

Modeling Taiwanese Speakers' Knowledge of Tone Sandhi in Reduplication

Abstract

Recent studies on productivity have shown that native speakers' phonological knowledge not only includes statistical patterns in the lexicon, but also patterns that cannot be gleaned from the lexicon. This is demonstrated in speakers' analytical bias towards abstract phonological representations (Davidson 2005), single-feature dependency (Moreton 2008), perceptually motivated phonological scales (Zuraw 2007), as well as their difficulties with exceptionless opaque patterns in wug tests (Zhang and Lai 2008). Based on the results of a wug test, we show in this paper that the opaque tone sandhi pattern in Taiwanese reduplication is a case in which the speakers' knowledge is a combination of more than, less than, and exactly what their lexicon informs them of: the phonetic effects of the sandhis are overlearned, the opaque tone sandhis are underlearned, and the lexical statistics are properly learned. We further argue that a substantively biased Maximum Entropy grammar (Wilson 2006) that encodes learning biases against lexical listing constraints *à la* the dual listing/generation theory of Zuraw (2000) and phonetically unmotivated patterns can model the simultaneous underlearning, overlearning, and proper learning of the lexical patterns by Taiwanese speakers.

Key words: Taiwanese tone sandhi, opacity, reduplication, wug test, productivity, Maximum Entropy grammar

1. Introduction

In a recent trend, modern phonological research is moving beyond the assumptions that (a) phonological patterns are invariant and exceptionless and (b) it suffices to study lexically manifested sound patterns observable from elicited data.

The abandonment of assumption (a) stems from long-standing observations that many phonological patterns are variable (e.g., Labov 1972, 1994), gradient (e.g., Bolinger 1961, Frisch et al. 2004), and full of exceptions (e.g., Zimmer 1969, Zuraw 2000); more importantly, speakers have phonological knowledge that reflects the variability, gradience, and patterns of exceptionality observed in the lexicon. For example, Frisch and Zawaydeh (2001) showed that Arabic speakers' knowledge of root-internal consonant cooccurrence restrictions, as indicated by wordlikeness ratings of novel forms, reflects gradient patterns in the lexicon that closely relate to consonant similarity. Hayes and Londe (2006) demonstrated that Hungarian speakers can apply the statistical knowledge of vowel harmony gleaned from lexical alternations to novel forms in a wug test (Berko 1958). Zuraw (2000) showed that although Nasal Substitution is riddled with exceptions in the Tagalog lexicon, Tagalog speakers are nonetheless able to detect patterns within the exceptions. For example, speakers exhibit a greater tendency toward Nasal Substitution of a front stem-initial consonant (*p* or *b*) than a back stem-initial consonant (*k* or *g*) for novel stems.

The validity of assumption (b) was challenged by Kenstowicz and Kisseberth (1979), who pointed out that *corpus-internal evidence* such as sound distribution, though highly practical, only indirectly reflects the speakers' knowledge of phonological patterning and should be complemented by *corpus-external evidence* gleaned from non-traditional sources. These sources — language games, speech errors, loanwords, and phonetic and psycholinguistic experimentation — demonstrate that the theories of phonology based on elicited data have psychological reality. Crucially, recent studies have shown that speakers' phonological knowledge does not always match the lexical patterns. There are areas in which speakers seem

to know more than the lexicon. For instance, Davidson (2005) showed that English speakers were not equally accurate in the production of non-native word-initial clusters: their accuracy rate did not correlate with the lexical frequency of these clusters in other positions, but could be accounted for if the speakers made stochastic phonotactic generalizations over abstract phonological representations. Zuraw (2007), in both a corpus study on loans and a web-based survey on novel words, demonstrated that Tagalog speakers possess knowledge of the splittability of word-initial consonant clusters that is not present in the lexicon, but that can be projected from perceptual knowledge, and they can apply the knowledge to infixation in stems with novel initial clusters. Moreton (2008) showed that English speakers can learn a height-height dependency in vowels better than a height-voicing dependency despite the facts that neither dependency is attested in English and that the two dependencies have comparable phonetic precursors, and he attributed the difference to the speakers' analytical bias in favor of single-feature dependencies. These cases present a "poverty of the stimulus" argument for the grammatical treatment of speakers' abstract phonological and phonetic knowledge.

Speakers sometimes also seem to know less than the lexical patterns. For example, Albright and Hayes (2006) argued that accidentally true generalizations in Navajo Sibilant Harmony are unproductive and propose a learning algorithm that prevents such generalizations from being encoded as high-ranking constraints in the grammar. More striking unproductive cases are found in languages with opaque phonological patterns. Sanders (2001) demonstrated that Polish speakers do not extend the counterbleeding interaction between o-Raising and Final Devoicing in nominative singular nouns to novel words in a wug test, even though there are many items in the Polish lexicon that illustrate this pattern. The opaque Taiwanese "tone circle" (see more details in §2) has also been shown repeatedly to be largely unproductive in wug tests despite its exceptionlessness in the language itself (Hsieh 1970, 1975, 1976, Wang 1993, Zhang et al. 2006, 2009a, b, Zhang and Lai 2008).

The goal of this paper is two-fold. Based on the results of a wug test, we first show that the tone sandhi pattern in Taiwanese reduplication is a case in which the speakers' knowledge is a

combination of more than, less than, and exactly what their lexicon informs them of. Regarding the opaque sandhi patterns, the speakers know less than the lexicon, and the exceptionless pattern in the lexicon is only partially productive in novel words. The speakers also know more than the lexicon, as there are indications from the wug-test that they disprefer a tone sandhi that changes a shorter tone to a longer tone (51 → 55) on nonfinal syllables. They could not have deduced this from lexical statistics, but could have deduced it from phonetics, since nonfinal syllables are phonetically shorter than final syllables due to the lack of final lengthening. Finally, transparent phonotactic generalizations such as “the rising tone 24 cannot occur on nonfinal syllables” are learned precisely from the lexicon and can be applied productively to novel words, and certain lexical frequency effects on tones also affect the speakers’ wug-test behavior.

Secondly, we argue that a substantively biased Maximum Entropy (Maxent) grammar (Wilson 2006) that encodes learning biases against lexical listing constraints *à la* the dual listing/generation theory of Zuraw (2000)¹ and phonetically unmotivated patterns can model the simultaneous underlearning, overlearning, and proper learning of the lexical patterns by Taiwanese speakers. With properly set parameters, the learner can learn a grammar that predicts the attested wug test behavior when the learning input only consists of real disyllabic Taiwanese words with or without reduplication with the appropriate lexical frequencies.

2. Tonal patterns in Taiwanese reduplication

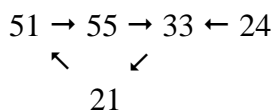
2.1. Taiwanese tone sandhi

Tone sandhi refers to tonal alternations conditioned by adjacent tones or the prosodic or morpho-syntactic position in which the tone occurs (Chen 2000, among others). Taiwanese tone sandhi is typical of Southern Min dialects of Chinese in that it is positionally conditioned and the

¹ The name “dual listing/generation model” was not given by Zuraw herself, but by Hayes and Londe (2006).

sandhi patterns are characterized by circular opacity:² tones in non-XP-final positions undergo sandhi, and four out of the five tones in the tonal inventory are involved in a circular chain shift, as in (1); the XP-final syllable preserves its tone (Chen 1987, Hsiao 1991, Lin 1994).^{3,4} “XP-final” and “non-XP-final” are sometimes abbreviated to “final” and “nonfinal” in the rest of the paper. As stated above, this “tone circle” pattern has been shown by earlier works to be generally unproductive when speakers were wug-tested with novel words (Hsieh 1970, 1975, 1976, Wang 1993, Zhang et al. 2006, 2009a, b, Zhang and Lai 2008), and this result is in line with other experimental works (e.g., Sanders 2001, Sumner 2003) that have demonstrated the lack of productivity of opaque phonological patterns. The sandhi 24 → 33, though still opaque in terms of the sandhi tone, is phonotactically transparent, as the ban on the rising tone 24 in nonfinal positions is a true generalization in Taiwanese. Wang (1993), Zhang et al. (2006, 2009a, b), and Zhang and Lai (2008) have shown that although Taiwanese speakers do not always give the expected sandhi tone 33 in novel words due to the opacity of the structural change, they are quite consistent in abiding by the phonotactic generalization that a rising tone cannot occur nonfinally.

(1) Taiwanese tone sandhi in nonfinal positions:



² We use “opacity” here in the Kiparskian sense (Kiparsky 1973). A phonological rule P of the form A→B/C__D is opaque if there are surface structures with any of the following characteristics: (a) instances of A in the environment C__D, or (b) instances of B derived by P that occur in environments other than C__D. Additional discussion on the opaque nature of Taiwanese tone sandhi, as prompted by an anonymous reviewer, to whom we are grateful, can be found in §6.

³ In this paper, we are only concerned with unchecked syllables, which are open syllables or syllables closed with a sonorant coda. The five tones given in (1) reflect the tonal inventory of this type of syllable in Taiwanese. Checked syllables — syllables closed with a /ʔ/ or /p, t, k/ — have a reduced tonal inventory (53 and 21) and different tone sandhi behaviors that also involve opacity. Detailed descriptions of their behaviors can be found in Chung (1996) and Cheng (1997). We do not attempt an analysis for these patterns here.

⁴ That the right edge of the tone sandhi domain coincides with the right edge of XP is an oversimplification of the tone group formation in Taiwanese. Chen (1987) argues that right edge of the tone sandhi domain is the right edge of every XP, except when XP is an adjunct c-commanding its head (p.131). Lin (1994) further refines Chen’s condition to say that XP is “not lexically governed” (p.248), where lexical government is defined on conditions of m-command and barriers. Details of the tone group formation in Taiwanese do not affect the claims made here.

2.2. Taiwanese reduplication

Reduplication is a productive morphological process in Taiwanese. The meaning of reduplication, depending on the part of speech being reduplicated, includes intensification, tentativeness, repetition, and continuity (Wu 1996: p.27-29). The object of study in this paper is the single reduplication of monosyllabic adjectives, which diminishes the intensity of the adjective. This reduplication is exemplified in (2) (from Chung 1996: p.129). As we can see in (2), the tonal pattern of this reduplication follows the general tone sandhi pattern — the initial reduplicated syllable undergoes tone sandhi according to (1).

(2) Taiwanese monosyllabic adjective reduplication:

Monosyllabic adjective	Single reduplication 'somewhat adj.'	Gloss
p ^h oŋ21	p ^h oŋ51-p ^h oŋ21	'blown-up'
nŋ51	nŋ55-nŋ51	'soft'
sin55	sin33-sin55	'new'
kau33	kau21-kau33	'thick'
tam24	tam33-tam24	'wet'

There are many possible theoretical treatments for this reduplication. We may treat the reduplicant as an abstract input morpheme RED and let the phonological grammar compute the output representation of RED, *à la* McCarthy and Prince's (1993, 1995, 1999) Prosodic Morphology. Under this approach, the reduplicant is better treated as a prefix due to the partly neutralizing sandhi in nonfinal position (55 and 24 are neutralized to 33): had the winning output been one with suffixing reduplication, we would have had to assume that there existed Input-Reduplicant (IR) correspondence in order to distinguish the reduplicated forms of /55/ and /24/, as both would have the base [33] and hence not distinguishable by Base-Reduplicant (BR) correspondence; this contradicts McCarthy and Prince's basic model of reduplicative

correspondence. It is also possible to treat the reduplication as morphological doubling *à la* Inkelas and Zoll (2005) and let cophologies derive the nonfinal and final syllables. Finally, Yip's (1998) approach whereby the reduplication is treated as compounding that results from a highly ranked REPEAT constraint, is also viable. For concreteness sake, we assume McCarthy and Prince's Prosodic Morphology and treat the final syllable as the base and the nonfinal syllable as the reduplicant in the output.

The different sandhi processes exhibited by this reduplication differ in two parameters. First, as discussed in §2.1, the sandhis in the tone circle are entirely opaque, while the 24 → 33 sandhi is phonotactically transparent. Second, within the opaque tone circle, the sandhis have different phonetic bases that relate to the duration of the tones. Studies by Lin (1988) and Peng (1997) have shown that the two falling tones 51 and 21 have considerably shorter intrinsic durations than 55, 33, and 24 in Taiwanese.⁵ Given that the sandhi occurs on nonfinal syllables, which are known to be shorter than final syllables due to the lack of final lengthening (Oller 1973, Klatt 1975, Wightman et al. 1992, among others), the 33 → 21 sandhi has a durational basis, as it is duration-reducing; the 51 → 55 sandhi is a durationally disadvantaged change, as it is duration-increasing; the other two sandhis are durationally neutral. Due to the differences in these two parameters, we expect that the speakers' knowledge of the sandhi patterns may exhibit both underlearning and overlearning from the lexicon: the opaque patterns may be underlearned despite their exceptionlessness in the lexicon as compared to the transparent sandhi, but the durational properties of the sandhis may cause an overlearning effect in that the productivity of

⁵ A reviewer requested clarification on whether the durational differences are (a) universal or only specific to Taiwanese, and (b) phonological or phonetic. We take these durational differences to be specific properties of Taiwanese tones, not universal properties of tones in general. The "phonology" vs. "phonetics" question, however, is not so easy to answer. Together with the even shorter duration of tones on checked syllables, these durational properties of the tones are predictable from tonal and segmental contexts and are thus clearly not contrastive. One may wish to consider these effects "phonetic," especially given the three levels of duration that must otherwise be encoded in phonology. But these durational effects have considerable effects on phonological behavior: the shorter duration on checked syllables induces a reduction of the tonal inventory, and as we will see later in the paper, the durational difference between 51, 21 and 55, 33, 24 influences the productivity of the tone sandhi patterns. This goes to the heart of the phonetics-phonology distinction debate. See Blevins (2004), Boersma (1998), Flemming (2001), Hansson (2008), Hayes and Steriade (2004), Hyman (2001), Ohala (1997), Steriade (2008), among many others, on different positions regarding this issue and Zhang (2002) on how the issue applies to tone.

sandhis with a stronger phonetic basis may be higher. Both of these effects have been shown in Zhang et al. (2006, 2009a, b) and Zhang and Lai (2008).

The relevance of the durational argument requires further clarification. Zhang (2002) provides a crosslinguistic survey that establishes the advantage afforded by prosodic-final positions for contour tone bearing. It should be recognized, however, that the sandhi domain in Taiwanese is not a prosodic domain, but a syntactic domain (Chen 1987, Lin 1994; see footnote 4). Chen (1987) specifically provides examples that show that the boundaries of the sandhi domains do not coincide with intonational phrase boundaries. This begs the question whether final lengthening can in fact be used to motivate the sandhis phonetically. However, unlike non-XP-final syllables, syllables that occur XP-finally have more opportunities to appear before a large prosodic boundary. Therefore, the average duration of an XP-final syllable in connected discourse will be greater than the same syllable in non-XP-final positions. Myers and Tsay (2008), for example, showed that the duration of a base tone in XP-final position is significantly longer than the duration of a sandhi tone in non-XP-final position even though the two syllables under comparison both appear medially in two sentences that are matched for the number of syllables and segmental contexts. A similar point was made by Zhang (2003/2009) to argue for the positive effect of word-final position on contour tone licensing despite its limited degree of lengthening as compared to phrase-final syllables.

Therefore, the tonal patterns of Taiwanese reduplication is an excellent test case for the interaction among the effects of lexical and non-lexical factors on phonological productivity, which we take to be a more accurate reflection of the phonological grammar than the corpus-internal evidence gleaned from the patterns in the language itself. Provided that such interaction can be demonstrated, it will also provide us with a novel case — one in which underlearning from the lexicon due to opacity coexists with overlearning and proper learning — to test both Zuraw's (2000) theory of patterned exceptionality and Wilson's (2006) substantively biased Maxent grammatical model. The productivity of reduplication as a morphological process also makes it a particularly easy grammatical configuration with which to wug test speakers.

Provided that the speakers can interpret the wug stimuli as real adjectives in Taiwanese, they are expected to be able to reduplicate the stimuli to form adjectives of a lesser degree.

The next section lays out the details of a wug test on the tone sandhi patterns in Taiwanese reduplication that aims to investigate the factors that affect their productivity. Section 4, then, proceeds to model the speakers' learning process that predicts the observed productivity using a Maxent model that encodes both dual listing/generation of Zuraw (2000) and substantive biases of Wilson (2006). Section 5 provides a general discussion of open questions raised by our analysis. Section 6 evaluates alternative analyses. Section 7 concludes the paper.

3. A wug test

3.1. Experimental methods

3.1.1. Stimuli construction

The design of our experiment followed the wug test paradigm pioneered by Berko (1958) and later widely used by researchers to test the productivity of regular and irregular morphological rules (e.g., Bybee and Pardo 1981, Albright 2002, Albright and Hayes 2003) and morphophonological alternations (e.g., Hsieh 1970, 1975, 1976, Wang 1993, Zuraw 2000, 2007, Albright et al. 2001, Hayes and Londe 2006, Pierrehumbert 2006). For our experiment, we asked native Taiwanese speakers to first listen to monosyllabic words through headphones, and then produce their reduplication forms.

Three types of monosyllabic words were used in the experiment: real words with **Actual Occurring** reduplication (AO), real words with no **Actual Occurring** reduplication (*AO), and **Accidental Gap** words (AG), which have a legal segmental composition and a legal tone in Taiwanese, but an accidentally non-existing segmental-tone combination in the Taiwanese

syllabary.⁶ For each type of word, all five tones that can appear on open and sonorant-closed syllables (21, 51, 55, 33, 24) were used, and eight different words were selected for each Word Type × Tone combination. This made a total of 120 test words. To facilitate the identification and measurement of the tone on the rhyme portion of the syllable, the onset of each of the test syllable was selected to be an obstruent. No other segmental or frequency controls were made in the stimuli construction. A sample of the word list is shown in (3), and the full word list is given in the Appendix. The word list was recorded by the second author, who is a phonetically trained native Taiwanese speaker.

(3) Sample word list for the wug test experiment:

Tone	AO			*AO			AG
21	怪	kuai	‘strange’	進	t̂sin	‘to enter’	t̂soo
51	飽	pa	‘full’	炒	t̂s ^h a	‘to stir fry’	p ^h au
55	芳	p ^h aŋ	‘fragrant’	區	k ^h u	‘area’	t̂san
33	鈍	tun	‘blunt’	電	tian	‘electricity’	k ^h ŋ
24	濕	tam	‘wet’	頭	t ^h au	‘head’	p ^h āi

3.1.2. Experimental set-up and data analyses

The experiment was conducted with SuperLab (Cedrus). Through a headphone, each subject first heard an instruction in Taiwanese, which explained in layperson’s terms and through examples that Taiwanese can mark the lesser intensity of an adjective by reduplication, and their task was to form the diminutive forms of the words they would hear from the headphone. They were also told that some of the words were not real words in Taiwanese, but they should form the diminutive forms just like they would for real words. During the test, the 120 stimuli were presented in randomized order to each subject, and after each monosyllabic stimulus, the subject

⁶ The naming of the different groups follows Hsieh (1970, 1975, 1976).

produced the reduplicated form. The subjects' responses were recorded by a Marantz solid state recorder PMD 671 at 16 bits and a 44.1KHz sampling rate.

3.1.3. Subjects

Our data were collected during two field trips to Taiwan by the second author in the summers of 2006 and 2007. The field site in 2006 was Chiayi, where the second author recruited ten speakers and conducted the experiment in the Phonetics Laboratory in the Institute of Linguistics at National Chung Cheng University. The field site in 2007 was a town named Paihe near Tainan, where the second author recruited four speakers and conducted the experiment in a quiet room. Two of the speakers in Chiayi did not understand the instructions and simply repeated each monosyllabic stimulus twice in two separate phrases with a large silent interval between the two syllables. These two speakers' data were not used in subsequent data analyses. Among the twelve speakers whose data we did analyze, two were male, ten were female, and they had an average age of 50.75 at the time of the experiment.

3.1.4. Data analyses

Due to the structure-preserving nature of Taiwanese tone sandhi (Tsay and Myers 2001, Tsay et al. 1999, Myers and Tsay 2008), the tonal responses to the experimental stimuli were transcribed in Chao letters by the three authors of the paper — one native Taiwanese speaker, one native Mandarin speaker, and one native American English speaker with phonetic training. The first two authors participated in the transcription for all subjects; the third author only transcribed the first set of subjects from Chiayi. The agreement based on first-time auditory impression among the authors was high — around 95%. With help from pitch tracks in Praat (Boersma and Weenink 2005), we agreed on virtually all tokens. For the handful of tokens on which we did not reach an agreement, we took the native Taiwanese author's judgment.

Due to our error in setting up the SuperLab experiment, one word each for the tones 21, 24, and 33 for the *AO group could not be used for analysis. We excluded these tokens for all subjects in subsequent statistical analyses.

3.1.5. Hypotheses

The hypotheses for the experiment are as follows.

First, the subjects' correct response rates⁷ for the sandhi patterns should be low for novel reduplications (real words without actual occurring reduplication (*AO) and accidental gaps (AG)) due to the opacity in the pattern, indicating underlearning from the lexicon.

Second, among the opaque sandhis, the durational property of the sandhi may have an effect on its productivity, such that a duration-increasing sandhi is less productive than a duration-reducing sandhi, indicating overlearning from the lexicon.

Third, for the phonotactically transparent sandhi 24 → 33, although the speakers might not always provide the correct sandhi tone 33 due to the output opacity, they should avoid a surface 24 on the initial syllable of the reduplicated form.

Fourth, lexical frequencies of both the individual tones and tonal melodies associated with reduplication may also affect productivity, in that more frequent tones or tonal melodies are more productive. Frequency information on base tones collected from a spoken Taiwanese corpus that includes 355,790 word tokens (Tsay and Myers 2005) is summarized in (4). In (4), the syllable type frequency of a tone refers to the number of different syllables in the Taiwanese syllabary that can carry the tone; the morpheme type frequency refers to the number of different monosyllabic morphemes that can carry the tone; and the token frequency is the number of occurrences of the tone in the entire corpus. Frequency information on the five tonal melodies

⁷ “Correct response” here simply means that the subject’s response agrees with the tone sandhi pattern laid out in (1). It is merely a convenient way to refer to the sandhi tones in the regular pattern in the Taiwanese lexicon, not a judgment on the “correctness” of the subject’s sandhi application. Our goal is to provide a model that accounts for the subjects’ sandhi behavior in the wug test without any judgment on whether their behavior is “correct” or not.

attested in reduplication is given in (5), based on 5630 tokens of reduplication representing 163 types in the corpus. It is crucial to note that none of these frequency scales correlates with the duration nature of the sandhi; i.e., it is not the case that a more frequent tone undergoes a durationally more advantageous sandhi.

The last two hypotheses are based on proper learning from the lexicon.

(4) Lexical frequencies of tones:

- a. Syllable type frequency: 55 > 51 > 24 > 21 > 33
- b. Morpheme type frequency: 55 > 24 > 51 > 33 > 21
- c. Token frequency: 55 > 24 > 33 > 51 > 21

(5) Frequencies of tonal melodies in reduplication:

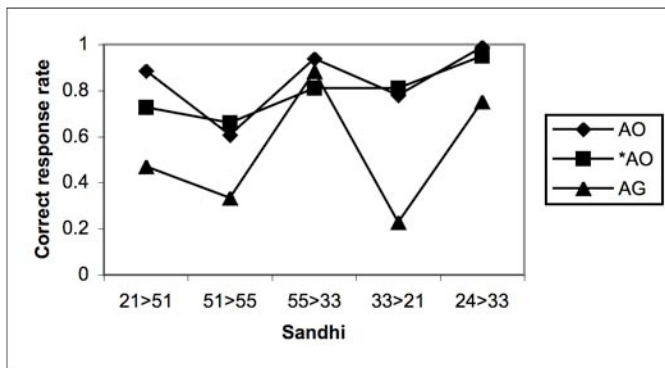
- a. Type frequency: 33-55 > 55-51 > 33-24 > 51-21 > 21-33
- b. Token frequency: 33-24 > 55-51 > 33-55 > 51-21 > 21-33

3.2. Results

The response rates for correct sandhi patterns are given in (6). The subjects have answered with the correct sandhi tone on the first syllable and kept the base tone on the second syllable unchanged in these cases. In (6) and following graphs, AO stands for existing Taiwanese syllables with actual occurring reduplications, *AO stands for existing Taiwanese syllables without actual occurring reduplications, and AG stands for accidental gaps in the Taiwanese syllabary. A Two-Way Huynh-Feldt Repeated-Measures ANOVA was conducted for Word Type (AO, *AO, AG) and Tone (21, 51, 55, 33, 24), and it indicated that the effect of Word Type is significant ($F(2.000, 22.000) = 70.481, p < 0.001$), so is the effect of Tone ($F(3.109, 34.202) = 15.132, p < 0.001$). The interaction between Word Type and Tone is also significant

($F(8.000, 88.000) = 9.230, p < 0.001$). Bonferroni posthoc tests indicated that the subjects performed the sandhis significantly less accurately (i.e., according to the pattern in (1)) for novel words AG than AO ($p < 0.001$) and *AO ($p < 0.001$), but there is no significant difference between AO and *AO ($p > 0.05$). Posthoc tests also showed that $55 \rightarrow 33$ and $24 \rightarrow 33$ have higher correct response rates than all other sandhis respectively (all $p < 0.05$), but there are no differences between $55 \rightarrow 33$ and $24 \rightarrow 33$ ($p > 0.05$), or among the other sandhis (all at $p > 0.05$).

(6) Correct response rates for sandhi patterns:



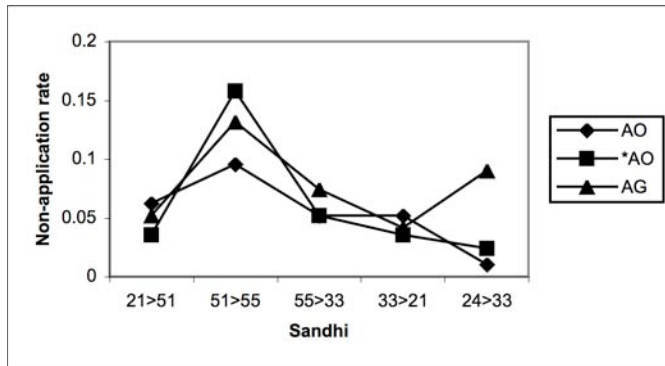
Among the responses that did not conform to the sandhi patterns in (1), two types of responses constitute the vast majority: (a) Non-application: the sandhi simply did not apply on the first syllable, i.e., the first syllable kept the base tone, and the second syllable also kept the base tone. (b) Template matching: the first syllable kept the base tone, but the second syllable underwent sandhi to result in an attested disyllabic melody for reduplication. For instance, if the monosyllabic stimulus was 51, then a subject’s response 51-51 would be an example of (a), and a response 51-21 would be an example of (b), as 51-21 is an attested tone pattern on reduplication, namely, for 21. The table in (7) lists the responses that fall under the “template matching” category for all base tones.

(7) “Template matching” responses:

Stimulus	Response
21	21-33
51	51-21
55	55-51
33	33-55, 33-24
24	—

The non-application rates of the sandhi patterns are given in (8). The effect of Word Type is not significant ($F(1.598, 17.574) = 0.996, p > 0.05$), nor so is the effect of Tone ($F(1.923, 21.156) = 2.808, p > 0.05$). The interaction between Word Type and Tone is also not significant ($F(7.062, 77.681) = 1.080, p > 0.05$). However, we observe two trends in the data: (a) the non-application rate is the greatest in AG words, followed by *AO and then AO; and (b) the durationally aberrant sandhi 51 → 55 has a higher non-application rate than the phonotactically transparent 24 → 33 and the durationally motivated 33 → 21.

(8) Non-application rates of sandhi patterns:

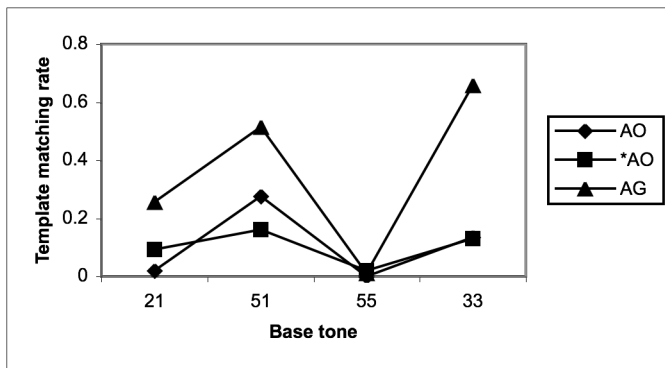


The template matching rates for different base tones are given in (9). For base tone 33, responses 33-55 and 33-24 were pooled together. Base tone 24 does not appear in the graph as there is no applicable template for 24. The effect of Word Type is significant ($F(2.000, 22.000) = 58.220, p < 0.001$), and so is the effect of Tone ($F(2.270, 24.967) = 12.707, p < 0.001$). The interaction between Word Type and Tone is also significant ($F(4.822, 53.047) = 9.602, p < 0.001$).

Posthoc tests indicated that the template matching rate is significantly higher in AG words than in AO and *AO words ($p < 0.001$), but there is no difference between AO and *AO ($p > 0.05$).

Posthoc tests also showed that base tone 51 elicited greater template matching responses than 21 ($p < 0.05$) and 55 ($p < 0.01$) and base tones 33 and 21 elicited greater template matching responses than 55 (both at $p < 0.01$).

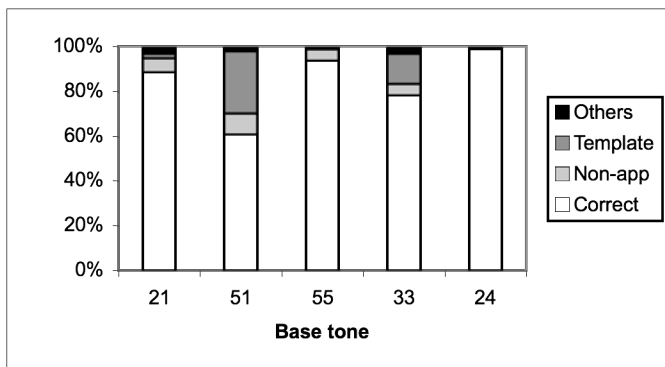
(9) Template matching rates for base tones:



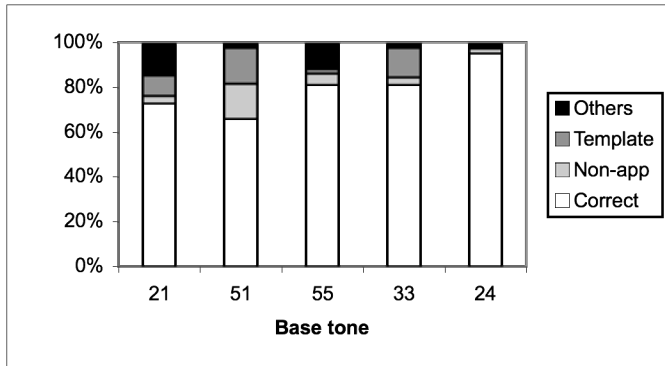
The rates for the three different categories of sandhi responses along with other changes are graphed together for AO, *AO, and AG in the bar-graphs in (10).

(10) Summaries for different categories of sandhi application:

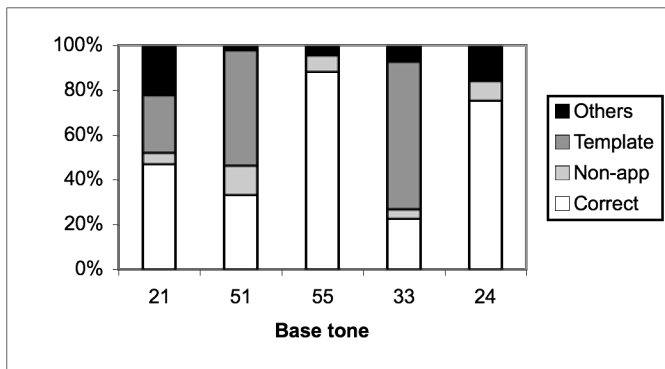
a. AO:



b. *AO:



c. AG:



3.3. Discussion

Our experimental results are considerably more complex than our hypotheses. But we did find various supports for our hypotheses, as discussed below.

First, although the opaque tone circle is not entirely unproductive in novel reduplications — for *AO and AG words, the correct response rates for sandhi tones in the tone circle in fact range from 20 to 90% — the sandhi productivity is significantly lower in non-existing syllables (AG) than real syllables (AO, *AO). This indicates that there is indeed underlearning from the lexicon when the input-output mapping is opaque.

Second, among the opaque sandhis, the durational property of the sandhi may have had an effect on productivity: the low productivity of 51 → 55 cannot be due to lexical frequency, as

both 55 and 51 as well as the tonal melody 55-51 occur relatively frequently, but could be due to the duration increasing nature of the sandhi. In addition, there is a trend for base 51 to have both a greater non-application rate and a greater template matching rate, possibly due to a combined result of preferring the short 51 and dispreferring the sandhi that changes the shorter 51 into the longer 55 in nonfinal position. These are instances of overlearning from the lexicon.

Comparing the two sandhis with the opposite durational effect — the duration increasing 51 → 55 and the duration reducing 33 → 21, we can see that the durational effect is asymmetrical: a sandhi that introduces additional duration, and hence additional markedness violation, in the output is clearly disfavored, as evidenced by the low productivity of 51 → 55, but a sandhi that reduces duration, and hence reduces markedness violations, in the output is only afforded marginal advantage, as suggested by the equally low productivity of 33 → 21. The low productivity of 33 → 21 is interpreted as resulting from its low type frequency, as discussed below. This asymmetrical effect is a gradient version of the “grandfather effect” *à la* McCarthy (2002), where a marked structure is permitted when it is originally in the input, but banned if it is derived in the output.⁸

Third, the phonotactic generalization that no rising tone can appear on nonfinal syllables is generally productive even in wug words. The sandhi 24 → 33 applied significantly more

⁸ An anonymous reviewer raises the possibility that the duration effect may not necessarily be manifested as the overall preference of having a shorter tone on nonfinal syllables, but as the preference for having an output in which the nonfinal syllable is shorter relative to the final syllable, thus mimicking final lengthening. Therefore, if the second syllable is a short 51 or 21, a 33 → 21 change may be more likely than if the second syllable is a longer 55, 33, or 24. This hypothesis cannot be directly tested in the experimental results here as the reduplicative nature of the outputs prevents us from being able to compare [21-33] with [21-51] and [21-21], as the latter forms are not the correct outputs of any reduplications. But indirect evidence suggests that this is likely not the case. If a short-long output in comparative terms is the primary manifestation of the duration effect in tone sandhi, then we would reasonably expect that in tone sandhi systems where nonfinal syllables undergo tonal neutralization, this neutralization would depend on the durational property of the final syllable in the form of an implicational statement: if neutralization occurs when the final syllable has longer duration, then it will also occur when the final syllable has shorter duration. But in typological studies of tones (e.g., Chen 2000, Hyman 2007, Hyman and Schuh 1974, Zhang 2002), no implicational statement as such has ever been established. On the other hand, Zhang (2002) established that if a nonfinal syllable can carry contour tones, then a final syllable can also carry contour tones when all else is equal, and the dispreference of contour tones on nonfinal syllables does not seem to be contingent upon the durational property of the final syllable. Moreover, in similar wug test experiments that test non-reduplicative disyllables (Zhang et al. 2006, 2009a, b), discrepancies in the rate of sandhi application based on the second syllable have not been reported.

accurately than other sandhis except for 55 → 33, and its non-application rate shows a trend of being lower than all other sandhis except for 33 → 21, which has a durational advantage. Moreover, given that there is no template matching response for base 24, the rate by which it changes into another tone in nonfinal position reaches 95.7%, which is considerably higher than that for any other base tone, which ranges from 55.4% (base 51) to 92.5% (base 55).

Lastly, we found some support for the relevance of lexical frequencies to productivity. The low productivity associated with 33 → 21 cannot be due to phonetic duration, as the sandhi is duration-reducing, but could potentially be related to the low lexical frequencies of the base tone 33 and the reduplicative melody 21-33. The higher rate of the 55 → 33 application also cannot be accounted for by the duration property of the tones, as 55 and 33 have comparable intrinsic duration, but may be due to the high lexical frequencies of the base tone 55. The frequency effects here are particularly well matched with the syllable type frequencies of lexical tones (4) and the type frequencies of disyllabic tonal melodies in reduplication (5), as for both frequency counts, 55 → 33 is the highest, and 33 → 21 is the lowest.⁹

Therefore, we contend that the results from our wug-test experiment show that the tone pattern of Taiwanese reduplication indeed represents a case in which under-, over-, and proper learning from the lexicon coexist. In the next section, we turn to the task of modeling the speakers' learning of the sandhi patterns in order to predict their variable behavior in the wug test.

⁹ This result is in general agreement with earlier research, which established that productivity was usually positively correlated with type frequency (Baayen 1992, 1993, Bybee 1985, 2001, Ernestus and Baayen 2003, Moder 1992, Pierrehumbert 2003, 2006, among others). However, it disagrees with Zhang et al. (2009a, b)'s results on tone sandhi productivity in Taiwanese disyllables, which showed a token frequency effect for speakers tested in Taiwan. We surmise that this discrepancy is due to the fact that Zhang et al. (2009a, b) tested the tone sandhi in disyllables, while we tested the sandhi in reduplication. The sandhis in disyllables are instantiated by many types of syllables, morphemes, and words, and the type frequency effect may have reached a ceiling. Reduplication, however, is only applicable to a few dozen syllable types for each tone, which may have caused the type frequency to show an effect despite the fact that the sandhi pattern is the same in reduplication as in disyllables. This issue warrants further studies.

4. A theoretical model of the speakers' wug-test behavior

Our theoretical model of the speakers' behavior is faced with the following challenges. First, it must be able to account for both the highly productive tone sandhi behavior in the Taiwanese lexicon and the variable tone sandhi behavior in wug words (underlearning). Second, it should be able to model the durational effect on productivity (overlearning). And finally, it needs to ensure that transparent and exceptionless generalizations are highly productive and lexical statistics can have its effect felt in the wug test (proper learning).

4.1. The Maximum Entropy (Maxent) model

The Maxent model of phonological grammar (Goldwater and Johnson 2003, Wilson 2006, Jäger 2007, Hayes and Wilson 2008) is equipped to handle the challenges laid out above. As a variant of Optimality Theory, it is inspired by conditional random fields in information theory (Della Pietra *et al.* 1997, Lafferty *et al.* 2001) and closely related to Harmonic Grammar of Smolensky (1986) and Smolensky and Legendre (2006).

In lieu of ranking the constraints on a linear scale, the Maxent model assumes that each constraint C_i has a weight of w_i . For a candidate y of a given input x , if $C_i(y|x)$ is the number of times y violates the constraint C_i , then the Harmonic Score of y given x is defined as e to the negative power of the weighted sum of the numbers of violations of all constraints, as shown in (11).

$$(11) \quad H(y | x) = \exp\left(-\sum_i w_i C_i(y | x)\right)$$

Given the Harmonic Score of y , the probability of y as the output to x is then the proportion of its Harmonic Score to the sum of all Harmonic Scores of x 's candidates, as shown in (12).

The probability distribution of all candidates defined as such maximizes the entropy of the distribution (13) — a concept commonly used in information theory.

$$(12) \quad p(y | x) = \frac{H(y | x)}{\sum_{y \in \Omega} H(y | x)}, \quad \Omega \text{ is the set of candidates for } x.$$

$$(13) \quad \text{Entropy}(p(y | x)) = \sum_{y \in \Omega} p(y | x) \log \frac{1}{p(y | x)}, \quad \Omega \text{ is the set of candidates for } x.$$

Given the constraint set, learning in a Maxent model from a set of training data D composed of input-output pairs is to determine the constraint weights that maximize the log probability of D , which equals to the log of the product of the probabilities of all input-output pairs in D , as in (14).

$$(14) \quad \log(p(D)) = \log\left(\prod_{y|x \in D} p(y | x)\right)$$

To prevent overfitting the training data, each weight is often considered to be associated with a regularizing Gaussian prior (Johnson et al. 1999, Martin et al. 1999, Chen and Rosenfeld 2000). The prior specifies μ as the default weight for a constraint, and σ^2 as the determinant of how severe the penalty is if the weight deviates from the default — the smaller the σ^2 , the greater the penalty. Learning, then, is to find the constraint weights that maximize the function combining $\log(p(D))$ and the penalty, as shown in (15). Crucially, learning *biases* can be encoded as different σ^2 s for different constraints, as we will see later in §4.

$$(15) \quad \log\left(\prod_{y|x \in D} p(y | x)\right) - \sum_i \frac{(w_i - \mu_i)^2}{2\sigma_i^2}$$

4.2. The dual listing/generation model

We will also base our theory on the dual listing/generation model of Zuraw (2000). This model assumes that existing forms are lexically listed and are protected by highly-ranked faithfulness constraints, but lower and stochastically-ranked constraints can encode both patterns of lexical statistics and phonetically-based generalizations. Transparent patterns in the lexicon can be derived through highly-ranked markedness constraints as well as lexical listing, and these high-ranking markedness constraints will also ensure that the patterns are productive in novel words. Opaque patterns, however, cannot be derived by high-ranking markedness constraints and rely on lexical listing. Speakers do make certain generalizations about the opaque input-output mappings and encode them as constraints. But these constraints are ranked relatively low in the grammar, preventing the patterns from being completely productive in novel words. In Maxent terms, the ranking values of constraints are expressed as weights, but the intuitions that grammatical patterns are derived from both lexical listing constraints and the interaction between markedness and faithfulness, and that opaque generalizations can be encoded as input-output listing constraints with low weights, are similar.

4.3. Constraints

4.3.1. Constraints governing reduplication

Let us now consider the inventory of relevant constraints for our grammatical model. A correspondence constraint that requires the tonal identity between the base and the reduplicant, as defined in (16), is plainly relevant.

- (16) IDENT-BR(Tone): The tones of the base and the corresponding reduplicant must be identical.

Assuming the Basic Model of reduplicative correspondence in McCarthy and Prince (1995), we also need an input-output mapping constraint regarding tone — IDENT-IO(Tone), as defined in (17).

- (17) IDENT-IO(Tone): The tones of the input and the base must be identical.

In our discussion of the lexical patterns of reduplication, we assumed the theory of Prosodic Morphology (McCarthy and Prince 1993, 1995, 1999) and argued that the reduplicant is to the left of the base due to the partially neutralizing tone sandhi that applies in nonfinal position. This indicates the relevance of a constraint that forces the reduplicant to be a prefix — ALIGN(RED, L, WORD, L), as defined in (18a). But the abundance of template matching cases in the wug test shows that it is also likely that the speakers have treated the initial syllable as the base and the final syllable as the reduplicant. This means that the constraint that forces the reduplicant to be a suffix — ALIGN(RED, R, WORD, R), as defined in (18b) — may also be relevant, even though it shows no effect in the lexicon.

- (18) a. ALIGN(RED, L, WORD, L) (abbr. ALIGN-L): The left edge of the reduplicant must be aligned with the left edge of the word.
- b. ALIGN(RED, R, WORD, R) (abbr. ALIGN-R): The right edge of the reduplicant must be aligned with the right edge of the word.

4.3.2. USELISTED constraints

With the assumption that speakers are able to use real words with existing reduplications correctly in real life, but do not always use the correct sandhi patterns in *AO and AG words as the experimental results indicate, the theory needs to encode three different levels of listedness: the listedness of reduplicated forms for real words with existing reduplications, the listedness of the allomorphic relation between final syllables and their corresponding nonfinal syllables for existing syllables (e.g., Final pa55 → Nonfinal pa33), and the listedness of the allomorphic relationship between final tones and nonfinal tones independent of segmental contexts (e.g., Final 55 → Nonfinal 33). The first level of listedness will account for the speakers' correct usage of existing reduplications. The second level will account for the fact that the sandhis apply more productively in real words without real reduplications (*AO) than wug words (AG) in the experiment, but is also partially responsible for the behavior of AO words in the experiment since an existing syllable for which the speaker did not find a listed reduplicated form, even though such a form exist, is equivalent to a *AO syllable. The third level of listedness accounts for the moderate degree of productivity of tone sandhis in AG words; without the listedness of the allomorphic relationship between final and nonfinal tones independent of segmental contexts, the opaque sandhis would be completely unproductive in AG words.

Therefore, for an existing syllable with an existing reduplications like /pa51/, its reduplication /RED, pa51/ has a listed lexical entry /pa55-pa51/, and the syllable itself also has a listed non-XP-final allomorph /pa55/. An existing syllable without reduplication such as /te21/ has a listed non-XP-final allomorph /te51/, but no listed reduplication if the wug test calls for the production of /RED, te21/. Each tone in the tonal inventory of Taiwanese also has a listed non-XP-final tonal allomorph; for example, /55/ has a listed allomorph /33/ to be used in non-XP-final positions.

To implement the three different types of listedness in the grammar, we define three classes of USELISTED constraints as in (19). The constraints in (19a) require that real words with

existing reduplications use the listed reduplicative forms; the constraints in (19b) require an existing syllable to use its listed allomorph in non-XP-final positions; and the constraints in (19c) require appropriate tonal allomorphs to be used for existing tones in non-XP-final positions regardless of the segmental contents.

(19) USELISTED constraints:

- a. USELISTED(σ_{51} - σ_{21}): Use the listed / σ_{51} - σ_{21} / for /RED, σ_{21} /.
 USELISTED(σ_{55} - σ_{51}): Use the listed / σ_{55} - σ_{51} / for /RED, σ_{51} /.
 USELISTED(σ_{33} - σ_{55}): Use the listed / σ_{33} - σ_{55} / for /RED, σ_{55} /.
 USELISTED(σ_{21} - σ_{33}): Use the listed / σ_{21} - σ_{33} / for /RED, σ_{33} /.
 USELISTED(σ_{33} - σ_{24}): Use the listed / σ_{33} - σ_{24} / for /RED, σ_{24} /.
- b. USELISTED(σ_{21}): Use the listed allomorph / σ_{51} / for / σ_{21} / non-XP-finally.
 USELISTED(σ_{51}): Use the listed allomorph / σ_{55} / for / σ_{51} / non-XP-finally.
 USELISTED(σ_{55}): Use the listed allomorph / σ_{33} / for / σ_{55} / non-XP-finally.
 USELISTED(σ_{33}): Use the listed allomorph / σ_{21} / for / σ_{33} / non-XP-finally.
 USELISTED(σ_{24}): Use the listed allomorph / σ_{33} / for / σ_{24} / non-XP-finally.
- c. USELISTED(21): Use the listed tonal allomorph /51/ for /21/ non-XP-finally.
 USELISTED(51): Use the listed tonal allomorph /55/ for /51/ non-XP-finally.
 USELISTED(55): Use the listed tonal allomorph /33/ for /55/ non-XP-finally.
 USELISTED(33): Use the listed tonal allomorph /21/ for /33/ non-XP-finally.
 USELISTED(24): Use the listed tonal allomorph /33/ for /24/ non-XP-finally.

The tableaux in (20) illustrate how these USELISTED constraints are evaluated against hypothetical inputs and candidates. If the input syllable is an existing word with an existing

reduplication (AO: / σ 21/), then it has a listed reduplicative form (/ σ 51- σ 21/), a listed non-XP-final allomorph (/ σ 51/), and the tone /21/ has a listed non-XP-final allomorph /51/. When the listed reduplication is used, none of the USELISTED constraints is violated; but when the nonfinal syllable maintains the base tone, for example, all three USELISTED constraints are violated. When anything other than the listed reduplication is used, USELISTED(σ 51- σ 21) is always violated. In addition, when the final syllable is used as the base, and the nonfinal reduplicant simply copies the base tone — [σ 21- σ 21] (R-B), it violates both USELISTED(σ 21) and USELISTED(21) as the correct allomorph for the final /21/ is not used in the nonfinal position. When the nonfinal syllable is used as the base, and the final reduplicant simply copies the base tone [σ 21- σ 21] (B-R), it likewise violates both USELISTED(σ 21) and USELISTED(21) for the same reason. as the correct allomorph for the base /21/ is not used in the final position. However, if the nonfinal syllable is used as the base, and the final reduplicant has a 33 syllable that allomorphically corresponds to the nonfinal 21 syllable — [σ 21- σ 33] (B-R), then no other USELISTED constraint is violated.

If the input syllable is an existing word without an existing reduplication (*AO: / σ 21/), then it lacks a listed disyllabic reduplication. Therefore, USELISTED(σ 51- σ 21) is vacuously satisfied by all candidates. Otherwise, the constraint evaluations for the four candidates are identical to those for the AO input.

Finally, if the input is a novel syllable, then it has neither a listed disyllabic reduplication nor a syllable-based allomorph. Thus, USELISTED(σ 51- σ 21), USELISTED(σ 21), and USELISTED(σ 33) are all vacuously satisfied. However, the USELISTED constraints that govern tonal allomorphs abstracted away from segmental contexts are still relevant: USELISTED(21) is violated by [σ 21- σ 21] (R-B and B-R), for example.

(20) The evaluation of USELISTED constraints:

AO: /RED, σ_{21} / Listed: / σ_{51} - σ_{21} / Listed: / σ_{51} -X/ Listed: /51-X/	USELISTED (σ_{51} - σ_{21})	USELISTED (σ_{21})	USELISTED (σ_{33})	USELISTED (21)	USELISTED (33)
σ_{51} - σ_{21} (R-B)					
σ_{21} - σ_{21} (R-B)	*	*		*	
σ_{21} - σ_{21} (B-R)	*	*		*	
σ_{21} - σ_{33} (B-R)	*				

*AO: /RED, σ_{21} / Listed: / σ_{51} -X/ Listed: /51-X/	USELISTED (σ_{51} - σ_{21})	USELISTED (σ_{21})	USELISTED (σ_{33})	USELISTED (21)	USELISTED (33)
σ_{51} - σ_{21} (R-B)					
σ_{21} - σ_{21} (R-B)		*		*	
σ_{21} - σ_{21} (B-R)		*		*	
σ_{21} - σ_{33} (B-R)					

AG: /RED, σ_{21} / Listed: /51-X/	USELISTED (σ_{51} - σ_{21})	USELISTED (σ_{21})	USELISTED (σ_{33})	USELISTED (21)	USELISTED (33)
σ_{51} - σ_{21} (R-B)					
σ_{21} - σ_{21} (R-B)				*	
σ_{21} - σ_{21} (B-R)				*	
σ_{21} - σ_{33} (B-R)					

The USELISTED constraints used here are different from USELISTED in Zuraw (2000) in the following respects. First, Zuraw only uses USELISTED for morphologically complex forms, not for allomorphs. Second, Zuraw assumes that each candidate is an input-output pairing, and her USELISTED constraint is defined as “The input portion of a candidate must be a single lexical entry.” (p. 50) We have adopted a simpler assumption: the candidate that is identical in form to the listed reduplication is necessarily derived from the listed form. Third, Zuraw uses only one USELISTED constraint, and encodes the strength of a lexical entry by a listedness value of 0 to 1 that is determined by the entry’s lexical frequency. The listedness value reflects the availability of the lexical entry in the derivation of the output. We, on the other hand, have adopted a family

of USELISTED constraints, one for each sandhi pattern, and the strength of each USELISTED constraint, reflected in its weight, is also determined by the lexical frequency of the pattern.

As an anonymous reviewer pointed out, our three levels of USELISTED constraints have a fair amount of duplication. A lexicalized disyllable, whether reduplicated or not, will be subjected to three different USELISTED constraints, as shown in the first tableau in (20). But we hasten to point out that the duplication is motivated by data: there are potentially three different levels of productivity for AO, *AO, and AG from real Taiwanese data and our experimental result, and we have aimed to capture these different levels using a unified grammatical model. Without any of the three types of USELISTED constraints, we are bound to lose an empirical distinction that speakers make.

4.3.3. A phonotactic constraint

The grammar also needs a markedness constraint *24-NONFINAL as in (21). It expresses a true phonotactic generalization in Taiwanese.

(21) *24-NONFINAL: Tone 24 cannot occur on non-XP-final syllables.

4.4. Learning biases as σ^2 values

Wilson (2006) argued that learning biases can be encoded as different σ^2 values of the Gaussian prior in a Maxent model. We capitalize on Wilson's claim to capture the underlearning and overlearning aspects of our Taiwanese speakers' behaviors here.

We assume the default weight μ for all constraints to be zero. The σ^2 s for IDENT-IO(Tone), IDENT-BR(Tone), ALIGN-L, ALIGN-R, and *24-NONFINAL are set to 1; the σ^2 values for USELISTED constraints, however, are $B_{Listed} \times B_{Duration}$, where B_{Listed} and $B_{Duration}$ are two coefficients representing the biases based on listedness and phonetic duration, respectively.

For each type of USELISTED constraints (disyllabic words, monosyllabic allomorphs, tonal allomorphs), we posit B_{Listed} to be 10 to the negative power of a logistic function, in which x represents the number of morphemes that the type of USELISTED constraints covers on average, as in (22a). As estimated from Tsay and Myers’s corpus, the average number of monosyllabic homophones is 2.1, and the average number of morphemes that a lexical tone may denote is 585.4. These values represent the x values for the USELISTED constraints for monosyllabic allomorphs and tonal allomorphs, respectively. The x value for the USELISTED constraints for disyllabic words is naturally 1. We can then calculate the B_{Listed} values for these constraints accordingly, as in (22b). The intuition behind this bias coefficient is that if USELISTED is the learner’s strategy to cope with exceptional patterns that cannot be captured by regular means, such as the MARKEDNESS » FAITHFULNESS ranking, then the learner is first of all cautious about positing exceptions, expressed in the model by assigning USELISTED constraints greater penalties if they deviate from the default ranking of 0; secondly, the learner is unwilling to treat massive amounts of data as exceptions, expressed in the model as greater penalties for USELISTED constraints that make generalizations.

(22) B_{Listed}

a. $B_{Listed} = 10^{-\frac{1}{1+e^{2-2x}}}$, x = the number of morphemes that the type of USELISTED constraints covers on average

b.

	x	B_{Listed}
USELISTED($\sigma\sigma$)	1	0.3162
USELISTED(σ)	2.1	0.1258
USELISTED(T)	585.4	0.1

The $B_{Duration}$ coefficient expresses a substantive bias *à la* Wilson (2006). It is in the form of a logistic function as in (23a), in which x represents the percentage of duration increase from the base tone to the sandhi tone. We base our duration data on Lin’s (1988) study of Taiwanese

tones. The values for tone duration (measured in isolation), x , and $B_{Duration}$ are summarized in (23b).

(23) $B_{Duration}$

a. $B_{Duration} = \frac{11}{10 + e^{7x}}$, $x = \frac{d_{sandhi} - d_{base}}{d_{base}}$

b.

Sandhi	Base dur. (ms)	Sandhi dur. (ms)	x	$B_{Duration}$
21 → 51	82	91	0.1098	0.9049
51 → 55	91	150	0.6484	0.1062
55 → 33	150	151	0.0067	0.9957
33 → 21	151	82	-0.4570	1.0955
24 → 33	155	151	-0.0258	1.0153

The $B_{Duration}$ coefficient has the following properties: (a) When there is no duration change between the base and sandhi tones, it has a value of 1, indicating no learning bias. (b) When there a duration increase from the base tone to the sandhi tone, it has a value of less than 1, indicating a bias against the promotion of the corresponding USELISTED constraints to higher weights. (c) When there is a duration decrease from the base tone to the sandhi tone, it has a value of greater than 1, indicating encouragement of the promotion of the corresponding USEDLISTED constraints to higher weights. (d) The magnitude of the learning bias against duration increase is considerably greater than the magnitude of the bias in favor of duration decrease. For instance, a 65% duration increase (51 → 55) reduces $B_{Duration}$ by around 90%, but a 45% duration reduction only increases $B_{Duration}$ by around 10%. This echoes the Comparative Markedness solution to the “grandfather effect” by treating derived markedness as less costly than markedness in the input (McCarthy 2002).

With B_{Listed} and $B_{Duration}$ in place, the σ^2 values for all constraints are summarized in (24).

(24) σ^2 values for all constraints:

Constraint	σ^2	Constraint	σ^2
UsLSTD(σ_{51} - σ_{21})	0.2862	UsLSTD(21)	0.0905
UsLSTD(σ_{55} - σ_{51})	0.0336	UsLSTD(51)	0.0106
UsLSTD(σ_{33} - σ_{55})	0.3149	UsLSTD(55)	0.0996
UsLSTD(σ_{21} - σ_{33})	0.3464	UsLSTD(33)	0.1096
UsLSTD(σ_{33} - σ_{24})	0.3211	UsLSTD(24)	0.1015
UsLSTD(σ_{21})	0.1139	*24-NONFINAL	1
UsLSTD(σ_{51})	0.0134	IDENT-IO(Tone)	1
UsLSTD(σ_{55})	0.1253	IDENT-BR(Tone)	1
UsLSTD(σ_{33})	0.1378	ALIGN-L	1
UsLSTD(σ_{24})	0.1277	ALIGN-R	1

4.5. Modeling the speakers

We turn to the modeling of the speakers' learning of the tone sandhi pattern in Taiwanese in this section. The goal is to feed the learner a representative of the Taiwanese lexicon, and the learner will consequently output a grammar that closely correlates to the speakers' wug test behavior.

Our modeling was conducted with Maxent Grammar Tool — a Java program that simulates Maxent grammar learning by Hayes, Wilson, and George (2009). The data fed to the learner are real reduplications, real disyllables, and real monosyllables according the syllable type frequency data from Tsay and Myers's corpus. The dataset includes 163 types of reduplications, 1389 types of disyllables, and 1389 types of monosyllables distributed among different base tones as in (25). This is an approximation of a real Taiwanese speaker's learning data. The input to the reduplication of a syllable σ with a tone T is /RED, σ T/. All tonal combinations in both R-B and B-R were considered as candidates; the complete candidate set for /RED, σ_{21} /, for example, includes 50 items — 25 possible tonal combinations, each either R-B or B-R, and the winner is [σ_{51} - σ_{21} (R-B)]. The input to the disyllables puts the base tone in nonfinal position and another syllable following it, e.g., / σ_{21} - σ_x /. The candidates for each input include each possible tone in

the tonal inventory in nonfinal position, with the tone of the final position either identical to or different from that of the input; the entire candidate set for / σ_{21} - σ_x /, therefore, is [σ_{51} - σ_x], [σ_{21} - σ_x], [σ_{55} - σ_x], [σ_{24} - σ_x], [σ_{33} - σ_x], [σ_{51} - σ_y], [σ_{21} - σ_y], [σ_{55} - σ_y], [σ_{24} - σ_y], [σ_{33} - σ_y], and the winner is [σ_{51} - σ_x]. For monosyllables, an input is a syllable with a specified tone, and its candidates include the same syllable with all tones in the inventory. For example, for the input / σ_{21} /, its entire candidate set is [σ_{21}], [σ_{51}], [σ_{55}], [σ_{24}], [σ_{33}], and the winner is [σ_{21}].

(25) Learning data fed to Maxent Grammar Tool

	Input	Output	Frequency
Reduplication	/RED, σ_{21} /	[σ_{51} - σ_{21}]	29
	/RED, σ_{51} /	[σ_{55} - σ_{51}]	34
	/RED, σ_{55} /	[σ_{33} - σ_{55}]	41
	/RED, σ_{33} /	[σ_{21} - σ_{33}]	26
	/RED, σ_{24} /	[σ_{33} - σ_{24}]	33
Disyllable	/ σ_{21} - σ /	[σ_{51} - σ]	255
	/ σ_{51} - σ /	[σ_{55} - σ]	293
	/ σ_{55} - σ /	[σ_{33} - σ]	316
	/ σ_{33} - σ /	[σ_{21} - σ]	246
	/ σ_{24} - σ /	[σ_{33} - σ]	279
Monosyllable	/ σ_{21} /	[σ_{21}]	255
	/ σ_{51} /	[σ_{51}]	293
	/ σ_{55} /	[σ_{55}]	316
	/ σ_{33} /	[σ_{33}]	246
	/ σ_{24} /	[σ_{24}]	279

The constraint weights learned by the Maxent model are given in (26).

(26) Constraint weights learned by Maxent model:

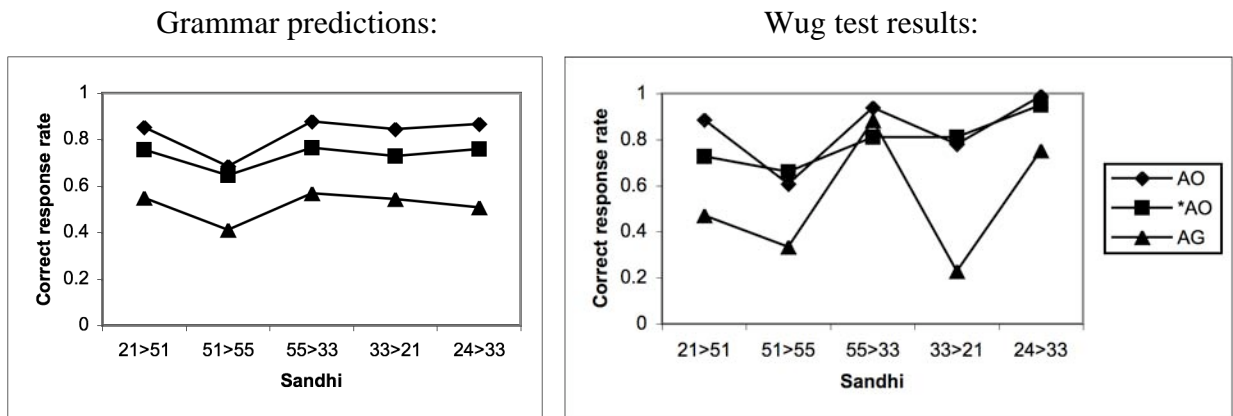
Constraint	Weight	Constraint	Weight
USELISTED(σ_{51} - σ_{21})	0.6167	USELISTED(21)	2.0633
USELISTED(σ_{55} - σ_{51})	0.1803	USELISTED(51)	1.0728
USELISTED(σ_{33} - σ_{55})	0.7960	USELISTED(55)	2.1935
USELISTED(σ_{21} - σ_{33})	0.7049	USELISTED(33)	2.1285
USELISTED(σ_{33} - σ_{24})	0.7152	USELISTED(24)	1.5459
USELISTED(σ_{21})	2.5962	*24-NONFINAL	3.2816
USELISTED(σ_{51})	1.3499	IDENT-IO(Tone)	3.0279
USELISTED(σ_{55})	2.7600	IDENT-BR(Tone)	0
USELISTED(σ_{33})	2.6783	ALIGN-L	2.8062
USELISTED(σ_{24})	1.9452	ALIGN-R	0

The constraint weights in (26) have the following six properties. First, the USELISTED constraints requiring real syllables with existing reduplications to use the listed reduplications have relatively low weights despite their higher σ^2 s. This results from the overall low type frequencies of reduplications as compared to non-reduplicative disyllables. This partially accounts for the fact that the tone sandhi is not 100% productive even in AO words in the experiment. Second, the USELISTED constraints that require an existing syllable to use its listed allomorph in nonfinal position are more highly weighted than the USELISTED constraints that require appropriate tonal allomorphs for existing tones regardless of the segmental content. In particular, for any tone T, USELISTED(σ T) has a higher weight than USELISTED(T). This is due to the lower σ^2 s imposed on the USELISTED(T) constraints. This property accounts for the higher sandhi productivity in *AO than AG words. Third, within each set of USELISTED constraints, the one that governs base 51 has by far the lowest weight as a result of the considerably lower σ^2 imposed on the constraint due to the phonetic disadvantage of 51 \rightarrow 55, the one that governs base 55 has the highest weight as a result of its high type frequency, and the one that governs base 33, despite its phonetic advantage, does not have the greatest weight, as the increase in σ^2 due to the phonetic advantage is minimal and does not compensate for its low type frequency. Fourth, *24-NONFINAL, as a true phonotactic generalization, has a relatively high weight. Fifth,

ALIGN(RED, L, WORD, L) has a considerably higher weight than ALIGN(RED, R, WORD, R), which has a weight of 0, indicating that there is a strong preference for the reduplicant to be a prefix. Given that all data fed to the learner have prefixing reduplication as the winning candidate, this is not surprising. Finally, IDENT-BR(Tone) has a weight of 0. This is because in the candidate sets that we considered, this constraint never “favored the winner,” as the winning candidates have either more or the same number of violations of this constraint than any given losing candidates.

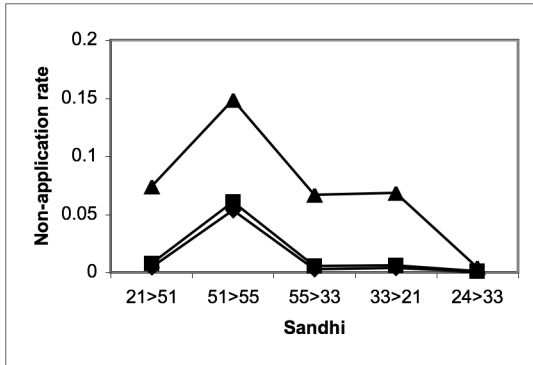
The rates of correct response, non-application, and template matching for the sandhi patterns predicted by the grammar in (26) are juxtaposed with the wug test results in (27), (28), and (29), respectively. The predicted rates for the three different types of sandhi application along with other changes are graphed together for AO, *AO, and AG in bar-graphs and juxtaposed with the wug test results in (30).

(27) Correct response rates for sandhi patterns — grammar prediction vs. wug test result:

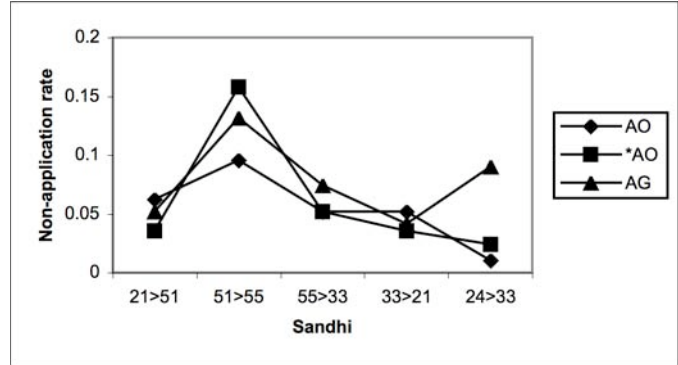


(28) Non-application rates of sandhi patterns — grammar prediction vs. wug test result:

Grammar predictions:

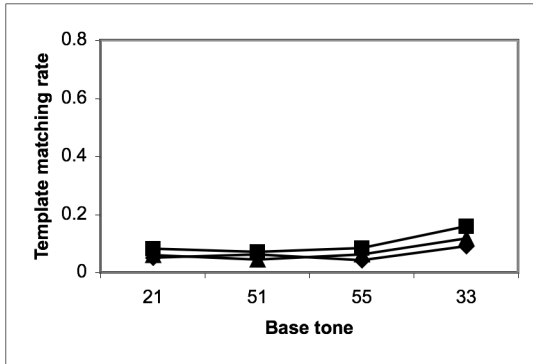


Wug test results:

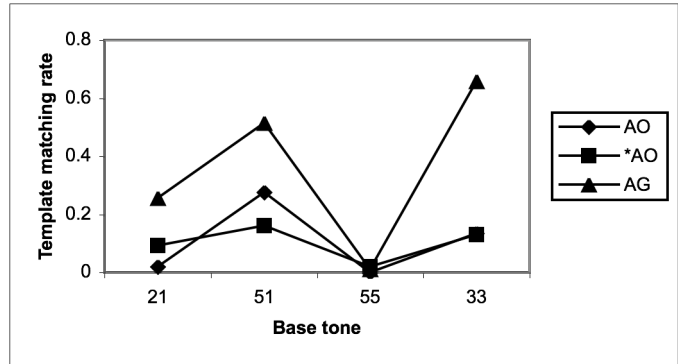


(29) Template matching rates for base tones — grammar prediction vs. wug test result:

Grammar predictions:



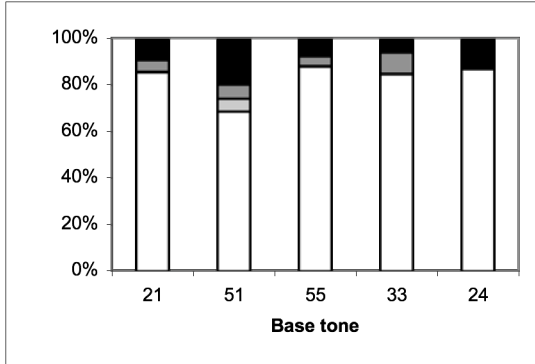
Wug test results:



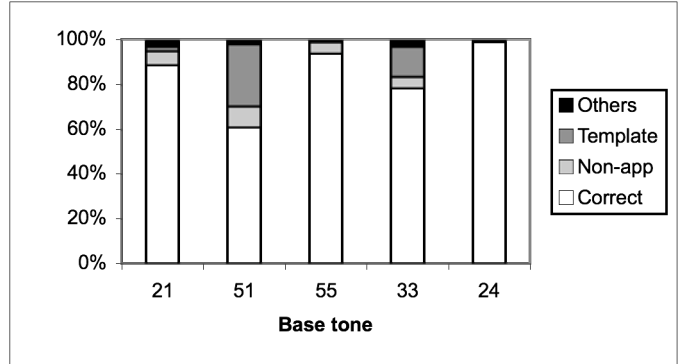
(30) Summaries for different categories of sandhi application — grammar predictions vs. wug test results:

a. AO:

Grammar predictions:

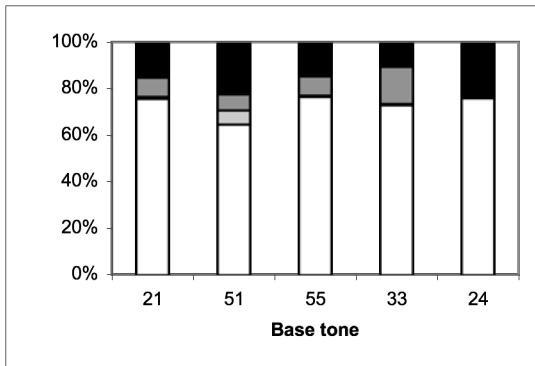


Wug test results:

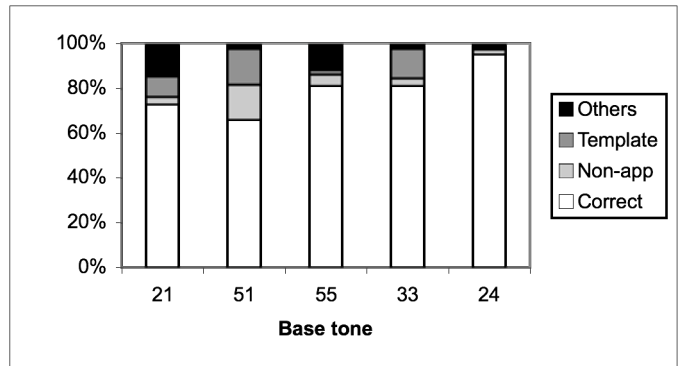


b. *AO:

Grammar predictions:

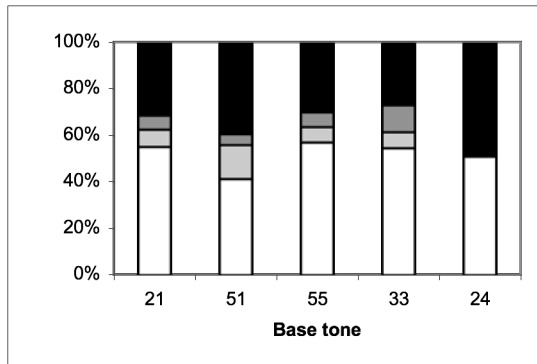


Wug test results:

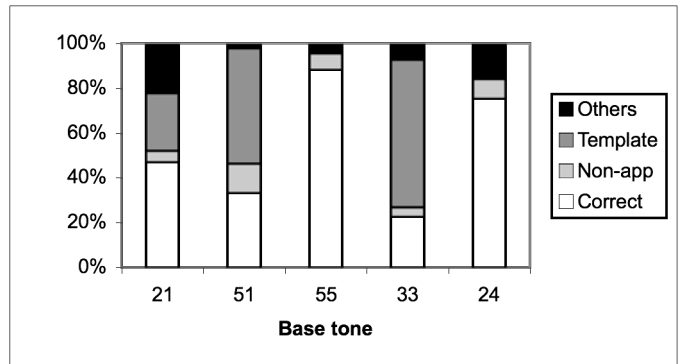


c. AG:

Grammar predictions:



Wug test results:



Let us first comment on the areas that the grammar did well. First of all, it captures the underlearning of the exceptionless sandhi patterns in the lexicon, as shown by the predicted high correct response rates for the sandhis in AO words, slightly lower rates in *AO words, and considerably lower rates in AG words (see (27)); the non-application rates are also predicted to be higher for AG than AO and *AO (see (28)).

Second, the overlearning of the phonetic effect beyond lexical statistics is also predicted by the grammar. In agreement with the wug test results, it predicts that the duration increasing sandhi 51 → 55 has the lowest correct response rate and highest non-application rate; moreover, it also predicts the gradient “grandfather effect” in that the duration reducing sandhi 33 → 21 does not show a clear productivity advantage, as the effect of duration reduction in the grammar is not strong enough to compensate for the negative effect of its low lexical frequency.

Third, the grammar predicts that the phonotactic generalization that no rising tone can appear on nonfinal syllables is highly productive. The rate by which the base tone 24 changes into another tone in nonfinal position is predicted to be 99.7%, as compared to those of other base tones, which range from 83.8% (base 33) to 90.1% (base 55).

Lastly, lexical frequency is also predicted to have an effect on productivity in that the most common sandhi pattern 55 → 33 is predicted to be the most productive despite its lack of

durational advantage, and that the low frequency of occurrence of 33 → 21 is predicted to have a strong enough negative effect to offset the effect of its durational advantage on its productivity.

In sum, our model is able to address all the major challenges that it sets out to model. Using the categories of “correct response,” “non-application,” “template matching,” and “others,” we conducted a regression analysis using the wug test results as the dependent variable and the grammars’ predictions as the independent variable. The analysis shows that the two variables are significantly correlated with each other: $R^2=.744$, adjusted $R^2=.739$, $\beta=0.862$, $p<0.001$.

Let us now move onto the areas where the grammar did not do so well. The most striking difference between the grammar’s predictions and the wug test results is that the rate of template matching responses are predicted to be low by the grammar. Consequently, the template matching rate differences among different tones in the experiment are generally not captured by the model. The only such difference that the model hints at is the greater rate of template matching rate for 33, due to the fact that there are two different templates (33-55 and 33-24) that fall under this category. The low template matching rates are by and large caused by the zero weight of ALIGN-R. Given that in the data fed to the learner, all reduplications are prefixing, the zero weight of ALIGN-R is unsurprising. The question, then, is why template matching has such a high rate of occurrence in the wug test result. It is possible that the subjects had a tendency to repeat the monosyllabic stimulus first, and then select a legal disyllabic tonal template for reduplication. The high rate of template matching responses, then, may not be due to the grammar having a hidden preference for suffixing reduplication, but an artifact of the experimental procedure. In our other wug-test studies of Taiwanese tone sandhi such as the formation of disyllabic words (Zhang et al. 2006, 2009a, b) and both single and double reduplications of monosyllabic adjectives (Zhang and Lai 2008), cases of template matching were very rare. It is thus likely that this pattern is a side effect of wug-testing this particular type of reduplication and should thus not be ascribed to any grammatical significance.

Another generalization from the wug test that the grammar fails to capture is the high correct response rate of 24 → 33. Although the grammar does predict that 24 cannot occur nonfinally, it

does not predict that 24 → 33 has the highest correct response rate as in the wug test. This is due to the fact that in the Maxent grammar, part of the work of the USELISTED constraints regarding base 24 is shared by *24-NONFINAL; this causes the weights of these USELISTED constraints to be suppressed during the learning process and consequently reduces the correct response rate of 24 → 33.

An additional problem that may have to do with the nature of the Maxent model is that the rates of the non-dominant variants are generally predicted to be lower than the observed rates. This is not only true in the template matching cases, but also in the non-application cases for AO and *AO. Unlike models of OT with constraint domination, the Maxent model does not in principle rule out candidates that are harmonically bound by other candidates. This means that when the candidate set for an input is large, as is the case for the reduplication forms here, the Maxent model usually predicts a small percentage for each candidate and consequently causes the real variants to have smaller percentages than desired.

We can also see that the predicted differences among different sandhis are generally smaller than the observed differences in the wug test. The small size of the frequency effect is due to the relatively small differences in syllable type frequencies among different tones, and the Maxent model is not able to amplify such effects. This is particularly obvious in 33 → 21, whose correct response rate in the wug test is considerably lower than all other sandhis except the durationally disadvantaged 51 → 55. But the predicted rate for 33 → 21, though in the right direction when compared to other sandhis, does not reach a low enough level (see (27)).

Finally, a word of caution is in order with regard to the sandhi responses for AO words. Unlike in the wug test, in real life the native speakers presumably use the sandhi patterns in (1) with close to 100% accuracy for the reduplications. This indicates that the wug results are likely confounded by the speakers' occasional inability to access the listed reduplication forms during the experiment. The experimental setting may have presented difficulties for *AO and AG words as well, but the wug test result for *AO and AG words should nonetheless reflect the general patterns of the speakers' knowledge as the experimental difficulty should apply across

the board to all tonal patterns in reduplication. Therefore, it is possible that the real grammar of the speakers will predict overall higher correct response rates for all sandhis, but the relative comparisons among different sandhis should be similar to the predictions of our grammar in (26).

These problematic areas aside, our learning model has successfully predicted a number of important generalizations that emerged from the wug test. The success of the model crucially relies on a number of its properties. The underlearning of exceptionless opaque sandhi patterns in the lexicon is achieved by deriving the opaque patterns via USELISTED constraints that directly state the input-output mappings of the sandhis and a learning bias against these constraints in the form of lower σ^2 s. Opaque patterns, however, may be partially productive due to the listing of the abstract input-output mappings, and the productivity of patterns is correlated with the listedness of such mappings reflected as weights of the USELISTED constraints. The overlearning of the phonetic knowledge despite the lack of such information from lexical statistics is modeled by a substantive learning bias against phonetically unmotivated patterns, again in the form of lower σ^2 s, *à la* Wilson (2006). Transparent and exceptionless generalizations from the lexicon are properly learned in the grammar as high-weighted markedness constraints, such as *24-NONFINAL. The proper learning of lexical statistics happens naturally for the learner, as the Maxent model inherently encodes the frequency effects by letting more frequent patterns to have greater pulls on constraint weights.

5. General discussion

We have so far proposed a learning model for Taiwanese speakers' phonological knowledge of tone sandhi in reduplication. The model is generally successful in capturing the emergence of patterns that are underlearned (opaque sandhis), overlearned (durational motivation of the sandhis), and properly learned (ban of 24 in nonfinal position; lexical frequency effect of the sandhis) from the lexicon. However, the model also leaves a number of open questions. We

make some exploratory remarks on these questions in this section, in the hopes that future research will provide fuller answers to them.

5.1. The listing of abstract input-output mappings and the nature of constraints in OT

In our learning model, the difference in productivity between real and wug words prompted our proposal that allomorphs must be listed with different levels of abstractness. Crucially, the abstract tonal allomorphs must be listed to account for the partial productivity of the sandhi patterns observed in AG words. The listing of abstract input-output mappings potentially poses a duplication problem for the theory of phonology: there are now two different mechanisms through which allomorphy can be derived — the listing of abstract input-output mappings and the MARKEDNESS » FAITHFULNESS ranking. For instance, Post-Nasal Voicing can now be derived by either a highly ranked USELISTED constraint that requires a voiceless consonant to use its voiced counterpart after a nasal or the *NÇ » IDENT(voice) ranking.

Our proposal, however, is based on empirical data: without the listing of abstract input-output tonal mappings as violable constraints, we cannot account for the partial productivity of the opaque tone circle observed in AG words. General mechanisms that derive opaque patterns within OT such as the Sympathy Theory (McCarthy 1999), OT with candidate chains (McCarthy 2007), and the encoding of contrast preservation permitted by opaque patterns (Lubowicz 2003) predict full productivity of these patterns in novel words, as no distinction is made between real and novel words with respect to candidate evaluations against the constraint system; but without either these mechanisms or abstract listing, opaque patterns are predicted to be categorically unproductive. Therefore, the duplication at least rests on the need for descriptive adequacy.

As we have mentioned, what we need is a strategy to keep the constraint weights on abstract allomorph listing low despite their exceptionlessness. This is achieved in the model by assigning greater penalties for these constraints in the form of smaller σ^2 values if they deviate from the default ranking of 0. Without this bias, the learner would readily promote the weights of these

constraints, and we would incorrectly predict full productivity of the opaque sandhi patterns in AG words. Transparent allomorphy, then, still derives from the MARKEDNESS » FAITHFULNESS ranking, and the role of the USELISTED constraints is invisible. But opaque allomorphy cannot be derived from any MARKEDNESS » FAITHFULNESS rankings, and the ranking (or weighting) of the abstract USELISTED constraints determines the productivity of the opaque pattern in novel words. It is likely that the productivity of opaque patterns varies crosslinguistically, which would mean that the ranking of these USELISTED constraints cannot be universally suppressed so low as to prevent partial or even complete productivity of the opaque patterns, but cannot be unrestrained as to predict complete productivity whenever the pattern is exceptionless in the lexicon. The intuition, however, is Optimality-Theoretic: crosslinguistic variation is predicted by differences in constraint ranking.

The coexistence of traditional markedness and faithfulness constraints with *ad hoc* USELISTED constraints that require the surface forms to use listed allomorphs coincides with Moreton's (2004) argument that the grammar is composed of an innate and "conservative" component of markedness and faithfulness constraints and a language-specific component with constraints that require particular lexical items to have particular surface representations in particular environments. These language-specific constraints are posited when the learner detects that the phonological pattern cannot be computed by the interaction of markedness and faithfulness. We share with Moreton (2004) the intuition that such constraints are necessary in the grammar in any case to deal with processes that target specific lexical items and morphological categories, that are suppletive, and that have lexical exceptions. Moreover, we may consider USELISTED as a universal template into which learners plug the specifics of their language, as suggested by one anonymous review. This further restricts the language specificity aspect of the grammatical model.

5.2. How phonology interacts with morphology and syntax

We must acknowledge that our analysis of the productivity patterns in Taiwanese reduplication rests on a number of assumptions on how phonology interacts with morphology and syntax. Although the reduplication dealt with here is clearly a morphological process, the general characterization of the sandhi positions in Taiwanese as “non-XP-final” is syntactically based. If we consider our treatment of the allomorph listing in reduplication to be extendable to other configurations where the same sandhis apply, then our approach is consistent with the “precompiled phrasal phonology” of Hayes (1990), which considers syntactically-determined allomorphs as deriving from lexical rules that refer to subcategorization frames with syntactic information within the lexicon (see also Tsay and Myers 1996). A preliminary analysis of general disyllabic tone sandhi in Taiwanese is presented in Zhang et al. (2009b). Our approach is also partly consistent with Paster (2006) and Bye (2007) in considering phonologically-conditioned suppletive allomorphy as deriving from the selection of allomorphs according to subcategorization frames, but deviates from them in not treating allomorph listing as a separate component from phonology (Paster 2006) or as a declarative component within phonology with inviolable constraints that takes place after H-Eval (Bye 2007). Rather, we consider allomorph listing as constraints that fully interact with the rest of phonology à la Burzio (2002) and MacBride (2004), as the listedness of the allomorphs is influenced by phonetics and lexical frequency, both of which must have a direct effect on phonology proper.

One issue that is worth noting is that in the model of tone sandhi in Taiwanese in general, given that the sandhi domain is syntactically defined, the only relevant USELISTED constraints will often be those that govern nonfinal allomorphs for existing syllables and nonfinal tonal allomorphs. In other words, many forms will be exactly like *AO in our experiment. Our model, then, faces the same problem of predicting too low a sandhi application rate as for the AO words in our experiment. Our response is similar to the one offered for AO words at the end of §4.5: the overall lower sandhi productivity for *AO words observed in the experiment may have

been partly caused by the nature of the experiment; this time, the complicating factor is that many of *AO syllables used in the experiment (see Appendix) are associated with a prominent nominal or verbal morpheme, and given that the subjects were instructed to form *adjectival* reduplications in the experiment, they might have been occasionally confused and simply repeated the syllable twice. Since our model aimed to match the experimental result, it consequently predicted lower sandhi application rates in *AO than the real grammar. We are grateful to an anonymous reviewer that prompted us to clarify this issue.

5.3. The source of variation and noise

A larger issue at play here is that we have chosen the grammar as the underlying cause for the variation and noise observed in the experimental results. This is an unconventional position, especially for generative phonologists that adhere to the competence/performance dualism. However, a large body of works in laboratory phonology have shown that factors traditionally considered to be in the realm of performance, such as word frequency and phonetic effects due to human physiology and perception, interact with grammatical factors in phonological systems and play important roles in phonological acquisition (see Pierrehumbert et al. 2000, for example). This casts doubt on a sharp divide between competence and performance. In line with the position of these researchers, we consider the variation patterns that we observed among different sandhi processes and the different levels of sandhi productivity among different word types to be part of the speakers' linguistic knowledge, not the result of performance factors. The sandhi effect, as we have argued, is due to the combined influence of lexical frequency and phonetic duration, both of which have been shown by earlier works to be relevant to phonological patterning (Bybee 2001, Pierrehumbert 2006, Zuraw 2000; Zhang 2002, Zhang and Lai 2010, etc.). The word type effect is treated by different levels of USELISTED constraints inspired by Zuraw (2000). One may wish to attribute the low sandhi productivity in wug words to performance factors due to the nature of the experiment, such as the foreignness of these

words, which consequently caused the speakers to put them in a different lexical stratum. But this position prevents two important questions to be answered. First, it provides no explanation for why certain sandhi processes, such as the phonotactically transparent 24 → 33, are considerably more productive than other sandhi processes, even in wug words. If we look beyond the sandhi patterns in question, the argument is even clearer. Zhang and Lai (2008) showed that Taiwanese speakers can compute a transparent floating High docking pattern in double reduplication completely productively in wug words despite their low rates of correct tone sandhi application. In the classic Berko (1958) study, she showed that the regular plural allomorphy is entirely productive when adult English are wug-tested. To attribute the low productivity observed here to performance misses the correlation between transparency and productivity — a true phonological generalization — entirely. Second, it provides no explicit mechanism to allow a new word to be gradually adapted into the Taiwanese lexicon and eventually undergo sandhi. Wang (1993), for example, documented loanwords from Mandarin Chinese and English that undergo tone sandhi according to the tone circle. A grammatical model based on the interaction between the speaker and the hearer and the hearer's Bayesian reasoning of the listedness of the sandhied form, however, is available from Zuraw (2000, 2003).

We do not deny that there is indeed performance noise in our experimental results. We have discussed in §5.1 and §5.2 that the correct sandhi rates for both AO and *AO words are underrepresented in our results, and we have argued in §4.5 that the high rates of template matching in the experiments may have been caused by an artifact of the experimental procedure. The real grammar of Taiwanese speakers needs to correct these issues. We have chosen to include these performance noises in our model simply because we do not know how to quantitatively decouple them from the true grammatical variations in a reliable and unbiased fashion. This shortcoming of the model is duly noted.

6. Alternative analyses

We review previous Optimality-Theoretic analyses of Taiwanese tone sandhi in this section. But before doing so, let us further clarify the opaque nature of the Taiwanese sandhi pattern.

As discussed in footnote 2, we use “opacity” here in the Kiparskian sense (Kiparsky 1973). A phonological rule P of the form $A \rightarrow B/C_D$ is opaque if there are surface structures with any of the following characteristics: (a) instances of A in the environment C_D , or (b) instances of B derived by P that occur in environments other than C_D . A sandhi process in the Taiwanese tone circle, e.g., $21 \rightarrow 51 / _ X]_{XP}$, satisfies condition (a) (instances of 21 appear in $_ X]_{XP}$ as the sandhi tone of 33 in this environment) and is therefore opaque.

Some patterns of opacity are problematic for traditional OT (Goldsmith 1993, Hermans and Oostendorp 1999, McCarthy 1999, 2007, McCarthy and Prince 1993, Prince and Smolensky 1993/2004, Roca 1997, Moreton 2004, among many others). In particular, Moreton (2004) provided a proof that circular chain shift cannot be captured by a “conservative” OT grammar that uses only IO-faithfulness and markedness constraints. This makes the understanding of Taiwanese tone sandhi particularly important for theories of phonology.

There are three general approaches under which Taiwanese tone sandhi can be analyzed within the OT framework. The first approach is to do away the opacity of the pattern either representationally or through acoustic evidence, and then analyze the pattern as a transparent pattern. The second approach is to relax the “conservativity” property of OT and allow constraints other than IO-faithfulness and markedness to enter into the system. The third approach is to treat the sandhi patterns as listed allomorphic relations, like what we have done here, but to couch the analysis in stochastic OT (Boersma and Hayes 2001) rather than Maximum Entropy. We discuss all three approaches in turn below.

6.1. Taiwanese tone sandhi as a transparent pattern

The most straightforward way to eliminate opacity in Taiwanese tone sandhi is to show that the derived sandhi tones are in fact acoustically not identical to their supposedly homophonous tones in XP-final position. For instance, the tone 51 as the result of the 21 → 51 sandhi in nonfinal position may not be identical to the base tone 51 in XP-final position. This hypothesis, however, has not been borne out in a series of phonetic studies (Chang 1988, Lin 1988, Du 1988, Tsay and Myers 2001, Tsay et al. 1999, Myers and Tsay 2008). In the most recent of these studies, Myers and Tsay (2008) carefully controlled for the segmental environments and the number of syllables in the stimulus sentences, used a large number of older speakers, who tended to be more fluent in Taiwanese than younger speakers, and subjected their acoustic results to stringent statistical testing. They found a durational difference between base tones that appear in XP-final position and sandhi tones that appear in nonfinal positions as expected, but no difference in either f₀ height or shape between the base and sandhi tones that are identical in transcription. Moreover, the speakers were not able to adjust the distinctiveness of the two tonal categories under comparison according to pragmatic conditions (absence vs. presence of a listener), further indicating complete neutralization (cf. Charles-Luce 1997).

Another option is to treat the homophonous base and sandhi tones as simply having distinct phonological representations. In Yip's (1980) rule-based analysis, for example, the base tone 33 is represented as in (31a), while the sandhi tone 33 as derived from 55 and 24 is represented as (31b) and (31c), respectively. This opens the possibility that the sandhi processes in fact interact transparently and can therefore be captured in traditional OT.

- (31) a. Base 33 b. Sandhi 33 derived from 55 c. Sandhi 33 derived from 24
- [+upper] [-upper] [-upper]
- \wedge \wedge |
- L L H H H

Moreton (2004) argues against this approach on two grounds. First, it is difficult to motivate the different phonological representations without phonetic evidence. Second, even if the phonological representations *can* be motivated, it is still implausible that we can come up reasonable markedness constraints to derive the sandhis. For instance, we would have to argue that the representation for the base 33 is considerably more marked than that for the sandhi 33 derived from 55 due to $33_{\text{base}} \rightarrow 21 \rightarrow 51 \rightarrow 55 \rightarrow 33_{\text{sandhi}}$ — a position difficult to defend.

In short, both empirical and theoretical attempts to eliminate opacity from Taiwanese tone sandhi have failed to provide convincing evidence for the transparency of the pattern. We must therefore look within OT to search for workable alternatives.

6.2. Adopting additional mechanisms in OT

Assuming that Taiwanese tone sandhi is truly opaque, we must adopt additional mechanisms beyond traditional IO-faithfulness and markedness to derive the sandhi patterns within OT. Two prominent additions to OT have been considered in the literature: anti-faithfulness (Alderete 2001) and contrast preservations (Lubowicz 2003). Horwood (2000) used the former, Barrie (2006) the latter, and Hsieh (2005) and Thomas (2008) a combination of the two.

The anti-faithfulness approach utilizes a highly ranked novel constraint that requires the sandhi tone to be different from the citation tone. Together with traditional IO-faithfulness constraints, this approach predicts that the tone sandhis will change the base tones minimally to satisfy the highly ranked anti-faithfulness constraint, and in the meantime incur minimum violations of the IO-faithfulness constraints. This prediction finds support if the tones 33, 21, 51, and 55 are represented as /L, h/, /L, hl/, /H, hl/, and /H, h/, respectively— each tone sandhi in the

circle 33 → 21 → 51 → 55 → 33 can then be interpreted as a change in the register or the contour, but not both (Horwood 2000).¹⁰

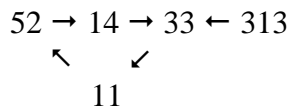
The contrast preservation approach utilizes constraints that ban the merge of two base tones (Hsieh 2005, Thomas 2008) or the neutralization of certain tone features such as register and contour (Barrie 2006). The chain shift is triggered by a highly ranked markedness constraint *RISE, which bans 24 in nonfinal position; the rest of the tonal inventory then shuffles to avoid further neutralization. The exact way in which the chain is formed is determined by lower ranked markedness constraints such as *CONTOUR and *HIGH (Barrie 2006) or faithfulness constraints such as IDENT(Pitch) and IDENT(Shape) (Thomas 2008).

There are two crucial problems to these generic approaches to the sandhi pattern. First, it provides no explanation for the low productivity of the sandhi processes in the wug test. Hsieh (2005) and Barrie (2006) presented two alternative interpretations for the low sandhi application rates in Hsieh's (1970, 1975, 1876) and Wang's (1993) wug tests: (a) the subjects were unwilling to apply the sandhi to a wug word if it results in a real word, and (b) the subjects might have trouble pronouncing phonotactic gaps in the Taiwanese syllabary. But in the current study, none of the reduplicated words in either *AO or AG is a real word, and the AG words are accidental, not principled gaps in the Taiwanese syllabary by having both a legal segmental composition and a legal tone. Accidental gaps are often used in nonce-probe tests to show the productivity of alternation patterns with great success; e.g., *wug*, *wug[z]*. These interpretations for the low productivity results, therefore, do not hold water. Moreover, these interpretations provide no explanation for the *differences* in productivity among different sandhis in wug words — a point already made in §5.3.

¹⁰ We thank an anonymous reviewer for bringing this work to our attention.

Second, as Chen (2000) pointed out, many Southern Min dialects of Chinese have essentially the same tone circle pattern as Taiwanese, but different tonal values at each point in the chain shift; e.g., Longxi, as in (32). The pattern similarity among these dialects originated from the proto-language that had the same tone circle pattern in the historical tonal categories. Given that the analyses based on anti-faithfulness and contrast preservation for Taiwanese crucially depend on the tonal values of Taiwanese, they are not easily generalizable to these other dialects, which means that the tonal grammar of these other dialects will be drastically different from Taiwanese despite the similarity of their overall patterns and their close genetic affinity. This seems counterintuitive.

(32) Lonxi tone sandhi (Chen 2000):



Therefore, as we have commented in §5.1, the best solution for the opaque Taiwanese tone circle, and possibly phonological opacity in general, is likely not a generic solution that requires additional machineries to be introduced into OT, such as the Sympathy Theory (McCarthy 1999), OT with candidate chains (McCarthy 2007), and the encoding of contrast preservation (Lubowicz 2003), which predict full productivity of these patterns. A fruitful first step is to show that these patterns are indeed psychologically real — a point made in Hsieh (1970, 1975, 1976) for rule-based phonology that resonates just as soundly today.

6.3. USELISTED in stochastic OT

In a similar experiment on the productivity of tonal patterns in Taiwanese double reduplication, Zhang and Lai (2008) showed similar results on the semi-productivity of tone sandhi, which occurs on the second syllable of the trisyllabic sequence; moreover, they also showed that a transparent pattern of floating High docking onto the first syllable is entirely productive in novel words. These results further support our treatment of opaque sandhi patterns as listed allomorphic mappings. Zhang and Lai (2008) used similar USELISTED constraints in their analysis, but couched the analysis in stochastic OT (Boersma and Hayes 2001) rather than Maxent. The crucial improvement of our analysis over Zhang and Lai (2008) is that we provided a model of learning that learns the grammar based on simulated inputs, while Zhang and Lai (2008) used the observed outputs in the experiment as the input to the Gradual Learning Algorithm and only used the algorithm as a constraint ranking tool to find a grammar that approximates those observed outputs. In fact, desired learning results are unobtainable in the stochastic OT model of Zhang and Lai (2008), which has no way of suppressing the promotion of all USELISTED constraints, including those for the syllable and tonal allomorphs; consequently, if the inputs to the algorithm only consist of forms that satisfy all USELISTED constraints, as real world inputs are, then all USELISTED constraints will be promoted to the top of the constraint hierarchy, predicting full productivity of the sandhi pattern. This problem is acknowledged in Zhang and Lai (2008) (p.213-214). Our solution here is to bias against the increase of weights of the more abstract USELISTED constraints by small σ^2 values in the Maxent model. It is conceivable that this can be done in stochastic OT as well, but no formal implementation has been proposed so far.

7. Conclusion

We have argued in this paper that Taiwanese speakers' phonological knowledge on the tone sandhi pattern in reduplication is the combined result of lexical statistics and a priori knowledge that they bring to the task of learning, which causes the speakers to know both more and less than the lexical patterns. The argument is based on the different levels of productivity observed for different word types and different tones in a wug test. The opaque tone circle is only variably productive in novel words, indicating that the speakers have underlearned this exceptionless pattern in the lexicon. The productivity of the opaque sandhis, however, also shows signs of both overlearning and proper learning from the lexicon: speakers prefer shorter tones as sandhi tones on nonfinal syllables, which could not have been deduced from the lexicon, and the lexical frequencies of tones and tonal melodies also have an effect on productivity. Finally, the transparent phonotactic generalization *24-NONFINAL is productively observed, even in novel words.

We modeled the Taiwanese speakers' learning of the sandhi patterns in reduplication using a Maximum Entropy grammar. We argued that the model needs to encode two crucial properties in order to achieve a grammar that closely predicts the speakers' wug test behavior. First, the model must include lexical listing constraints that list the input-output mappings of tone sandhis. This echoes the dual listing/generation theory of Zuraw (2000). The opaque mappings are essentially treated as patterned exceptions and listed in the lexicon with various degrees of abstractness. The regular lexical behavior is due to the combined effect of USELISTED constraints that govern existing words as well as more abstract USELISTED constraints, while the more variable behavior in wug words is due to the lack of effect of USELISTED constraints for existing words. This property of the model addresses the underlearning problem and helps the model predict partial productivity of the sandhi patterns despite the fact that the patterns are fully

productive in the learner's input. Second, the model needs to have a substantive learning bias *à la* Wilson (2006). The learning bias mitigates against a USELISTED constraint by suppressing its weight if it has either one of the following two properties: (a) it covers an input-output mapping beyond a lexical item, and (b) it governs a durationally disadvantaged sandhi. The former also contributes crucially to the modeling of the underlearning behavior, while the latter captures the overlearning of the phonetic knowledge despite the lack of such information in lexical statistics. The proper learning of transparent and exceptionless generalizations happens naturally in the model as there is no bias against the weight promotion of such constraints. The proper learning of lexical statistics also happens naturally for the learner, as the Maxent model inherently encodes the frequency effects by letting more frequent patterns to have greater pulls on constraint weights.

In sum, we hope that we have demonstrated through this study that (a) we can arrive at a more precise estimate of the speakers' phonological knowledge via experimental means than simply studying the patterns in the language itself, and (b) a theory that incorporates the gradience of phonetics and lexical statistics is a step in the right direction in modeling speakers' phonological knowledge.

Acknowledgments

This research is inspired by Hsieh (1970, 1975, 1976) and Wang (1993), and we owe a special thank to them. We are indebted to Jane Tsay and James Myers for their help throughout this project, particularly in providing us with a Taiwanese frequency corpus and hosting us at National Chung Cheng University in Chiayi, Taiwan for part of our data collection. We also thank Paul Boersma and Mietta Lennes for their assistance with Praat scripts and Bruce Hayes for his help with the Maxent Grammar Tool software. For helpful discussions and written comments on various aspects of the manuscript, we thank Andries Coetzee, Bruce Hayes, René Kager, Charles Chien-Jer Lin, James Myers, and Joe Pater. Earlier versions of this work were

presented at the 16th annual conference of the International Association of Chinese Linguistics at Beijing University, and we are grateful to the audience at the conference for their feedback. All remaining errors are strictly our own. This research is partly supported by research grants from the National Science Foundation (0750773), the Chiang Ching-Kuo Foundation for International Scholarly Exchange, and the University of Kansas General Research Funds (2301760).

References

- Albright, Adam (2002). Islands of reliability for regular morphology: evidence from Italian. *Language* 78: 684-709.
- Albright, Adam, Argelia Andrade, and Bruce Hayes (2001). Segmental environments of Spanish diphthongization. In Adam Albright and Taehong Cho (eds.), *UCLA Working Papers in Linguistics* 7 (*Papers in Phonology* 5): 117-151.
- Albright, Adam and Bruce Hayes (2003). Rules vs. analogy in English past tense: a computational/experimental study. *Cognition* 90: 119-161.
- Albright, Adam and Bruce Hayes (2006). Modeling productivity with the Gradual Learning Algorithm: the problem of accidentally exceptionless generalizations. In Gisbert Fanselow, Caroline Féry, Ralf Vogel, and Matthias Schlesewsky (eds.), *Gradience in grammar: generative perspectives*, 185-204. Oxford University Press, Oxford, UK.
- Alderete, John (2001). Dominance effects as trans-derivational anti-faithfulness. *Phonology* 18: 201-253.
- Baayen, R. Harald (1992). Quantitative aspects of morphological productivity. In Geert E. Booij and Jaap van Marle (eds.), *Yearbook of Morphology* 1991. Dordrecht: Kluwer Academic Publishers. 109-149.
- Baayen, R. Harald (1993). On frequency, transparency, and productivity. In Geert E. Booij and Jaap van Marle (eds.), *Yearbook of Morphology* 1992. Dordrecht: Kluwer Academic Publishers. 181-208.

- Barrie, Michael (2006). Tone circles and contrast preservation. *Linguistic Inquiry* 37.1: 131-139.
- Berko, Jean (1958). The child's learning of English morphology. *Word* 14: 150-177.
- Blevins, Juliette (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge University Press, Cambridge, UK.
- Boersma, Paul (1998). *Functional phonology: Formalizing the interaction between articulatory and perceptual drives*. Holland Academic Graphics, The Hague.
- Boersma, Paul and Bruce Hayes (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32.1: 45-86.
- Boersma, Paul and David Weenink (2005). *Praat: doing phonetics by computer* (Computer program). Retrieved November 2005, from <http://www.praat.org/>.
- Bolinger, Dwight Le Merton. (1961). *Generality, gradience and the all-or-none*. Mouton, The Hague.
- Burzio, Luigi (2002). Surface-to-surface morphology: when your representations turn into constraints. In Paul Boucher (ed.), *Many Morphologies*, 142-177. Cascadilla Press, Somerville, MA.
- Bybee, Joan L. (1985). *Morphology: A study of the relation between meaning and form*. Benjamins, Philadelphia.
- Bybee, Joan L. (2001). *Phonology and language use*. Cambridge University Press, Cambridge, UK.
- Bybee, Joan and Elly Pardo (1981). On lexical and morphological conditioning of alternations: A nonce-probe experiment with Spanish verbs. *Linguistics* 19: 937-968.
- Bye, Patrick (2007). Allomorphy — selection, not optimization. In Sylvia Blaho, Patrik Bye, and Martin Krämer (eds.), *Freedom of Analysis?* Mouton de Gruyter, Berlin. 63-91.
- Chang, Yueh-Chin (1988). Sandhi tonal des syntagmes dissyllabiques du Min-nan parlé a Taiwan (Tone sandhi in disyllabic compounds in Minnan spoken in Taiwan). *Cahiers de Linguistique Asie Orientale* 17: 193-234.

- Charles-Luce, Jan (1997). Cognitive factors involved in preserving a phonemic contrast. *Language and Speech* 40.3: 229-248.
- Chen, Matthew Y. (1987). The syntax of Xiamen tone sandhi. *Phonology Yearbook* 4: 109-150.
- Chen, Matthew Y. (2000). *Tone sandhi: patterns across Chinese dialects*. Cambridge University Press, Cambridge, UK.
- Chen, Stanley F. and Ronald Rosenfeld (2000). A survey of smoothing techniques for maximum entropy models. *IEEE Transactions on Speech and Audio Processing* 8.1: 37-50.
- Cheng, Robert L. (1997). *Taiwanese and Mandarin structures and their developmental trends in Taiwan: Taiwanese phonology and morphology*. Yuan-Liou Publishing Co., Taipei, Taiwan.
- Chung, Raung-fu (1996). *The segmental phonology of Southern Min in Taiwan*. The Crane Publishing Co., Taipei, Taiwan.
- Davidson, Lisa (2005). Phonology, phonetics, or frequency: influences on the production of non-native sequences. *Journal of Phonetics* 34: 104-137.
- Della Pietra, Stephen, Vincent J. Della Pietra, and John D. Lafferty (1997). Inducing features of random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19: 180-191.
- Du, Tsai-Chwun (1988). *Tone and stress in Taiwanese*. Ph.D. dissertation, University of Illinois, Urbana Champaign.
- Ernestus, Mirjam and R. Harald Baayen (2003). Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79: 5-38.
- Flemming, Edward (2001). Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18.1: 7-44.
- Frisch, Stefan A., Janet B. Pierrehumbert, and Michael B. Broe (2004). Similarity avoidance and the OCP. To appear in *Natural Language and Linguistic Theory* 22: 179-228.

- Frisch, Stefan A. and Bushra Zawaydeh (2001). The psychological reality of OCP-place in Arabic. *Language* 77.1: 91-106.
- Goldsmith, John (ed.) (1993). *The last phonological rule: reflections on constraints and derivations*. University of Chicago Press, Chicago.
- Goldwater, Sharon and Mark Johnson (2003). Learning OT constraint ranking using a maximum entropy model. In Jennifer Spenader, Anders Eriksson, and Osten Dahl (eds.), *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*. 111-120.
- Hansson, Gunnar Ólafur (2008). Diachronic explanations of sound patterns. *Language and Linguistics Compass* 2.5: 859-893.
- Hayes, Bruce (1990). Precompiled phrasal phonology. In Sharon Inkelas and Draga Zec (eds.), *The syntax-phonology connection*, 85-108. Center for the Study of Language and Information, Stanford, CA, and University of Chicago Press, Chicago.
- Hayes, Bruce and Zsuzsa Cziráky Londe (2006). Stochastic phonological knowledge: the case of Hungarian vowel harmony. *Phonology* 23: 59-104.
- Hayes, Bruce and Donca Steriade (2004). Introduction: The phonetic bases of phonological markedness. In Hayes, Bruce, Donca Steriade, and Robert Kirchner (eds.), *Phonetically based phonology*. Cambridge University Press, Cambridge, UK. 1-33.
- Hayes, Bruce and Colin Wilson (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39.3: 379-440.
- Hayes, Bruce, Colin Wilson, and Benjamin George (2009). *Maxent Grammar Tool*. Java program downloaded from <http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool/>, May 22, 2009.
- Hermans, Ben and Marc van Oostendorp (eds.) (1999). *The derivational residue in phonology*. John Benjamins, Amsterdam and Philadelphia.
- Horwood, Graham (2000). *Anti-identity and Taiwanese tone sandhi*. Ms., Rutgers University.

- Hsiao, Yuchau E. (1991). *Syntax, rhythm and tone: a triangular relationship*. The Crane Publishing Co., Taipei, Taiwan.
- Hsieh, Feng-fan (2005). Tonal chain-shifts as anti-neutralization-induced tone sandhi. In Sudha Arunachalam, Tatjana Scheffler, Sandhya Sundaresan, and Joshua Tauberer (eds.), *Proceedings of the 28th Annual Penn Linguistics Colloquium*. Penn Linguistics Circle, University of Pennsylvania, Philadelphia. 99-112.
- Hsieh, Hsin-I (1970). The psychological reality of tone sandhi rules in Taiwanese. In *Papers from the 6th Meeting of the Chicago Linguistic Society*, 489-503. Chicago, IL.
- Hsieh, Hsin-I (1975). How generative is phonology. In E. F. Koerner (ed.), *The transformational-generative paradigm and modern linguistic theory*, 109-144. John Benjamins, Amsterdam, The Netherlands.
- Hsieh, Hsin-I (1976). On the unreality of some phonological rules. *Lingua* 38: 1-19.
- Hyman, Larry (2001). On the limits of phonetic determinism in phonology: *NC revisited. In Elizabeth Hume and Keith Johnson (eds.), *The role of speech perception in phonology*. Academic Press, New York. 141-186.
- Hyman, Larry (2007). Universals of tone rules: 30 years later. In Tomas Riad and Carlos Gussenhoven (eds.), *Tones and tunes, vol. 1: Typological studies in word and sentence prosody*. Mouton de Gruyter: Berlin and New York. 1-34.
- Hyman, Larry and Russell Schuh (1974). Universals of tone rules: Evidence from West Africa. *Linguistic Inquiry* 5: 81-115.
- Inkelas, Sharon and Cheryl Zoll (2005). *Reduplication: doubling in morphology*. Cambridge University Press, Cambridge, UK.
- Jäger, Gerhard (2007). Maximum Entropy Models and Stochastic Optimality Theory. In Annie Zaenen, Jane Simpson, Tracy Holloway King, Jane Grimshaw, Joan Maling, and Chris Manning (eds.), *Architectures, rules, and preferences — variations on themes by Joan W. Bresnan*. CSLI Publications, Stanford, CA. 467-479.

- Johnson, Mark, Stuart Geman, Stephen Canon, Zhiyi Chi, and Stefan Riezler (1999). Estimators for stochastic “unification-based” grammars. *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*: 535-541.
- Kenstowicz, Michael and Charles Kisseberth (1979). *Generative Phonology*. Academic Press, San Diego.
- Kiparsky, Paul (1973). Abstractness, Opacity, and Global Rules. In Osamu Fujimura (ed.), *Three Dimensions of Linguistic Theory*. TEC Company, Ltd., Tokyo, Japan. 57-86.
- Klatt, Dennis H. (1975). Vowel lengthening is syntactically determined in connected discourse. *Journal of Phonetics* 3: 129-140.
- Labov, William (1972). *Sociolinguistic patterns*. University of Pennsylvania Press, Philadelphia.
- Labov, William (1994). *Principles of linguistic change: Internal factors*. Blackwell, Oxford, UK and Cambridge, USA.
- Lafferty, John, Andrew McCallum, and Fernando Pereira (2001). Conditional random fields: probabilistic models for segmenting and labeling sequence data. *International Conference on Machine Learning. ICML'01*. Morgan Kaufmann, San Francisco, CA. 282-289.
- Lin, Hwei-Bing (1988). *Contextual stability of Taiwanese tones*. Ph.D. dissertation, University of Connecticut.
- Lin, Jo-wan (1994). Lexical government and tone group formation in Xiamen Chinese. *Phonology* 11: 237-275.
- Lubowicz, Ania (2003). *Contrast preservation in phonological mappings*. Ph.D. dissertation, University of Massachusetts, Amherst.
- MacBride, Alexander Ian (2004). *A constraint-based approach to morphology*. Ph.D. dissertation, UCLA.
- Martin, S. C., H. Ney, and J. Zaplo (1999). Smoothing methods in maximum entropy language modeling. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 1: 545-548.

- McCarthy, John (1999). Sympathy and phonological opacity. *Phonology* 16: 331-399.
- McCarthy, John (2002). Comparative markedness. In Angela C. Carpenter, Andries W. Coetzee, and Paul de Lacy (eds.), *University of Massachusetts Occasional Papers in Linguistics 26: Papers in Optimality Theory II*. GLSA, Amherst, MA. 171–246.
- McCarthy, John (2007). *Hidden generalizations: phonological opacity in Optimality Theory*. Equinox, London.
- McCarthy, John and Alan Prince (1993). *Prosodic Morphology I: constraint interaction and satisfaction*. Unpublished ms., University of Massachusetts, Amherst and Rutgers University.
- McCarthy, John and Alan Prince (1995). Faithfulness and reduplicative identity. In Beckman, Jill, Laura Walsh Dickey, and Suzanne Urbanczyk (eds.), *University of Massachusetts Occasional Papers in Linguistics 18: Papers in Optimality Theory*. GLSA, Amherst, MA. 249-384.
- McCarthy, John and Alan Prince (1999). Faithfulness and identity in Prosodic Morphology. In Rene Kager, Harry van der Hulst, and Wim Zonneveld (eds.), *The prosody-morphology interface*. Cambridge University Press, Cambridge, UK. 218-309.
- Moder, Carol Lynn (1992). *Productivity and categorization in morphological classes*. Ph.D. dissertation, SUNY Buffalo.
- Moreton, Elliot (2004). Non-computable functions in Optimality Theory. In John J. McCarthy (ed.), *Optimality Theory in phonology*. Blackwell Publishing, Malden, MA. 141-164.
- Moreton, Elliott (2008). Analytical bias and phonological typology. *Phonology* 25: 83-127.
- Myers, James and Jane Tsay (2008). Neutralization in Taiwan Southern Min tone sandhi. In Yuchau E. Hsiao, Hui-Chuan Hsu, Lian-Hee Wee, and Dah-An Ho (eds.), *Interfaces in Chinese Phonology: Festschrift in Honor of Matthew Y. Chen on His 70th Birthday*. Academia Sinica, Taiwan. 47-78.
- Ohala, John J. (1997). The relation between phonetics and phonology. In William Hardcastle and John Laver (eds.), *The handbook of phonetic sciences*. Blackwell, Oxford. 674-694.

- Oller, D. Kimbrough (1973). The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America* 54: 1235-1247.
- Paster, Mary Elizabeth (2006). Phonological conditions on affixation. Ph.D. dissertation, UC Berkeley.
- Peng, Shui-Hui (1997). Production and perception of Taiwanese tones in different tonal and prosodic contexts. *Journal of Phonetics* 25: 371-400.
- Pierrehumbert, Janet B. (2003). Probabilistic phonology: discrimination and robustness. In Rens Bod, Jennifer Hay, and Stefanie Jannedy (eds.), *Probabilistic linguistics*. MIT Press, Cambridge, MA. 177-228.
- Pierrehumbert, Janet B. (2006). The statistical basis of an unnatural alternation. In Louis Goldstein, Douglas H. Whalen, and Catherine Best (eds.), *Laboratory Phonology VIII, Varieties of Phonological Competence*, 81-107. Mouton de Gruyter, Berlin.
- Pierrehumbert, Janet B., Mary E. Beckman, and D. R. Ladd (2000). Conceptual foundations of phonology as a laboratory science. In Noel Burton-Roberts, Philip Carr, and Gerard Docherty (eds.), *Phonological knowledge: conceptual and empirical issues*. Oxford University Press, Oxford, UK. 273-304.
- Prince, Alan and Paul Smolensky (1993/2004). *Optimality Theory: constraint interactions in generative grammar*. Ms., Rutgers University and University of Colorado, Boulder. Published in 2004 by MIT Press, Cambridge, MA.
- Roca, Iggy (ed.) (1997). *Derivations and constraints in phonology*. Clarendon Press: Oxford.
- Sanders, Nathan (2001). Preserving synchronic parallelism: diachrony and opacity in Polish. *Papers from the 37th Meeting of the Chicago Linguistic Society*. 501-515.
- Smolensky, Paul (1986). Information processing in dynamical systems: foundations of Harmony Theory. In David E. Rumelhart and James L. McClelland (eds.), *Parallel Distributed Processing: explorations in the microstructure of cognition*. MIT Press/Bradford Books, Cambridge, MA. 194-281.

- Smolensky, Paul, and Géraldine Legendre (2006). *The harmonic mind: from neural computation to Optimality-Theoretic grammar*. MIT Press, Cambridge, MA.
- Steriade, Donca (2008). The phonology of perceptibility effects: the P-map and its consequences for constraint organization. In Kristin Hanson and Sharon Inkelas (eds.), *The nature of the word: Essays in honor of Paul Kiparsky*. MIT Press, Cambridge, MA. 151-180.
- Sumner, Megan (2003). *Testing the abstractness of phonological representations in Modern Hebrew weak verbs*. Ph.D. dissertation, State University of New York at Stony Brook.
- Thomas, Guillaume (2008). An analysis of Xiamen tone circle. In Natasha Abner and Jason Bishop (eds.), *Proceedings of the 27th West Coast Conference on Formal Linguistics*. Cascadilla Proceedings Project, Somerville, MA. 422-430.
- Tsay, Jane, Jan Charles-Luce, and Yung-Song Guo (1999). The syntax-phonology interface in Taiwanese: acoustic evidence. *Proceedings of the 14th International Congress of Phonetic Sciences, vol. 3*. 2407-2410.
- Tsay, Jane and James Myers (1996). Taiwanese tone sandhi as allomorph selection. *Proceedings of 22nd Meeting of the Berkeley Linguistics Society*, 394-405.
- Tsay, Jane, and James Myers (2001). Processes in the production of Taiwanese tone sandhi: an acoustic phonetic study. In Lianhong Cai, Tongchun Zhou, and Jianhua Tao (eds.), *Proceeding of the 5th National Conference on Modern Phonetics*. Tsinghua University, Beijing. 233-237.
- Tsay, Jane and James Myers (2005). *Taiwanese spoken corpus*. National Chung Cheng University, Taiwan.
- Wang, H. Samuel (1993). Taiyu biandiao de xinli texing (On the psychological status of Taiwanese tone sandhi). *Tsinghua Xuebao (Tsinghua Journal of Chinese Studies)* 23: 175-192.
- Wightman, Collin W., Stefanie Shattuck-Hufnagel, Mari Ostendorf, and Patti J. Price (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 91: 1707-1717.

- Wilson, Colin (2003). Experimental investigation of phonological naturalness. In Gina Garding and Mimura Tsujimura (eds.), *Proceedings of the 22nd West Coast Conference on Formal Linguistics*, 533-546. Cascadilla Press, Somerville, MA.
- Wilson, Colin (2006). Learning phonology with substantive bias: an experimental and computational study of velar palatalization. *Cognitive Science* 30: 945-982.
- Wu, Tzuling (1996). *A morphological analysis of reduplication in Taiwan Southern Min*. MA thesis, National Tsinghua University, Taiwan.
- Yip, Moira (1980). *The tonal phonology of Chinese*. Ph.D. dissertation, MIT.
- Yip, Moira (1998). Identity avoidance in phonology and morphology. In Steven LaPointe, Diane Brentari, and Patrick Farrell (eds.), *Morphology and its relations to phonology and syntax*. CSLI Publications, Stanford, CA. 216-246.
- Zhang, Jie (2002). *The effects of duration and sonority on contour tone distribution — a typological survey and formal analysis*. Routledge, New York.
- Zhang, Jie (2003/2009). Contour tone distribution is not an artifact of tonal melody mapping. *Studies in the Linguistic Sciences* 33.1/2: 73-132.
- Zhang, Jie (2004a). The role of contrast-specific and language-specific phonetics in contour tone distribution. In Hayes, Kirchner, and Steriade (eds.), *Phonetically based phonology*, 157-190. Cambridge University Press, Cambridge, UK.
- Zhang, Jie (2004b). Contour tone licensing and contour tone representation. *Language and Linguistics* 5: 925-68.
- Zhang, Jie and Yuwen Lai (2008). Phonological knowledge beyond the lexicon in Taiwanese double reduplication. In Yuchau E. Hsiao, Hui-Chuan Hsu, Lian-Hee Wee, and Dah-An Ho (eds.), *Interfaces in Chinese Phonology: Festschrift in Honor of Matthew Y. Chen on His 70th Birthday*. Academia Sinica, Taiwan. 183-222.
- Zhang, Jie and Yuwen Lai (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology* 27.1.

- Zhang, Jie, Yuwen Lai, and Craig Turnbull-Sailor (2006). Wug-testing the “tone circle” in Taiwanese. In Donald Baumer, David Montero, and Michael Scanlon (eds.), *Proceedings of the 25th West Coast Conference on Formal Linguistics*, 453-461. Cascadilla Proceedings Project, Somerville, MA.
- Zhang, Jie, Yuwen Lai, and Craig Sailor (2009a). Effects of phonetics and frequency on the productivity of Taiwanese tone sandhi. *Papers from the 43rd Meeting of the Chicago Linguistic Society*. Chicago, IL.
- Zhang, Jie, Yuwen Lai, and Craig Sailor (2009b). Opacity, phonetics, and frequency in Taiwanese tone sandhi. In *Current issues in unity and diversity of languages: Collection of papers selected from the 18th International Congress of Linguists*. Linguistic Society of Korea. 3019-3038.
- Zimmer, Karl E. (1969). Psychological correlates of some Turkish morpheme structure conditions. *Language* 45.2: 309-321.
- Zuraw, Kie (2000). *Patterned exceptions in phonology*. Ph.D. dissertation, UCLA.
- Zuraw, Kie (2003). Probability in language change. In Rens Bod, Jennifer Hay, and Stefanie Jannedy (eds.), *Probabilistic linguistics*. MIT Press, Cambridge, MA.
- Zuraw, Kie (2007). The role of phonetic knowledge in phonological patterning: corpus and survey evidence from Tagalog infixation. *Language* 83: 277-316.

Appendix: Stimuli for the Wug-Test

Tone	AO			*AO			AG
21	膨	p ^h oŋ	‘inflated’	進	tsin	‘to enter’	tsoo
	臭	ts ^h au	‘smelly’	過	kue	‘to cross’	ts ^h ue
	碎	ts ^h ui	‘shattered’	帝	te	‘a chunk’	tsuan
	破	p ^h ua	‘broken’	吐	t ^h oo	‘to vomit’	ke
	拜	pai	‘to pray to’	氣	k ^h ui	‘air’	pan
	怪	kuai	‘strange’	退	t ^h e	‘to back up’	p ^h ai
	笑	ts ^h io	‘to laugh’	教	ka	‘to teach’	t ^h ua
	夠	gau	‘enough’	透	t ^h au	‘to mix’	tsua
51	飽	pa	‘full’	炒	ts ^h a	‘to stir fry’	tsaŋ
	短	te	‘short’	轉	tŋ	‘to turn’	ts ^h ui
	快	kin	‘fast’	體	t ^h e	‘body’	tsue
	脯	poo	‘dried’	表	piau	‘cousin’	k ^h ŋ
	少	tsio	‘few’	跑	tsau	‘to run’	pai
	普	p ^h o	‘ordinary’	手	ts ^h iu	‘hand’	p ^h au
	鬼	kui	‘impish’	請	ts ^h ã	‘to treat’	t ^h ŋ
	歹	p ^h ai	‘bad’	統	t ^h oŋ	‘to unite’	tsam
55	翹	k ^h iao	‘perked up’	尊	tsun	‘to respect’	tsan
	光	kŋ	‘bright’	區	k ^h u	‘area’	ts ^h ui
	深	ts ^h im	‘deep’	郊	kau	‘suburb’	tsue
	輕	k ^h in	‘light’	公	kaŋ	‘public’	k ^h ŋ
	膏	ko	‘sticky’	珍	tin	‘precious’	pě
	甜	fi	‘sweet’	添	t ^h iam	‘to add’	p ^h oo
	芳	p ^h aŋ	‘fragrant’	車	ts ^h ia	‘car’	t ^h oo
	清	ts ^h iŋ	‘pure, clear’	偷	t ^h au	‘to steal’	tsam
33	鈍	tun	‘blunt’	電	tian	‘electricity’	tsau
	近	kin	‘near’	病	pě	‘sickness’	ts ^h ui
	重	taŋ	‘heavy’	樹	ts ^h iu	‘tree’	tsui
	大	tua	‘big’	市	ts ^h i	‘city’	k ^h ŋ
	靜	ts ^h ŋ	‘quiet’	豆	tau	‘bean’	pa
	癢	tsiũ	‘itchy’	娶	ts ^h ua	‘to marry’	p ^h ai
	厚	kau	‘thick’	渡	too	‘to ferry’	t ^h oo
	硬	tiŋ	‘hard’	陣	tin	‘battle array’	tsaŋ
24	濕	tam	‘wet’	材	tsai	‘material’	tsa
	長	tŋ	‘long’	賠	pue	‘to compensate’	ts ^h ue
	鹹	kiam	‘salty’	頭	t ^h au	‘head’	tsue
	齊	ts̄e	‘orderly’	球	kiu	‘ball’	k ^h ŋ
	柴	ts ^h a	‘slow witted’	環	k ^h uan	‘bracelet’	pŋ
	肥	pui	‘fat’	陳	tan	‘old’	p ^h ai
	槌	t ^h ui	‘foolish’	奔	ts̄oŋ	‘to run quickly’	t ^h ua
	牢	tiau	‘sturdy’	茶	te	‘tea’	tsam