f], that is, [pj] [pʰj] [mj] but *[fj]. This disfavour of [f] following a high front vocoid is also found in other Chinese languages such as Wu, Cantonese, Hakka, Xiang, and Gan. This constraint is considered as unnatural, or phonetically unmotivated at least, because there is no sensible phonetic grounding. There might be co-occurrence constraint on Labial onset and Labial glide, as the feature [Lab] cannot occur twice in sequence. However, a Labial onset is expected to freely host a Coronal glide [j], as [fj] does occur in many languages such as English. This paper will discuss speaker's distinct performance regarding natural vs. unnatural phonotactic constraints.

Many theories believe that grammar is biased towards phonetic naturalness. One strong argument comes from Becker et al. (2007, 2011), showing that Turkish speakers are unable to generalise certain unnatural patterns exhibited in their lexicon. Turkish laryngeal alternation can partly be predicted by the place of articulation of the stem-final stop, by word-length, and by the preceding vowel quality. However, in novel word tasks, the productivity of this laryngeal alternation is only affected by place of articulation and word-length. Speakers tend to ignore the vowel quality effect, even though this information is provided by the lexicon. This finding is questioned by Hayes et al. (2009), claiming that an unnatural vowel harmony pattern in Turkish is learnable, but dispreferred. They in turn propose a biased Universal Grammar view: leaning a natural pattern is easier than an unnatural one (Hayes & White, 2013; Pater & Tessier, 2005; Pycha, Nowak, Shin, & Shosted, 2003; Wilson, 2003, 2006). Under this view, UG is linked with phonetic naturalness, and forms that violate natural phonotactic constraints are predicted to receive lower scores than those violating unnatural ones in non-word acceptability judgements. Quantitatively, the number of systematic gaps largely exceeds the accidental ones in Chinese syllabaries. However, intuitively, some accidental gaps mentioned above are in fact much less 'accidental' than they should be. First, they are not necessarily judged as more acceptable than systematic gaps (Myers, 2002). Second, those unnatural phonotactic constraints may recur in various Chinese languages. Why do such unnatural phonotactics exists?

Although a widely held assumption is that phonology should be phonetically grounded, synchronic phonology does allow a large number of unnatural patterns to occur (Anderson, 1981; Blevins, 2004). Contrary to the belief that phonology is shaped and determined by phonetic substance, Evolutionary Phonology leaves all such phonetic information to diachronic sound change; synchronic phonologies cannot refer to any phonetic details (Blevins, 2004). Many synchronic phonological rules make no phonetic sense, labelled as 'crazy rules' (Bach & Harms, 1972). Johnsen (2012) uses diachronic model to explain a synchronically unnatural retroflection pattern in Norwegian. A similar method is also adopted on a synchronic 'crazy' yet very productive rule of Sardinian (Scheer, 2015). The surface [l] → [ʁ] alternation is completely unnatural. Nevertheless, based on diachronic evidence and dialectal variation, this rule can be broken down as [l] → [ɫ] → [w] → [gʷ] → [ɣʷ] → [ʁ], with each step being grounded by phonetics. Phonotactic constraints which operate against phonetic naturalness are found in languages such as Tswana, Tarma Quechua and Berawan, where the phonetically more marked voiceless rather than voiced stops occur after a nasal or intervocalically, showing a post-nasal devoicing effect (Begus & Nazarov, 2017; A. W. Coetzee & Pretorius, 2010; Hyman, 2001). Kawahara (2008) also reports that in Japanese,

natural and unnatural patterns can co-occur in a single module of phonological grammar. Thus, synchronic unnaturalness can be explained by diachronic rule telescoping, and has a chance to enter speaker's grammar. In some cases, unnatural constraints may even be over-phonologised by speakers. Kager & Pater (2012) find that the Dutch lexicon contains very few sequences of a long vowel followed by a consonant cluster whose second member is a non-coronal. This pattern is too complicated to be phonetically natural, yet speakers presented special sensitivity to it in well-formedness judgement tasks. Kirby & Yu (2007)'s well-formedness judgement data show that Cantonese speakers marked the nonce forms violating an unnatural Consonant-Vowel Coronal co-occurrence condition as less acceptable than forms violating a much more natural Labial co-occurrence restriction on Onsets and Codas. It seems that unnatural patterns can be productively internalised by native speakers, at least on an inductive basis.

This paper discusses the role of unnatural phonotactic constraints in Standard Chinese. Are they simply unimportant lexical accidents (badly supported by lexical statistics), or are they implicitly known by speakers and thus belong to the grammar? Section 2 discusses the restrictive syllabary of Standard Chinese and its natural and unnatural phonotactic constraints. Section 3 reviews the current debate on the source of natural vs. unnatural patterns. Section 4 presents statistical analyses of the data extracted from a huge Chinese non-word judgement experiment conducted by Myers & Tsay (2015), using lexical frequency and grammaticality parameters to predict participants' reaction time and lexical decisions. Section 0 offers a discussion about the results, showing that besides lexical statistics, formal grammaticality has independent influence on linguistic performance. More importantly, an unnatural phonotactic constraint exhibits similar behaviour as other natural ones. Finally, Section 6 concludes.

2 Chinese Phonotactics
2.1 Systematic Gaps vs. Accidental Gaps

Due to a highly restrictive CGVX syllable structure (Duanmu, 2007), it is feasible to work out a list of theoretically possible syllables by factorial combination of all phonemes of the four syllable constituents. Here C stands for onset consonant, G for glide, V for nuclear vowel, and X for either an off-glide or a nasal coda. Taking SC as an example, the phoneme inventory of SC is shown below:

(1)   Standard Chinese Phoneme Inventory
    C: p pʰ m f t tʰ n l ts tsʰ s tʂ tʂʰ ʂ ʐ k kʰ x
    G: j w ɥ
    V: a ə i u y
    X: i u n ŋ

In CGVX template, only the nuclear vowel V is obligatory, whereas other three constituents are optional. This gives rise to $(18+1) \times (3+1) \times 5 \times (4+1) = 1,900$ theoretically possible syllables in SC, among which only 404 syllables actually occur. Syllabic consonant rhymes such as [ʐ̩ n̩] and other unproductive rhymes like [ɚ] will be ignored.

Also, if we take tonal contrasts into consideration, there will be $1,900 \times 4 = 7,600$ distinct forms. However, I omit the discussion of tonal distinction here due to the following reasons. First, experimental results suggest that tonal accidental gaps are judged much more acceptable than other segmental gaps (Kirby & Yu, 2007; Myers, 2002; H. S. Wang, 1998). Second, other psycholinguistic studies show that tones are less important than previously assumed in language processing (e.g. Taft & Chen, 1992; Ye & Connine, 1999). For example,

when presented with non-words and asked to change them into real words by substituting a vowel, a consonant, or its lexical tone, speakers' responses show that changes to vowels and consonants were more detrimental than tonal change, suggesting that tonal information is of lower priority than consonants and vowels (Wiener & Turnbull, 2016). Third, many tonal gaps are merely due to historical sound change. They are ubiquitous in other dialects and can be easily filled in by loan words, neologism, syllable combination in fast speech, and onomatopoetic words. Fourth, as we will see, tonal gaps often will not violate the principle phonotactic constraints we propose, if well defined. Tonal features are perceptually less salient than other segmental features. Thus, to avoid too much elaboration, tonal contrasts will not be considered.

Comparing theoretically possible syllables with actual existing syllables, we obtain a list of missing syllables. As mentioned in Section 1, the number of missing syllables is extremely high (1,496 for SC), and raises much theoretical attention on how to explain these missing forms. If we look at the syllable table, some patterns of those missing syllables can be immediately determined. For example, labial consonants [p pʰ m f] cannot co-occur with a labial glide [w], which leaves a huge blank in the syllable table. In the meantime, there are some sporadic, accidental gaps, such as *[fai] and [ʐ̩ai]: the rhyme [ai] can co-occur with all other 16 onsets except for [f] and [ʐ]. However, there is no consensus on how to treat these two types of syllable gaps, not only for SC, but also for other languages in general. Some believe that there is categorical distinction between the two, while others do not draw a clear line (Boersma & Hayes, 2001). However, much effort still has been made to figure out the phonological constraints on missing syllables, as stated in Coetzee (2008):

> What is required is a way to distinguish between accidental lexical gaps and principled lexical gaps. It is not easy to formalize a principle that can be used for this purpose. However, as a first approximation I suggest the following: lexical gaps that can be stated in terms of general phonological properties should be treated as principled.

Similar arguments are provided in Frisch & Zawaydeh (2001). Indeed, efforts searching for phonotactic constraints turn out to be very successful, at least in Chinese languages. For example, Duanmu & Yi (2015) use only four constraints to explain 90% of the missing syllables in Lanzhou Mandarin. However, one must remain wary about the term "general phonological properties", as it may still seem ambiguous and uncertain. Therefore, a clearer definition on the distinction between systematic gaps and accidental gaps is needed, yet we cannot ignore the descriptive power of the phonotactic constraints proposed for various languages.

Following Coetzee (2008)'s proposal, I use Standard Chinese as an example to show how natural (principled) phonotactic constraints help to distinguish systematic gaps and accidental gaps. Kessler & Treiman (1997) find that in English, onset-vowel combinations are statistically random, while vowel-coda sequences may be either underrepresented or overrepresented. This provides evidence for rhyme as a concrete syllable constituent, instead of nucleus and coda being independent units. Sinologists employ a traditional initial-final analysis for the study of Chinese phonology. Under this traditional method, glide is grouped as a rhyme constituent instead of onset (H Samuel Wang & Chang, 2001). The phonological system of *fanqie* was based on grouping glides with rhymes. Basically, *fanqie* method was a guideline to show the pronunciation of Chinese characters in Middle Chinese, widely used in ancient Chinese phonology books. It uses two characters to represent the pronunciation of a syllable for a new character. For example, the syllable [duŋ] is represented by [dɤ] + [huŋ], taking the onset of the first syllable and the rhyme of the second. When it comes to syllables

with prenuclear glides, the glide is usually parsed with the second syllable, meaning that the glide is not part of the onset. For instance, [du] + [njɛn] will give the pronunciation of [djɛn].

Here we adopt a similar method, by first considering missing rhymes, i.e., missing GVX forms, and then looking at missing onset-rhyme combinations, i.e., missing syllables.[1] One missing rhyme is equivalent to 19 missing syllables, since each missing rhyme indicates that it cannot be combined with any onset consonant.

Duanmu (2007) proposes three constraints, namely Rhyme Harmony, Merge, and G-Spreading, to account for the missing GVX forms in SC.

(2)   Duanmu (2007)'s constraints
     a.   Rhyme Harmony: VX cannot have opposite values in [round] or [back], e.g. *[yi], *[yu]
     b.   Merge: Two tokens of the same feature merge into one long feature, e.g. [ii] = [i]
     c.   G-Spreading: A high nuclear vowel spreads to the glide G, e.g. [u] = [wu]

However, this proposal is not satisfactory. First, some good forms, such as [iŋ] [un], are mistakenly ruled out, because he analyses [ŋ] and [n] as [+back] and [-back]. Second, this analysis leaves too many gaps as accidental: 23 GVX forms in number. The current analysis also proposes three constraints, adapted from Lin (1989) and Duanmu & Yi (2015).

(3)   Standard Chinese Rhyme Constraints
     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.

These three constraints can rule out most missing GVX forms, leaving only seven as accidental gaps: /ɥa ɥaŋ ɥən ɥəŋ un jən jəŋ/.[2]

As for Onset-Rhyme combinations, one simple constraint can explain most missing forms.

(4)   Articulator Dissimilation:
           C and G must have different articulators.

As a result, apart from the seven missing rhymes, the following syllables are considered as accidental gaps in SC.

(5)   Accidental Syllable Gaps in Standard Chinese
           ʐa   ʐai   fai   fau   pəu   pʰəu   nə
           əi   təi   tʰəi   tsʰəi tʂəi   tʂʰəi   tən   tʰən   nən   lən   əŋ
           pja   pʰja   mja   pjaŋ   pʰjaŋ mjaŋ   pjəu   pʰjəu   mjəu
           fi   fin   fiŋ   fja   fjan   fjaŋ   fjau   fjə   fjəu
           twa   tʰwa   nwa   lwa   tswa   tsʰwa   swa   tʂʰwa   ʐwa
           twai tʰwai nwai   lwai   tswai tsʰwai swai   ʐwai

---

twaŋ	tʰwaŋ	nwaŋ	lwaŋ	tswaŋ	tsʰwaŋ	swaŋ	ʐwaŋ
nwəi	lwəi	nwən	ʂuŋ
ty	tʰy	tɥan	tʰɥan	nɥan	lɥan	tɥə	tʰɥə	tyn	tʰyn	nyn
lyn	tyŋ	tʰyŋ	nyŋ	lyŋ

This analysis attributes all systematic gaps to Obligatory Contour Principle (Leben, 1973; McCarthy, 1986), in which identical features or segments are not permitted to occur in sequence. This similarity avoidance effect is also reported in Arabic (Frisch, Pierrehumbert, & Broe, 2004; Frisch & Zawaydeh, 2001), Muna (A. W. Coetzee & Pater, 2008), and many other languages.

        After a quick scan, we notice that many gaps contain a [y] or its glide counterpart [ɥ], which seems to reflect the fact that [y] is an uncommon sound cross-linguistically. Among all 1,900 theoretically possible syllables, 404 are actually found, leaving 1,496 missing forms. After applying four phonotactic constraints, $7 \times 19 + 81 = 214$ missing forms remain unaccounted for. That is, 86% of the missing syllables are marked as systematic gaps, while the remaining 14% are accidental gaps. Given the fact that all four constraints proposed here are derived from general phonological properties, this phonotactic analysis deals with SC very nicely.

2.2 An Unnatural Phonotactic Constraint

Apart from these systematic gaps ruled out by phonetically driven OCP constraints, there are a few lexical gaps that behave not as accidental as we would expect. Let's focus on one of them. In SC and many other Chinese languages, labial onsets are free to combine with the coronal glide [j], except the labial fricative [f], as shown in (6).

(6)   *[fj] in Standard Chinese
    a.   [pjan]   变   'change'   b.   [pjau]   表   'watch'
    c.   [pʰjan]   片   'slice'   d.   [pʰjau]   票   'ticket'
    e.   [mjan]   面   'noodle'   f.   [mjau]   秒   'second'
    g.   *[fjan]   h.   *[fjau]

Synchronically, it is hard to find a phonetic motivation for this phonotactic constraint. However, unnatural processes can often be explained by sequential natural sound changes (Begus & Nazarov, 2017). Therefore, a historical account for *[fj] is in order. According to linguistic reconstruction, all contemporary onset [f] is derived from historical Mid Chinese *[pj], *[pʰj], *[bj] sequences by a labiodentalisation rule roughly stated in (7).

(7)   Mid Chinese Labiodentalisation[3]
           p pʰ b → f / _ j V[+back]

The sequence [fj] lacks historical source during the evolutionary history of Chinese languages. In a sense, if the diachronic trace is still transparent to synchronic grammar, contemporary [f] contains [j] already and therefore does not permit another [j] to follow. As Evolutionary Phonology (Blevins, 2004) would claim, the rule of (7) is natural, yet its result on synchronic grammar is unnatural: a *[fj] constraint. However, contemporary speakers are unlikely to have access to this sound change hundreds of years ago, and still apply that in the

---

[3] Exact applying condition for this rule is still a debatable topic in historical Chinese phonology, yet it does not concern us in current discussion.

grammar. Thus, a more possible hypothesis is that speakers are aware of this pattern, but in a synchronic way; any historical information is opaque.

On the other hand, there is a possibility that the acoustic properties of [fj] sequence is perceptually less salient; therefore speakers tend to avoid such sequence to occur in the lexicon. This would suggest that a constraint like *[fj] is phonetically natural. The details of such perceptual deficiencies are worth further investigation.

3 How to Treat Accidental Gaps
3.1 The Role of Naturalness

Speakers' judgements on non-words are gradient (Albright, 2009; Berent, Lennertz, Smolensky, & Vaknin-Nusbaum, 2009; Berent, Steriade, Lennertz, & Vaknin, 2007; Chomsky & Halle, 1965; A. Coetzee, 2008). Many would assume that systematic gaps will be marked as less acceptable in wug tests (Berko, 1958), and indeed this is reported in gradient acceptability experiments on English (A. W. Coetzee, 2009). This is because the syllable inventory of English is fuzzy; speakers do not have a clear distinction between words and non-words. Anecdotally, for Chinese, due to its small syllable inventory, native speakers are very sensitive to word/non-word distinction. One would expect that Chinese speakers will behave less gradiently and simply judge all non-words as equally unacceptable. However, previous acceptability experiments demonstrate that apparent accidental gaps are judged better than other apparent systematic gaps (Myers, 2002; Myers & Tsay, 2005; H. S. Wang, 1998), suggesting that even in Chinese, non-word judgement is gradient. Accidental gaps are supposed to be rated more highly than systematic gaps.

Nonetheless, some may argue that there is no clear-cut distinction among systematic and accidental gaps (Boersma & Hayes, 2001; Frisch, Large, & Pisoni, 2000). Things might become trickier when extending this type of analysis on other languages with less restrictive syllable structure and constituency, e.g. English. This is probably why no comprehensive analysis of English phonotactic constraints is available (but see Hayes (2012) for an attempt). It is not feasible to cope with so many phoneme slots and choices. Alternatively, computational phonotactic learner is developed to automatically figure out phonotactic constraints based on MaxEnt Grammar, by feeding lexicon data and the feature matrix (Hayes & Wilson, 2008). However, machine-learned constraints show some interesting patterns. Besides those would have been proposed by human linguists, the model also learns many constraints that make no phonological sense. For example, Hayes & Wilson's model learns a constraint *[+round, +high][-cons, -son] for English, which seems completely unnatural in both phonetic and phonological sense (Hayes & White, 2013).

Phonotactic constraints can be subdivided into natural vs. unnatural ones, each playing distinct roles in phonological grammar. Return to our discussion of SC. If we study the missing syllables in more detail, we can propose more fine-grained and ad-hoc constraints, say, *[Cor, -cont][ən], which is a true reflection of the lexicon. Hayes et al. (2009) have shown that such detailed unnatural constraints are noticed by the speakers and they would make use of them in wug tests. However, it is unlikely to treat them as principled phonotactic constraints. According to Hayes & White (2013), phonotactic constraints should either be typologically well-attested or phonetically-grounded. The detailed constraints often meet neither of the criteria and thus should be marked as unnatural. Very often these unnatural phonotactics show 'surfeit of the stimulus' effect: despite being statistically significant in the lexicon, speakers tend to not generalise these unnatural phonotactic trends into nonce words. Synchronic approaches such as Phonetically Based Phonology (e.g. Hyman 2001; Hayes et al. 2004) would claim that cognitive system is prone to learn natural interactions. Unnatural

processes such as these unnatural phonotactic constraints cannot enter synchronic grammar (Becker, Eby Clemens, & Nevins, 2017; Becker et al., 2011; Becker, Nevins, & Levine, 2012; Pycha et al., 2003), or, at least natural processes are favoured by acquisition while unnatural patterns require extra intention to learn (Hayes et al., 2009; Hayes & White, 2013; Wilson, 2006). Meanwhile, detailed constraints only work for very few missing syllables. Hayes & Wilson's model employs a generality heuristic when searching for phonotactic constraints: priority is given to the constraints which are able to cover a large portion of possible forms. In sum, the above studies suggest that phonetically natural constraints are psychologically real, whereas unnatural constraints are dispreferred and less noticed by speakers. I call this argument the naturalness view.

However, it also seems unfair to ignore all unnatural constraints. Some phonetically justified constraints may have lexical exceptions, lacking systematicity, and speakers may not even notice such constraints. For instance, there is a curious gap in SC: syllables both begin and end with a Coronal sound are disfavoured when the nuclear vowel is high front or mid. This appears to be a long-distance OCP effect. However, this constraint has several lexical exceptions, e.g. [lei, nen, lin] etc. Despite its naturalness, there is no evidence that native speakers will judge these gaps to be more unacceptable than pure accidental gaps. On the other hand, some accidental gaps are more noticed by speakers, and speakers strongly resist such gaps to enter the lexicon. This can be exemplified by the restricted distribution of SC rhyme [ia]. In SC, this rhyme can occur alone, or after Coronal affricates, while it is missing after all other onsets. This gap is genuinely accidental, as Hayes & Wilson (2008)'s phonotactic learner generates 0 penalty scores for these gaps (i.e. completely acceptable as a real lexical item) and there is no obvious natural phonetic explanation for these constraints. However, non-word acceptability tests show that these gaps receive lower judgements than expected, behaving more like those systematic gaps ruled out by natural constraints (Myers, 2002).

Diachronic approach such as Ohala's listener-based sound change model (Ohala, 1981) or Evolutionary Phonology (Blevins, 2004) would suggest that synchronic patterns came into existence as a consequence of a series of diachronic events, and each diachronic stage is fully grounded in phonetic sense, either articulatorily or perceptually. In this case, synchronically natural patterns are certainly expected, and synchronically unnatural patterns can be decomposed into series of natural sound changes. Unnatural processes, just as natural ones, all have a discernible historical explanation, and may retain their productivity. For example, an unnatural constraint is reported for English, which allows only coronal consonant to follow the diphthong /au/, that is, *shout*, *town*, but *laup or *lauk (Hammond, 1999). Though being unnatural, this constraint experiences a natural history: modern /au/ vowel derived from historical monophthong *[u:]. Further, velar and labial sounds after historical *[u:] often underwent lenition, e.g. [b] → [β], [g] → [ɣ], and subsequently vocalised as part of preceding vowel. On the other hand, historical long vowel *[u:] was frequently shortened to short vowel [u] or merged with *[o:] before velars and labials. Each of these steps was unrelated and grounded by phonetic motivation, yet the result that only modern [au][Cor] patterns survive is accidental and may appear to be unnatural. Moreover, some unnatural constraints are recurrent across languages. If we accept synchronic approach, we would expect that unnatural patterns to be deteriorating among world languages. However, there is no evidence to support such declining tendency (Buckley, 2000). For instance, the specific constraint *[fj], which bans labial-dental fricative to occur with coronal glide, holds true for seven major Chinese languages. If unnatural constraints are dispreferred, it is hard to explain why this pattern is recurrent across languages. In addition, Myers (2002) finds that the acceptability of these unnatural gaps is even lower than the gaps ruled out by natural constraints. Unnatural phonotactics, or in this case, accidental gaps, have the potential to be noticed by speakers and

then probably be internalised as part of the phonological grammar, just as other natural constraints. There is a chance for over-phonologisation of unnatural constraints. This can be known as diachrony view.

3.2 The Role of Lexical Statistics

Both naturalness and diachrony view acknowledge that phonotactic constraints are part of the grammar, and this grammar guides speakers' judgements on non-words. However, another potential source for phonological patterns arises from frequency distribution. High frequency results in higher grammaticality (Frisch & Zawaydeh, 2001). It is also argued that structures frequently presented in the lexicon might be phonologised even without phonetic grounding (Bybee, 2001). Also, nonce-words containing well-attested clusters are perceived more acceptable than those with low-frequency or unattested clusters (Coleman & Pierrehumbert, 1997). Therefore, if certain forms are perceived as less acceptable, it may simply due to its lack of support from the lexicon, not due to grammatical naturalness or any lexical-external explanations. 'Grammar' thus becomes the cognitive organisation of one's linguistic experience. If this view is taken to its extreme, one may argue that phonological grammar is a projection of lexical statistics (Hay et al., 2003; Ohala, 1986), or an analogy of the lexicon (Daelemans & van den Bosch, 2005). Myers (2002) finds that speakers' judgements show no significant difference between systematic gaps and accidental gaps. Violations of apparent principled phonotactic constraints are treated the same as of language-specific ad-hoc constraints (contrary to the findings in the current study, as will be discussed below). This indicates that a grammatical approach to phonology might not be necessary; analogical approach may suffice to explain non-word judgement. Non-word repetition tasks investigating the acquisition of novel phonotactic constraints show that speakers learn those constraints from recent experience of the experimental lexicon, suggesting that phonological knowledge can be reduced to statistical knowledge of patterns instantiated in the lexicon (Daland et al., 2011; Dell, Reed, Adams, & Meyer, 2000; Onishi, Chambers, & Fisher, 2002).

   Numerous studies have shown that non-word acceptability judgements are related with lexical statistics (Bailey & Hahn, 2001; Coleman & Pierrehumbert, 1997; Frisch et al., 2000; Hay et al., 2003; Vitevitch & Luce, 1999, 1998). Other than English, such lexical effect on non-word gradient acceptability is also found in SC (Myers & Tsay, 2005) and Cantonese (Kirby & Yu, 2007). Different models have been proposed to quantify this lexical influence. First, speakers may evaluate how similar the non-word is with other known words, which is known as Neighbourhood Density. This is defined as the number of words generated by substituting, deleting, or adding a single phoneme together with their summed frequency (Bailey & Hahn, 2001; Greenberg & Jenkins, 1964). For example, the form *lat* has abundant lexical neighbours in English (e.g. *cat*, *lap*), while *zev* has sparse neighbourhood density. Second, speakers may decompose a non-word into substrings and calculate the combinatory probability of the phonemes of each substring. This method is called Phonotactic Probability (Jusczyk & Luce, 1994; Vitevitch & Luce, 1998). A widely accepted way to calculate phonotactic probability is position-specific biphone probability (Vitevitch & Luce, 2004). Given a sequence *st* found in the third and fourth positions of a word in English, we need to first sum up the frequency of all English words that contain *st* as their third and fourth segments, and divide that by the frequency count of all words that contain biphones in third and fourth positions. These two factors are usually confounded. since both are derived from lexical frequencies, high value in one factor may cause the other factor to be high as well. Nevertheless, many studies have proved that phonotactic probability and neighbourhood density have distinct effect on acceptability judgement (Bailey & Hahn, 2001). Myers & Tsay (2005) demonstrate that phonotactic probability boosts lexical decision in both word and non-

word scenario, while neighbourhood density only shows positive effect in real words; in non-words the effect is negative. This suggests that phonotactic probability is more like a pre-lexical grammatical process, while neighbourhood density is closely connected to the lexicon. This study also indicates that phonotactic probability is always a facilitator in lexical processing, while neighbourhood density, sometimes, can exhibit inhibitory effect, as too many neighbours may slow down the process of locating lexical item. Generally, phonotactic probability and neighbourhood density are considered as non-grammatical effect in language processing, since they are directly derived from lexical statistics (Inkelas, Orgun, & Zoll, 1997). However, some argue that phonotactic probability involves phonological generalisation, and thus should be viewed as part of the grammar. For example, Albright (2009) proposes a feature-based model which generates biphone probability by calculations over combinations of natural classes. Hayes & Wilson (2008)'s Phonotactic Learner also employs phonological features to calculate the optimal phonotactic constraint set that maximises the probability of input lexical items. On the other hand, the role of neighbourhood density is much more analogical rather than grammatical, as it only measures the similarity of a form with other lexical words.

Thus, one can argue that gradient acceptability of the missing syllables is a direct consequence of lexical statistics. Those missing forms which contain highly probable segment sequences or, more generally, have denser lexical neighbours, will be marked as more acceptable. However, this approach often fails to compare the acceptability between two missing forms, say [dl] and [bz]. Both sequences are missing from English and therefore having low phonotactic probability and sparse neighbourhood density (Albright, 2009). In addition, lexical frequency account says very little about the naturalness bias. A missing form is more acceptable than another because it is more word-like, not because it is more natural, or more grammatical. Experimental results also indicate that formal grammatical constraints like OCP do affect acceptability judgements, even when neighbourhood density and phonotactic probability are controlled (A. Coetzee, 2008; Frisch & Zawaydeh, 2001). A compromise approach between pure frequency-based and pure formal-grammar is in order: learners use frequency information only at the initial stage of learning. After being familiar with a patterns, grammar sets apart from lexical statistics and is ready for independent evolvements (A. Coetzee, 2008).

3.3 Summary

So far, we have met two theoretical questions. First, how do speakers treat natural vs. unnatural patterns? Do they prefer natural forms relative to unnatural ones, as predicted by naturalness view? Or diachrony view is more likely, that unnatural patterns are treated the same, or sometimes, even psychologically more real than natural patterns? Various experiments have confirmed that natural constraints are preferred (e.g. Berent et al. 2007; Carpenter 2010). Therefore, it is reasonable to believe that natural constraints form part of the grammar, and indeed, all constraints proposed for SC phonotactics in (3) and (4) are natural. Much of the debate has been focusing on unnatural patterns. Becker et al. (2011) argue that they are unlearnable, while others propose for soft bias, saying that unnatural patterns do exert some effect in acquisition and performance, but they are dispreferred (Hayes et al., 2009; Hayes & White, 2013; Wilson, 2006). Yet there is also evidence to show that unnatural constraints are being noticed by speakers. For example, Kager & Pater (2012) provide experimental evidence that Dutch speakers are actively aware of a phonotactic constraint which is too complex to be natural. There is no evidence that language-specific ad-hoc constraints cannot be more psychologically real than natural ones.

Second, suppose that speakers have implicit knowledge about phonotactic constraints, either natural or unnatural, we are also interested in the nature of this knowledge, i.e. is it a grammatical effect, or it is generalised from lexical analogy? Maybe it is more likely to accept natural constraints as part of the grammar. However, are unnatural constraints also part of the grammar, acting like other natural, principled constraints, or they are just accidentally true, an analogical reflection of lexical statistics? Also, we are interested in whether all unnatural constraints behave the same, that is, receiving similar wordlikeness judgements? If not, is it purely due to lexical statistics? Or it is because certain unnatural constraints are more prominent than the others and therefore guide speakers' performance in such judgements?

In order to test the source of unnatural phonotactics, this paper will focus on one specific constraint, *[fj], as mentioned earlier. There is no straightforward phonetic evidence to explain this constraint, yet this pattern is widely found across different Chinese languages.

4 Statistical Analyses
4.1 Hypotheses

To evaluate the role of unnatural phonotactic constraints, this study compares speakers' wordlikeness judgement performance of *[fj] violating gaps with (i) systematic (i.e. natural) gaps, and (ii) other accidental (i.e. unnatural) gaps. Previous experimental studies have shown that speakers will be slower to reject a token as non-word if it is more well-formed (Berent, Shimron, & Vaknin, 2001). It is therefore hypothesised that *[fj] violating gaps receive less positive responses and are rejected more quickly by participants than other accidental gaps. This will confirm that *[fj] behaves less accidentally and is learned well as if it were a natural constraint. Furthermore, to test whether the variance in performance can be simply explained by trends of the lexicon, lexical statistics will be included in the model. Suppose that *[fj] indeed is an unnatural constraint, if *[fj] effect remains significant after taking lexical statistics into consideration, then at least some unnatural phonotactics can be internalised as part of the grammar, which favours the diachrony view.

4.2 Data Sources

This study draws data from a phonological acceptability judgement mega study run on 110 Mandarin native speakers (Myers & Tsay, 2015). All 3,274 theoretically possible but actually missing syllables that are representable under Zhuyin Fuhao system were presented to the participants in written stimuli to judge their acceptability. Each participant was required to judge all missing syllables. Since this study was conducted in Taiwan, the written stimuli were in the form of Zhuyin Fuhao symbols, a phonetic spelling system used in Taiwan, equivalent to Mainland China's Pinyin. Participants were asked to judge each stimulus as 'like Mandarin', or 'not like Mandarin' by pressing keys on the keyboard as quickly as possible, and their response and reaction time were recorded. Even though SC speakers are said to possess a sharp intuition between words and non-words (Myers, 2002; H. S. Wang, 1998), 13.42% of all responses were positive despite the fact that all syllables presented to participants were non-words.

Not all missing syllables included in the original data are selected for current analysis. First, original data distinguish lexical tones of each missing syllable, which are merged in current study. For example, the missing syllable *[pja] is labelled as [pja1], [pja2], [pja3], [pja4] in the original dataset, and participants offer judgements to each of them. Now that lexical tones are merged, these four missing forms are treated as identical, yet their reaction times are preserved, as if participants judge [pja] for four times. This is true for all syllables

being considered so the tonal distinctions of reaction time are balanced across all syllables. Second, due to different phonemic analysis, many syllable forms in the original dataset are non-contrastive under the current phonemic analysis of (1), and therefore being eliminated from further consideration. For instance, ㄝ [je] and ㄛ [jo] are counted as different forms in Myers & Tsay (2015)'s missing syllable list. However, in fact, SC has only one mid vowel phoneme. The mid vowel /ə/ is realised as [e] after the glide [j], so [jo] is impossible and redundant. In other words, the constraints covered in this study are all morphophonotactic constraints which operate on the phonemic input segments, rather than surface phonetic output segments. After merging lexical tones and ruling out impossible forms, we identify 360 non-lexical syllables from the original 3,274.[4] Among them, 128 syllables are labelled as accidental, and the remaining 232 are systematic gaps, i.e., ruled out by the constraints proposed in (3) and (4). Further, we label 12 *[fj] violating syllables out of all others as out theoretical interest group.

As mentioned before, non-word judgement data can be explained by lexical statistics. Two models have been introduced as the predictors for the reaction time variance in this study. The first one is Hayes & Wilson (2008)'s Maximum Entropy Phonotactic Learner, which is a constraint-based phonotactic learner producing a set of feature-based constraints given a feature matrix and a lexicon for training. The learner attempts to identify the constraint set and a set of constraint weights that maximise the probability of the input forms. The training lexicon is derived from Tsai (2000)'s Mandarin Syllable Frequency Counts for Chinese Characters. It iterates through the 111,417 words included in an online Chinese lexicon Libtabe (Hsiao, Tsai, Hsieh, Yeh, & Tan, 2013) and counts the frequency of occurrence for each pronunciation of each character. As a result, the frequency was type frequency. For example, if a character is pronounced "hui4" in 402 words in the lexicon, the frequency will be 402. The pronunciations with their frequencies are then merged with tones (by adding up the type frequencies of four lexical tones) and reinterpreted under the phonemic analysis in (1). In the end this generates a corpus of 399 lexical entries with type frequency. The feature definition of each segment follows the analysis in Duanmu (2007). The number of constraints to be learned is set as 400. After obtaining the constraint set, we apply this grammar to the testing data, which is the 360 non-lexical syllables mentioned earlier. Each missing syllable is thus assigned a penalty score to reflect its lexical well-formedness. The second lexical statistics predictor is Neighbourhood Density. We use the standard definition of Levenshtein Edit Distance, that is, by calculating the number of lexical neighbours differing in only one phoneme (either adding, deleting, or substituting) from the test item for all 360 non-lexical syllables (Luce & Pisoni, 1998). The corpus data is the same as the one used for Phonotactic Learner, and the searching algorithm is provided by Hall et al. (2017).

4.3 Results

First, we use the independent variables to predict reaction time. Here we exclude the 13.42% of positive responses, since positive responses may be involved with a distinct processing pattern than of non-word judgement. The statistical analysis employs a powerful technique named mixed-effects linear regression (Baayen, Davidson, & Bates, 2008), which simultaneously incorporates random effects (the participants and test items) and fixed effects (in this case, our four independent predictors: penalty scores generated by phonotactic learner,

---

[4] This number is different from 1,496 as discussed before. This is because Myers & Tsay (2015)'s data does not include any syllable violating *HH, since rhymes violating this constraint like [iu], [yi] cannot be represented by Zhuyin Fuhao system, and indeed a large number of missing syllables is under this constraint.

neighbourhood density, whether being systematic or accidental gaps, and whether being *[fj] violating gaps). Analyses were calculated by *lmer* function of *lme4* package (Bates, Mächler, Bolker, & Walker, 2014) in the R statistical programming language. No logarithmic transformation was applied to reaction time, because log-likelihood suggests that a model without such transaction has a better fit. Although Barr et al. (2013) recommend that one should preserve the maximal random effects by involving both random intercept and possible slopes, due to convergence failures we consider only random intercepts for participants and test items and random slope for *type = fj*, since that is our theoretical interest. The current model is a fully crossed one, which includes four independent variables: *penalty*, *neighbourhood density*, *being a systematic gap*, *being a *[fj] gap*, and all their interactions, including *penalty : neighbourhood density*, *penalty : systematic*, *neighbourhood density : systematic*, *penalty : fj*, *neighbourhood density : fj*, *penalty : neighbourhood density : systematic*, *and penalty : neighbourhood density : fj*. Notice that all *[fj] violating syllables are also accidental gaps, which means *type = fj* does not interact with *type = systematic*. All these interactions are removed from the model.

	β	SE(β)	t	p (t=z)	p (LRT)[5]
(Intercept)	774.51	33.00	23.47		
penalty	-0.01	1.42	-0.01	.9938	.9933
neighbourhood density	10.46	0.77	13.56	<.0001*	<.0001*
type = systematic	7.31	16.47	0.44	.6575	.6556
type = fj	29.07	79.18	0.37	.7136	.7144
penalty : *neighbourhood density*	-1.08	0.08	-12.97	<.0001*	<.0001*
penalty : *type = systematic*	-1.61	1.58	-1.02	.3094	.3066
ND : *type = systematic*	-4.93	1.03	-4.81	<.0001*	<.0001*
penalty : *type = fj*	-2.86	5.86	-0.49	.6257	.6256
ND : *type = fj*	-12.19	4.87	-2.51	.0123*	.0123*
penalty : *ND* : *type = systematic*	0.44	0.10	4.58	<.0001*	<.0001*
penalty : *ND* : *type = fj*	0.73	0.38	1.94	.0531	.0532

Table 1 Results of mixed-effects linear regression on reaction time

In general, the results accord with our hypothesis. Penalty score is negatively related to reaction time (Figure 1a): syllables with higher penalty scores are rejected more quickly, that is, more unacceptable. Neighbourhood density is reported to lower the acceptability of a non-lexical word when it gets higher (Myers & Tsay, 2005), however the current model shows that syllables with denser neighbours are rejected more slowly, i.e. more acceptable (Figure 1b). This is in line with Kirby & Yu (2007)'s finding on Cantonese.

---

[5] Exact *p*-values could not be obtained by *lmer* function, since there is controversy on how to decide the degrees of freedom when calculating them. Here we report two methods of calculation. The first method assumes infinite degrees of freedom so that *t* value can be equivalent to *z* scores. The second method uses *anova* function to run a likelihood ratio test on full model vs. models missing each parameter.
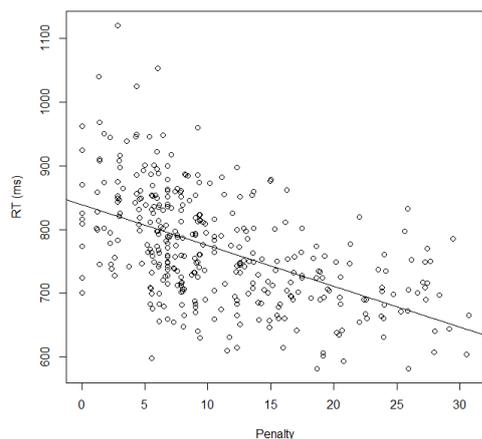
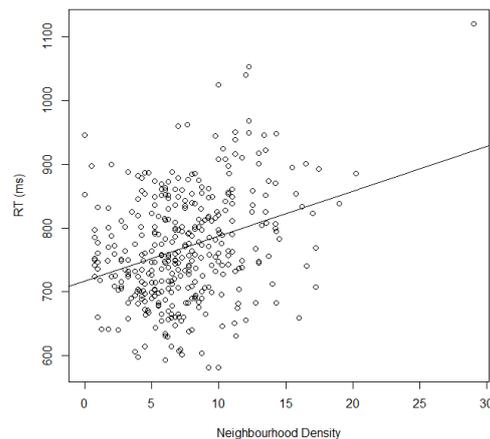**Figure 1a** Penalty score effect on reaction time

Figure 1b Neighbourhood density effect on reaction time

Although being systematic gap alone does not affect reaction time significantly, its interaction with neighbourhood density reaches significance, supporting the fact that systematic gaps are rejected more quickly than accidental ones (Figure 2a). This also suggests that grammatical effect still holds even after taking lexical statistics into consideration. Finally, Figure 2b shows that the twelve *[fj] gaps behave significantly different from other accidental gaps: they are rejected much more quickly than they should be. Also, the regression results show that interaction of neighbourhood density and *[fj] gaps reaches significance, even though, again, being *[fj] alone does not. This indicates that the unnatural constraint *[fj] works pretty much the same as other phonetically natural constraints. Speakers are aware that *[fj] gaps are less acceptable than other 'pure' accidental gaps, and treat them as if they are systematic gaps.
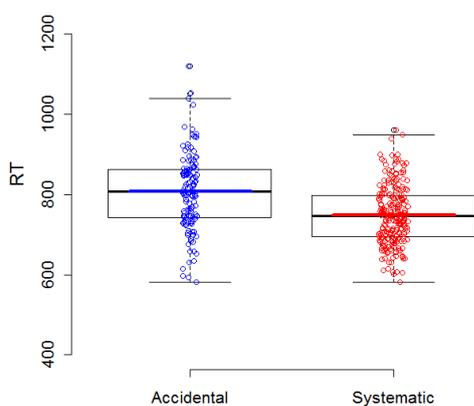

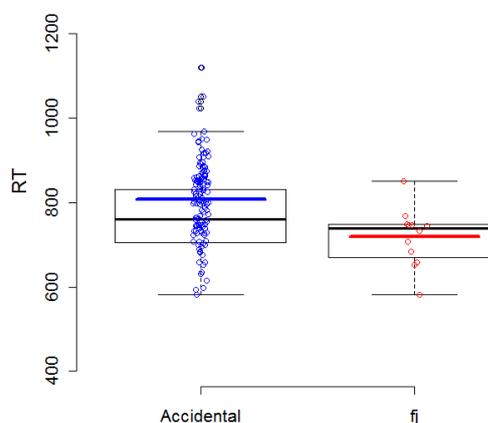**Figure 2a** Reaction time distribution of systematic vs. accidental gaps

Figure 2b Reaction time distribution of accidental vs. *[fj] violating gaps

Next, we consider participants' lexical responses, using the same independent variables to predict the yes (like Mandarin) or no (not like Mandarin) responses on test items. An examination on individual response pattern reveals that most participants rejected all test items (70 out of 110), whereas three participants accepted all. For those who offered both positive and negative responses, their response pattern tended to be extreme as well, either rejecting nearly all the syllables or accepting nearly all. This indicates that subjects may have different understandings of the experiment task. Therefore, we excluded 70 all-negative and 3

all-positive participants, which left 37 participants' data for analyses. This all-in or all-out behaviour made random effect models hard to converge. Therefore, analyses were done by logistic regression without random effects, using *glm* function of *lme4* package.

Maximal model with fully crossed parameters was first attempted, yet the results showed that most interaction effects were not significant. We first took out the most complicated three-way interaction parameters, and the results showed that still the four interactions involving the two categorical parameters were not significant. The model was then reduced to parameters without interactions, except penalty score vs. neighbourhood density.

	β	SE(β)	t	p
(Intercept)	-0.7212	0.0308	-23.442	
penalty	-0.0051	0.0020	-2.514	.0119*
neighbourhood density	0.0010	0.0028	3.557	.0004*
type = systematic	-0.0487	0.0223	-2.183	.0290*
type = fj	-0.0302	0.0528	-0.572	.5673
penalty : *neighbourhood density*	-0.0007	0.0002	-3.054	.0023*

Table 2 Results of linear logistic regression on response

The results re-confirm our hypotheses: similar with the case of reaction time, higher penalty score lowers the acceptability (Figure 3a), and neighbourhood density increases it (Figure 3b).
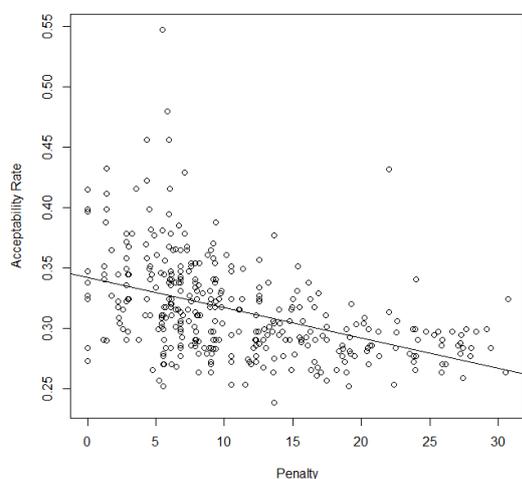

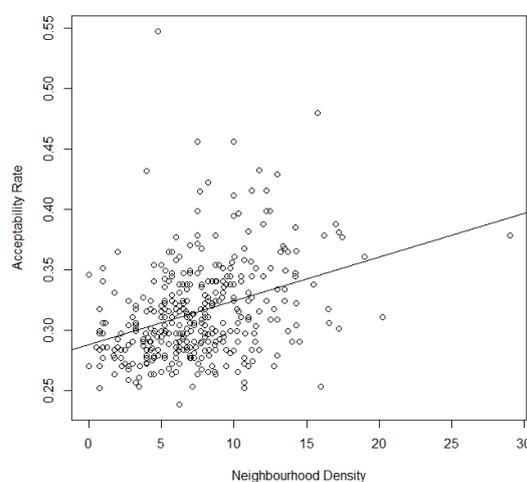**Figure 3a** Penalty score effect on acceptability rate

Figure 3b Neighbourhood density effect on acceptability rate

Overall, the acceptability rates stay at low level. This is because participants rejected most of the test items. According to regression analyses, being a systematic gap or a *[fj] violating gap both result in lower acceptability, as illustrated by Figure 4a and Figure 4b. However, the *[fj] factor does not reach significance in the case of acceptability rate.
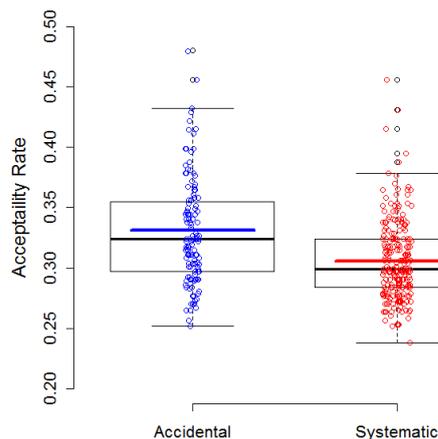
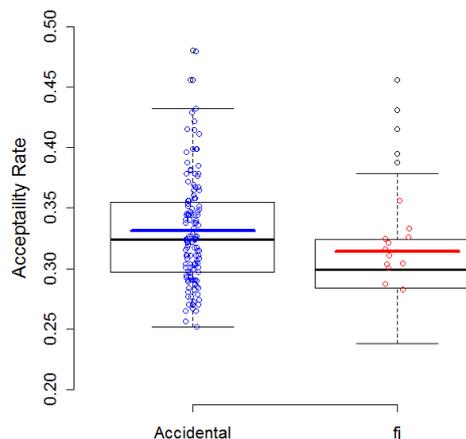**Figure 4a** Acceptability rate distribution of accidental vs. systematic gaps

Figure 4b Acceptability rate distribution of accidental vs. *[fj] violating gaps

5 Discussion

The analyses indicate that neighbourhood density has a tight correlation with speakers' non-word judgements, both in terms of reaction time and acceptability rate. This critical role of neighbourhood density in non-word judgement has been established in plenty of previous studies, including English (Bailey & Hahn, 2001; Luce & Pisoni, 1998), Arabic (Frisch & Zawaydeh, 2001), Cantonese (Kirby & Yu, 2007), and of course, Standard Chinese (Myers & Tsay, 2005). Neighbourhood density is the only factor that reaches significance alone in predicting reaction time, and most interactions involving neighbourhood density are also significant. In contrast, penalty score's predictive ability is inferior to neighbourhood density, which will be further discussed in later section.

The results also show that in addition to the influence from lexical trends and usage frequencies, speakers are also aware of the distinction made by grammaticality: non-words ruled out by the OCP constraints proposed in (3) and (4) were judged systematically worse than other accidental missing forms, even when lexical statistics are incorporated (Berent et al., 2007; A. Coetzee, 2008; Frisch & Zawaydeh, 2001; Moreton, 2008). This result is supported by both reaction time and acceptability rate. In addition, a phonetically unnatural constraint is proved to be over-phonologised: syllables violating a *[fj] constraint behaved distinctly from other accidental gaps and were rejected more quickly as if they were systematic gaps. This finding contradicts with the naturalness view, which assumes unnatural patterns are dispreferred in terms of perceptibility and learnability (Hayes & White, 2013; Wilson, 2006). Unnatural phonotactics like *[fj] can be part of phonological grammar and guide linguistic performance (Kager & Pater, 2012).

5.1 Unnatural Phonotactics as Part of Linguistic Knowledge

The model reported in Table 1 is the maximal model where the major effect of *fj is covered by other interaction parameters. If we simplify the model and only consider the four main factors without their interactions, the effect of *fj would be clearer.

	β	SE(β)	t	p ($t=z$)	p (LRT)
(Intercept)	861.10	31.59	27.26		
penalty	-5.78	0.59	-9.75	<.0001*	<.0001*
neighbourhood density	-1.10	0.25	-4.38	<.0001*	<.0001*
type = systematic	-30.02	9.03	-3.32	.0009*	.0009*
type = fj	-62.90	22.23	-2.83	.0049*	.0039*

Table 3 Simplified mixed-effects linear regression on reaction time

This shows that *fj gaps significantly reduce participants' reaction time and are rejected more often than other 'pure' accidental gaps, suggesting that unnatural phonological processes caused by diachronic development can retain some productivity in synchronic grammar. For instance, English speakers successfully apply the [k]~[s] alternation (electri[k]~electri[s]ity) in made-up Latinate words, less so in semi-Latinate words, and almost never for non-Latinate stimuli (Pierrehumbert, 2006). She argues that despite [k]~[s] alternation is an unnatural pattern, it can still be learned from statistical inference. Similar findings are reported in Japanese Lyman's Law (Kawahara, 2008).

Theorists who hold a naturalness view towards phonological knowledge often assume that unnatural patterns are biased and subsequently dispreferred in perception and learnability (Hayes et al., 2009; Hayes & White, 2013; Wilson, 2006), some even argue that unnatural patterns are unlearnable (Becker et al., 2011). In their view, natural processes are backed by Universal Grammar; speakers have free access to them without extra work. Unnatural processes are possible via dissimilation invoking or hypercorrection (Ohala, 1981), but they are treated as consequences of historical development and therefore require additional learning. Nevertheless, the current finding suggests that unnatural phonotactics is not necessarily dispreferred. At least some unnatural constraints are equally likely to be noticed by speakers as other phonetically natural patterns, such as OCP based co-occurrence constraints. If so, the bias against unnatural patterns is removed, which means unnatural patterns are no longer 'historical accidents' per se. All phonological patterns can be viewed as the results of historical developments, and unnatural processes are as legit as natural ones. Of course, articulatory, acoustic, or perceptual factors still hold to support the fact that natural patterns are typologically much more frequent than unnatural ones, but this does not prevent unnatural patterns from entering the grammar. There is no synchronic device that enforces naturalness. This diachrony view of phonological knowledge calls a diachronic analysis for all synchronic patterns. Due to the majority of sound changes is phonetically motivated, it is no surprise that most synchronic patterns are also natural. However, unnatural patterns do sometimes occur, and they behave no differently as natural ones. In naturalness view, whenever such unnatural pattern is determined, one needs to seek for additional explanations. But now, all patterns are accounted for by historical developments, no need to draw the distinction between natural vs. unnatural pattern in the synchronic grammar again.

This does not mean that synchronic patterns will encode all the intermediate stages of diachronic development. Rather, it adopts a flattened, or telescoped way. Once a pattern is phonologised from sound change, it loses its connection with phonetics and can be subject to non-phonetic conditioning (Hyman, 2001). In our case, SC speakers clearly have no access to the fact that synchronic [f] potentially already contains a glide [j] and thus forbids it to be combined with another [j]. Speakers are not equipped with any historical knowledge behind the pattern, but still apply this phonotactic constraint fluently (Scheer, 2015).

5.2 Phonotactic Learner: A Mixture of Lexical Statistics and Grammar

Taking a close look at Figure 1a and Figure 3a, we notice that penalty score behaves categorically: the regressive relation is much more salient when penalty score is at low level. When penalty score is relatively high, more specifically, above 13, the scatter plot starts to look random and linear relation is hard to determine.
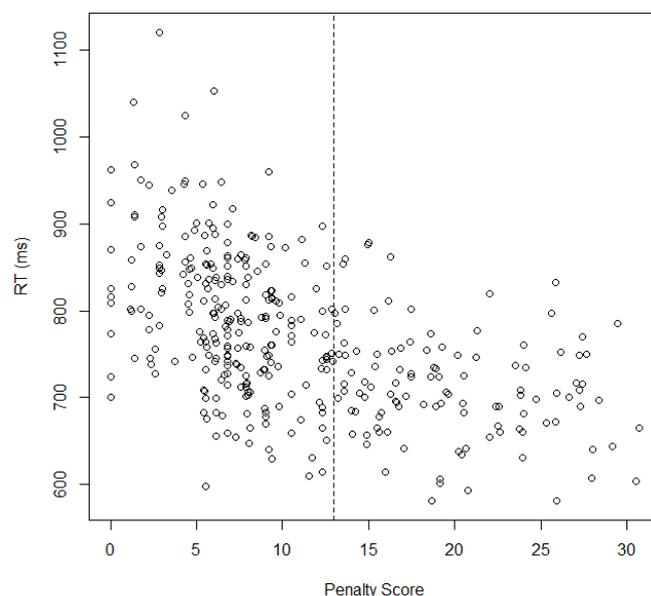

Figure 5 Spitting penalty score effect on reaction time

As we can see, the distribution on the left-hand side of the cut-off is much denser, meaning a more salient correlation. This categorical nature of penalty score has influence on the relation between neighbourhood density and reaction time (or acceptability rate) as well. If we split the data into two parts, one with low penalty score (<13), the other with high penalty score (≧13), and make two separate plots on neighbourhood density vs. reaction time, the results would be similar: when penalty score is low, there is a strong correlation between ND and RT, whereas no discernible linear relation when penalty score is high.
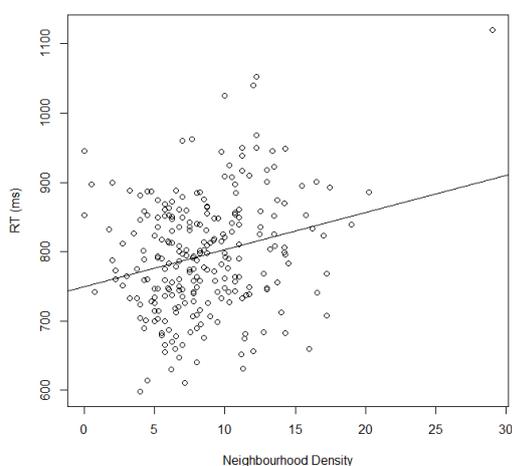

Figure 6a Neighbourhood density effect on reaction time when penalty score is low (<13)
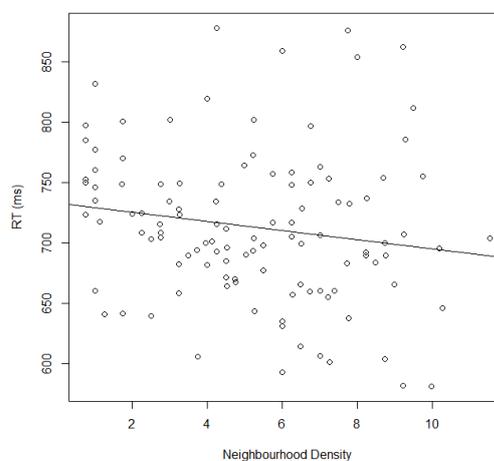
Figure 6b Neighbourhood density effect on reaction time when penalty score is high (≧13)

This categorical behaviour of penalty score results from the fact that Phonotactic Learner is not only based on lexical statistics. Apart from the input lexicon, Phonotactic Learner loads in additional non-lexical information such as feature system and constraint structure when calculating the most representative constraints on feature combinations. This means that it does not make generalisations simply out of the lexicon, but in the meantime also presumes certain grammatical features. On the contrary, neighbourhood density is a pure analogical model, which only calculates the similarity of a word with the lexicon.

This methodological difference between Phonotactic Learner and neighbourhood density explains their distinct behaviour on predicting speakers' performance. Being a pure analogical model, neighbourhood density shows a much more gradient effect on acceptability, whereas the penalty scores offered by Phonotactic Learner present a categorical tendency, because in essence, Phonotactic Learner is a mixture of grammaticality and lexical statistics. Gradient effect is only found when penalty score is low, that is, more grammatical. When the ungrammaticality accumulates to a certain level, speakers judge those forms as all bad, and reject them equally quickly without considering their wordlikeness (See Section 5.3 for more discussion).

5.3 What is Grammar? How Does It Influence Phonotactic Processing?

Traditionally, grammar is defined as formal abstract principles that guides the construction of utterances. It could be constraint ranking in Optimality Theory, a set of rules that derive surface form from underlying representation, or, in our case, phonotactic constraints that categorise legal syllables from illegal ones. In the past few decades, various experiments have been looking at the frequency effect on linguistic performance. Their success leads to a new interpretation, that 'grammar is the cognitive organisation of one's experience with language' (Bybee, 2001). In other words, there is no abstract principles about languages, the real grammar is simply a derivation of frequency statistics. Lexical statistics itself is grammar.

Indeed, it is not easy to disentangle pure grammar and lexical statistics. For example, the statistical results in Table 1 show that there is a noticeable interaction between neighbourhood density (standing for lexical statistics) and being a systematic gap (standing for pure grammar). More specifically, if a syllable shares plenty of neighbours, then it is more likely to be an accidental gap rather than ungrammatical systematic gap.

However, if we contribute all variation in linguistic performance to lexical statistics, it is hard to explain why speakers present distinct behaviours on structures which they have no experience (A. Coetzee, 2008). In English, forms like [spVp] and [skVk] are both missing from the lexicon, yet speakers still systematically prefer [skVk] over [spVp], which suggests that a mini constraint ranking like *spVp >> *skVk >> Faithfulness >> *stVt is at work. Speaker's well-formedness knowledge about zero frequency forms is a strong indicator for independent formal grammar.

Moreover, in Section 5.2 it has been demonstrated that due to the dual behaviour of penalty score, neighbourhood density is less predictive when penalty score is high (i.e., more ungrammatical). This is also true for the pure grammatical division of systematic vs. accidental gaps: neighbourhood density explains a lot of variance for accidental gaps, while for ungrammatical systematic gaps, the distribution is much more random-like. Daland et al. (2011)'s findings are relevant here, in which they report that pure lexical statistics models are able to predict the well-formedness of non-words with attested clusters, i.e. more grammatical words. In contrast, a more grammar-oriented model like MaxEnt Phonotactic Learner treats all non-words with attested clusters as equally acceptable, and shows a good fit in predicting and unattested (ungrammatical) forms. These findings would suggest that formal grammar

and lexical statistics do work in different ways. During phonotactic processing, grammaticality shows a filter effect. When speaker hears a non-word, he or she first consults the phonological grammar to decide its grammaticality. If the target is judged bad by the grammar (systematic gap), it is rejected directly without referring to neighbourhood density. This explains the right-hand side of Figure 5 and Figure 6b, where highly ungrammatical forms are quickly rejected, no matter it is in dense or sparse neighbours. If the target passes grammaticality filter (accidental gap), then speaker will move onto lexical statistics, looking up its neighbourhood density and offering gradient responses accordingly (Figure 6a). This grammaticality filter effect is backed up with neurolinguistic evidence, where ungrammatical sequence evoke a distinct N400 effect, regardless of whether such sequences are lexically well-attested or not (White & Chiu in press).

In sum, despite the interaction between grammaticality and lexical statistics, there is still reason to believe that they play distinct roles in phonotactic processing and other general linguistic performance.

6 Conclusion

This paper has answered the two theoretical questions raised in Section 3.3. First, unlike what has been proposed by Becker et al. (2011), unnatural phonotactics can certainly be noticed and learned by speakers, and thus enter the phonotactic knowledge. In addition, previous studies such as Wilson (2006) and Hayes & White (2013) argue for a bias for unnatural phonotactics. However, the SC case of *[fj] constraint demonstrates that unnatural phonotactic constraints are not necessarily dispreferred by speakers. They can be equally, and even more, psychologically real than other natural constraints. Second, the extreme lexicalist view which attributes all phonotactic patterns to frequency statistics is too strong. Lexical statistics and grammaticality have independent influence on phonotactic processing, even though there can be much overlap between them. The nature of phonotactic knowledge consists of both grammaticality and lexical statistics.

References

Albright, A. (2009). Feature-based generalisation as a source of gradient acceptability. *Phonology*, *26*(1), 9.

Anderson, S. R. (1981). Why Phonology Isn't "Natural." *Linguistic Inquiry*, *12*(4), 493–539.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Bach, E., & Harms, R. (1972). How do languages get crazy rules? In R. P. Stockwell & R. K. S. Macaulay (Eds.), *Linguistic change and generative theory* (pp. 1–21). Bloomington: Indiana University Press.

Bailey, T. M., & Hahn, U. (2001). Determinants of Wordlikeness: Phonotactics or Lexical Neighborhoods? *Journal of Memory & Language*, *44*(4), 568.

Bao, Z. (1990). *On the nature of tone*. Massachusetts Institute of Technology.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2014). Fitting linear mixed-effects models using lme4. Retrieved from https://arxiv.org/abs/1406.5823

Becker, M., Eby Clemens, L., & Nevins, A. (2017). Generalization of French and Portuguese plural alternations and initial syllable protection. *Natural Language and Linguistic Theory*, *35*(2), 299–345.

Becker, M., Ketrez, N., & Nevins, A. (2011). the Surfeit of the Stimulus: Analytic Biases Filter Lexical Statistics in Turkish Laryngeal Alternations. *Language*, *87*(1), 84–125.

Becker, M., Nevins, A., & Levine, J. (2012). Asymmetries in generalizing alternations to and from initial syllables. *Language*, *88*(2), 231–268.

Begus, G., & Nazarov, A. (2017). Lexicon against Naturalness : Unnatural Gradient Phonotactic Restrictions and Their Origins. *Ms.*, 1–44.

Berent, I., Lennertz, T., Smolensky, P., & Vaknin-Nusbaum, V. (2009). Listeners' knowledge of phonological universals: evidence from nasal clusters. *Phonology*, *26*(1), 75.

Berent, I., Shimron, J., & Vaknin, V. (2001). Phonological Constraints on Reading: Evidence from the Obligatory Contour Principle. *Journal of Memory and Language*, *44*(4), 644–665.

Berent, I., Steriade, D., Lennertz, T., & Vaknin, V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition*, *104*(3), 591–630.

Berko, J. (1958). A child's learning of English morphology. *Word*, *14*(November), 150–77.

Blevins, J. (2004). *Evolutionary Phonology*. Cambridge: Cambridge University Press.

Boersma, P., & Hayes, B. (2001). Empirical Tests of the Gradual Learning Algorithm. *Linguistic Inquiry*, *32*(1), 45–86.

Buckley, E. (2000). On the Naturalness of Unnatural Rules. *Proceedings from the Second Workshop on American Indigenous Languages.*, *9*, 1–14.

Bybee, J. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.

Carpenter, A. C. (2010). A naturalness bias in learning stress. *Phonology*, *27*(3), 345–392.

Chao, Y.-R. (1968). *A Grammar of Spoken Chinese*. Berkeley and Los Angeles: University of California Press.

Chomsky, N., & Halle, M. (1965). Some controversial questions in phonetical theory. *Journal of Linguistics*, *1*(2), 97–138.

Coetzee, A. (2008). Grammaticality and Ungrammaticality in Phonology. *Language*, *84*(2), 218–257.

Coetzee, A. W. (2009). Grammar is both categorical and gradient. *Phonological Argumentation: Essays on Evidence and Motivation*, 1–36.

Coetzee, A. W., & Pater, J. (2008). Weighted Constraints and Gradient Restrictions on Place Co-Occurrence in Muna and Arabic. *Natural Language & Linguistic Theory*, *26*(2), 289–337.

Coetzee, A. W., & Pretorius, R. (2010). Phonetically grounded phonology and sound change: The case of Tswana labial plosives. *Journal of Phonetics*, *38*(3), 404–421.

Coleman, J., & Pierrehumbert, J. (1997). Stochastic phonological grammars and acceptability. *Computational Phonology Third Meeting of the ACL Special Interest Group in Computational Phonology*.

Daelemans, W., & van den Bosch, A. (2005). Memory-based language processing. *Computational Linguistics*, *11*(3), 287–296.

Daland, R., Hayes, B., White, J., Garellek, M., Davis, A., & Norrmann, I. (2011). Explaining sonority projection effects. *Phonology*, *28*(2011), 197–234.

Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: a study of the role of experience in language production. *J Exp Psychol Learn Mem Cogn*, *26*(6), 1355–1367.

Duanmu, S. (1990). *A Formal Study of Syllable, Tone, Stress, and Domain in Chinese Languages*. Massachusetts Institute of Technology.

Duanmu, S. (2003). The syllable phonology of Mandarin and Shanghai. *Proceedings of the 15th North American Conference on Chinese Linguistics (NACCL 15)*, 86–102.

Duanmu, S. (2007). *The phonology of standard Chinese. The Phonology of the world's languages* (2nd ed.). Oxford ; New York: Oxford University Press.

Duanmu, S. (2008). *Syllable Structure: The Limits of Variation*. Oxford ; New York: Oxford University Press.

Duanmu, S. (2011). Chinese Syllable Structure. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell Companion to Phonology* (Vol. 5). Blackwell Publishing.

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.

Frisch, S., Large, N. R., & Pisoni, D. B. (2000). Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language*, *42*(4), 481–496.

Frisch, S., Pierrehumbert, J. B., & Broe, M. B. (2004). Similarity avoidance and the OCP. *Natural Language and Linguistic Theory*, *22*(1), 179–228.

Frisch, S., & Zawaydeh, B. A. (2001). The Psychological Reality of OCP - Place in Arabic. *Language*, *77*(1), 91–106.

Greenberg, J. H., & Jenkins, J. J. (1964). Studies in the psychological correlates of the sound system of American English. *Word*, *20*(July), 157–177.

Hall, K. C., Allen, B., Fry, M., Mackie, S., & McAuliffe, M. (2017). Phonological CorpusTools. Retrieved from http://phonologicalcorpustools.github.io/CorpusTools/

Halle, M. (1962). Phonology in generative grammar. *Word*, *18*, 54–72.

Hammond, M. (1999). *The Phonology of English: a prosodic Optimality-Theoretic approach*. Oxford: Oxford University Press.

Hay, J., Pierrehumbert, J., & Beckman, M. E. (2003). Speech perception, well-formedness and the statistics of the lexicon. In J. Local, R. Ogden, & R. Temple (Eds.) (pp. 58–87). Cambridg2: Cambridge University Press.

Hayes, B. (2012). The role of computational modeling in the study of sound structure. In *Conference on Laboratory Phonology*. Stuttgart.

Hayes, B., Kirchner, R., & Steriade, D. (2004). *Phonetically Based Phonology*. (B. Hayes, R. Kirchner, & D. Steriade, Eds.). Cambridge: Cambridge University Press.

Hayes, B., Siptar, P., Zuraw, K., & Londe, Z. (2009). Natural and Unnatural Constraints in Hungarian Vowel Harmony. *Language*, *85*(4), 822–863.

Hayes, B., & White, J. (2013). Phonological naturalness and phonotactic learning. *Linguistic Inquiry*, *44*(1), 45–75.

Hayes, B., & Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, *39*(3), 379–440.

Hsiao, P.-H., Tsai, C.-H., Hsieh, T.-H., Yeh, W., & Tan, K.-S. (2013). Libtabe Lexicon. Retrieved from https://sourceforge.net/projects/libtabe/

Hyman, L. M. (2001). The limits of phonetic determinism in phonology: *NC revisited. In E. Hume & K. Johnson (Eds.), *The role of speech perception phenomena in phonology* (pp. 141–185). New York: Academic Press.

Inkelas, S., Orgun, C., & Zoll, C. (1997). The implications of lexical exceptions for the nature of grammar. In I. Roca (Ed.), *Derivations and constraints in phonology* (pp. 393–418). Oxford: Clarendon Press.

Johnsen, S. S. (2012). A diachronic account of phonological unnaturalness. *Phonology*, *29*(3), 505–531.

Jusczyk, P. W., & Luce, P. A. (1994). Infants′ Sensitivity to Phonotactic Patterns in the Native Language. *Journal of Memory and Language*, *33*(5), 630–645.

Kager, R., & Pater, J. (2012). Phonotactics as phonology: knowledge of a complex restriction in Dutch. *Phonology*, *29*(1), 81–111.

Kawahara, S. (2008). Phonetic naturalness and unnaturalness in Japanese loanword phonology. *Journal of East Asian Linguistics*, *17*(4), 317–330.

Kessler, B., & Treiman, R. (1997). Syllable Structure and the Distribution of Phonemes in English Syllables. *Journal of Memory and Language*, *37*(3), 295–311.

Kirby, J. P., & Yu, A. C. L. (2007). Lexical and phonotactic effects on wordlikeness judgments in Cantonese. *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007)*, *21*(August), 1389–1392.

Leben, W. (1973). *Suprasegmental Phonology*. MIT.

Lin, Y.-H. (1989). Autosegmental treatment of segmental processes in Chinese phonology.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing Spken Words: The Neighborhod Activation Model. *Ear and Hearing*, *19*(1), 1–36.

McCarthy, J. J. (1986). OCP Effects: Gemination and Antigemination. *Linguistic Inquiry*, *17*(2), 207–263.

Moreton, E. (2008). Analytic bias and phonological typology. *Phonology*, *25*(1), 83–127.

Myers, J. (1995). Nonlocal Dissimilation in Mandarin Syllables. In *The 7th North American Conference on Chinese Linguistics*. University of Wisconsin at Madison.

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

Ohala, J. J. (1981). The listener as a source of sound change. *Papers from the Parasession on Language and Behavior Chicago Linguistic Society*.

Ohala, J. J. (1986). Consumer's guide to evidence in phonology. *Phonology Yearbook*, *3*(1986), 3–26.

Onishi, K. H., Chambers, K. E., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition*, *83*(1).

Pater, J., & Tessier, A. (2005). Phonotactics and alternations: Testing the connection with artificial language learning. *In: K. Flack and S. Kawahara (Eds.)*, 1–16.

Pierrehumbert, J. B. (2006). The statistical basis of an unnatural alternation. *Laboratory Phonology 8, Varieties of Phonological Competence*, 81–107.

Pycha, A., Nowak, P., Shin, E., & Shosted, R. (2003). Phonological rule-learing and its implications for a theory of vowel harmony. *Proceedings of the 22nd West Coast Conference on Formal Linguistics (WCCFL 22)*, (January 2003), 101–114.

Scheer, T. (2015). *How Diachronic is Synchronic Grammar?* (P. Honeybone & J. Salmons, Eds.), *The Oxford Handbook of Historical Phonology*. Oxford: Oxford University Press.

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

Vitevitch, M. S., & Luce, P. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, *40*(3), 374–408.

Vitevitch, M. S., & Luce, P. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 481–487.

Vitevitch, M. S., & Luce, P. A. (1998). *When words compete: Levels of processing of spoken words*.

Psychological Science (Vol. 9).

Wang, H. S. (1998). An experimental study on the phonotactic constraints of Mandarin Chinese. In B. K. T'sou (Ed.), *Studia Linguistica Serica* (pp. 259–268). Hong Kong: Language Information Sciences Research Center, City University of Hong Kong.

Wang, H. S., & Chang, C. (2001). On the Status of the Prenucleus Glide in Mandarin Chinese. *Language and Linguistics*, *2*(2), 243–260.

White, J., & Chiu, F. (n.d.). Disentangling phonological well-formedness and attestedness: An ERP study of onset clusters in English.

Wiener, S., & Turnbull, R. (2016). Constraints of Tones, Vowels and Consonants on Lexical Selection in Mandarin Chinese. *Language and Speech*, *59*(1), 59–82.

Wiese, R. (1997). Underspecification and the description of Chinese vowels. In J. Wang & N. Smith (Eds.), *Studies in Chinese Phonology* (pp. 219–249). Berlin; New York: Mouton de Gruyter.

Wilson, C. (2003). Experimental Investigation of Phonological Naturalness. *West Coast Conference on Formal Linguistics 22 (WCCFL22)*, 101–114.

Wilson, C. (2006). Learning phonology with substantive bias: an experimental and computational study of velar palatalization. *Cognitive Science*, *30*(5), 945–982.