

Phonological Knowledge beyond the Lexicon in Taiwanese Double Reduplication*

Jie Zhang and Yuwen Lai
The University of Kansas

Recent studies on productivity have shown that native speakers' phonological knowledge not only includes statistical patterning of irregularities reflected 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) and perceptually motivated phonological scales (Zuraw 2007), as well as their difficulties with exceptionless opaque patterns in wug tests (Zhang et al., to appear). Based on the results of a wug test, we show in this paper that the tone pattern in Taiwanese double 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. We also provide a stochastic OT grammar based on the dual listing/generation model of Zuraw (2000) to account for our wug test results.

Key words: Taiwanese tone sandhi, opacity, double reduplication, wug test, productivity, stochastic OT, lexical listing

1. Introduction

The productivity of a linguistic pattern refers to its applicability to new items (Bybee 2001). Recent research on phonological productivity has shown that speakers often have detailed knowledge that reflects the statistical patterning of irregularity to lexical phonological generalizations (Zuraw 2000, Frisch & Zawaydeh 2001, Albright

* 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 our data collection. We also thank Paul Boersma and Mietta Lennes for helping us with Praat scripts, Craig Sailor for processing some of the data, and the editors of this volume and an anonymous reviewer for their helpful comments. Earlier versions of this work were presented at the 15th annual conference of the International Association of Chinese Linguistics in conjunction with the 19th North American Conference on Chinese Linguistics at Columbia University, the 12th Mid-Continental Workshop on Phonology at the University of Iowa, and the 27th West Coast Conference on Formal Linguistics, and we are grateful to the audiences at these conferences for their feedback. All remaining errors are strictly our own. This

2002, Albright & Hayes 2003, Pierrehumbert 2006, among others). For example, Zuraw (2000) shows that although Nasal Substitution is riddled with exceptions in the Tagalog lexicon, Tagalog speakers are nonetheless able to detect patterns within the exceptions, such as the greater tendency for a front stem-initial consonant (*p* or *b*) to undergo Nasal Substitution than a back stem-initial consonant (*k* or *g*), and apply the patterns to novel stems in a wug-test (Berko 1958). Hayes & Londe (2006) demonstrate that Hungarian speakers can also apply the statistical knowledge of vowel harmony to novel forms. A notable theory that accounts for such behavior of the speakers is Zuraw's (2000) dual listing/generation model.¹ The theory combines lexical listing, which accounts for the irregular behavior of existing words, with statistical generalizations of the irregularity patterns expressed in stochastic Optimality-Theoretic terms, which account for the frequency matching between the patterns in the lexicon and the speakers' varied outputs in the wug test.

However, speakers' phonological behaviors in wug-tests do not always match the lexical statistics of input-output mappings. There are areas in which speakers seem to know more than the lexical statistics. For instance, Davidson (2005) shows that when English speakers were asked to produce non-native word-initial clusters, they were not equally accurate in the production of different clusters, and the accuracy does not correlate with the lexical frequency of these clusters in other positions, but can instead 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, demonstrates 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 phonetic knowledge, and they can apply the knowledge to infixation in stems with novel initial clusters. These cases present a "poverty of the stimulus" argument for the grammatical treatment of abstract phonological and phonetic knowledge of the speakers. Zuraw's (2000) theory can be easily adapted to account for these cases provided that the stochastic OT constraints can encode phonological representations on abstract levels and biased phonetic knowledge.

There are also instances where speakers seem to know less than the lexical statistics. For example, Albright & Hayes (2006) argue that accidentally true generalizations in

research is supported by an award from the University of Kansas General Research Funds in 2005-2006 (#2301760) and a research grant from the Chiang Ching-Kuo Foundation for International Scholarly Exchange in 2006-2007. An earlier and shorter version of the paper appears in the *Proceedings of the 27th West Coast Conference on Formal Linguistics*, Cascadilla Proceedings Project.

¹ The name "dual listing/generation model" is in fact not given by Zuraw herself, but by Hayes & Londe (2006).

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, however, are found in languages with opaque phonological patterns. Sanders (2001) demonstrates 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, to appear).

We argue in this paper that the tone pattern in Taiwanese double reduplication is a case in which the speakers’ knowledge is a combination of more than, less than, and exactly as what their lexicon informs them, and we base our argument on a wug test in which native Taiwanese speakers were asked to produce the reduplicated forms of monosyllabic words. Regarding the tone circle, the speakers know less than the lexicon, and the exceptionless pattern in the lexicon is largely unproductive in novel words. The speakers also know more than the lexicon, as there are indications from their wug-test behavior that they prefer shorter tones on non-final syllables, which they could not have deduced from the lexicon, but could have deduced from phonetics, as non-final syllables are phonetically shorter than final syllables due to lack of final lengthening. Finally, transparent phonotactic generalizations on tones and the transparent behavior of floating tone docking are learned precisely from the lexicon and apply productively to wug words, and certain lexical frequency effects on tones also affect the speakers’ wug-test behavior.

We further argue that Zuraw’s theory can also be applied to cases in which the speakers underlearn from the lexicon due to opacity. Essentially, the opaque mappings are treated as exceptions and listed in the lexicon, which accounts for the lexical behavior. But lower ranked stochastic constraints that encode the phonetic and frequency biases will determine the behavior of novel words. The transparent patterns, however, can be derived productively for both lexical and novel items.

2. Tonal patterns in Taiwanese double 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 sandhi patterns are characterized by circular opacity:

assume that in both types of reduplication, the final syllable is the base and the non-final syllables are the reduplicants.

(2) Taiwanese monosyllabic adjective reduplications:

Monosyllabic adjective	Single reduplication 'somewhat adj.'	Double reduplication 'very adj.'	Gloss
p ^h oŋ21	p ^h oŋ51-p ^h oŋ21	p ^h oŋ51-p ^h oŋ51-p ^h oŋ21	'blown-up'
ŋŋ51	ŋŋ55-ŋŋ51	ŋŋ55-ŋŋ55-ŋŋ51	'soft'
sin55	sin33-sin55	sin35-sin33-sin55	'new'
kau33	kau21-kau33	kau35-kau21-kau33	'thick'
tam24	tam33-tam24	tam35-tam33-tam24	'wet'

The tonal pattern of single reduplication follows the general tone sandhi pattern—the initial reduplicated syllable undergoes tone sandhi according to (1). Double reduplication, however, has two different patterns: when the base tone is 21 or 51, the two reduplicants both have the sandhi tone from the pattern in (1); but when the base tone is 55, 33, or 24, the first syllable has a rising tone 35, while the second syllable observes the pattern in (1). These reduplicative tonal patterns are summarized in (3).

(3) Tonal patterns in Taiwanese reduplication:

Monosyllabic adjective	Single reduplication	Double reduplication
21	51 -21	51-51 -21
51	55 -51	55-55 -51
55	33 -55	35-33 -55
33	21 -33	35-21 -33
24	33 -24	35-33 -24

2.3 Generalizations about tonal patterns in double reduplication

Assuming that the final syllable represents the base of reduplicated forms, the tonal patterns of Taiwanese double reduplication can be summarized as follows: the tone on the second syllable in a doubly reduplicated form is derivable from the base tone according to the sandhi pattern in (1); the tone on the first syllable can be derived from the same sandhi tone via the insertion of a floating High tone—the High tone is inserted (vacuously) to the left edge of 51 (base 21) and 55 (base 51), but to the right edge of 33 (base 55 and 24) and 21 (base 33). The edge of insertion of the floating High can be determined by the correspondence between the two reduplicated syllables: the High is inserted to the left edge if it does not disrupt the identity between the two reduplicated syllables at the left edge; otherwise the High is inserted on the right. These observations

have been made in various forms in Cheng (1973), Yip (1990), Wu (1996), Myers & Tsay (2001), and in particular, Lin (2004). Notice that we have treated the “2” in “21” and the “3” in “33” as identical “Mid” tones here, as acoustic studies have shown that the mid level and low falling tones in Taiwanese start from the same pitch (Lin 1988, Peng 1997). Hence, the insertion of a High tone at the right edge of these tones renders identical 35s.

Among these generalizations, the insertion of the floating High is entirely transparent, provided that the sandhi tone is known. The sandhi tones themselves, however, may exhibit both underlearning and overlearning from the lexicon: previous studies have indicated that speakers may have underlearned the pattern due to their opaque nature, but the different phonetic bases among the different sandhi processes may also provide the speakers with knowledge that they cannot glean from the lexicon. The phonetic bases relate to the duration of the tones. Studies by Lin (1988) and Peng (1997) show that the two falling tones 51 and 21 have considerably shorter intrinsic durations than 55, 33, and 24 in Taiwanese. Given that the sandhi occurs on non-final 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 an anti-duration change, as it is duration-increasing; the other two sandhis are durationally neutral. That the duration properties of the sandhis have an effect on their productivity in Taiwanese have been shown in Zhang et al. (2006, to appear).

Therefore, the tonal patterns of Taiwanese double 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 internal evidence gleaned from the patterns in the language itself. Provided that such interaction can be demonstrated, it will provide us with a novel case—one in which underlearning from the lexicon due to opacity coexists with overlearning and proper learning—to test Zuraw’s (2000) theory of patterned exceptionality.

The next section lays out the details of a wug-test study on the tonal patterns in Taiwanese double reduplication, which aims to investigate the various factors on the productivity of these patterns experimentally. Section 4, then, proceeds to model the productivity using a grammatical model based on Zuraw (2000) and argues that, like patterned exceptions, underlearned opacity can be derived via lexical listing, and overlearned phonetic effects and properly learned transparent patterns and lexical statistics can be derived via other stochastically ranked constraints.

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 & Pardo 1981, Albright 2002, Albright & Hayes 2003, Pierrehumbert 2006) and morphophonological alternations (e.g., Hsieh 1970, 1975, 1976, Wang 1993, Zuraw 2000, 2007, Albright et al. 2001, Hayes & Londe 2006). The basic method was that we asked native Taiwanese speakers to first listen to monosyllabic words through a headphone, and then produce the monosyllable, the single reduplication, and the double reduplication in turn.

Three types of monosyllabic words were used in the experiment: real words with actual occurring single and double reduplications (AO), real words with no actual occurring reduplications, either single or double (*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.⁴ For each type of word, all five tones that appear on open and sonorant-closed syllables (21, 51, 55, 33, 24) were used, and eight different words were used for each word type × tone combination. This made a total of 120 test words. A sample of the word list is shown in (4). 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.

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

Tone	AO			*AO			AG
21	怪	kuai	'strange'	進	tsin	'to enter'	tsɔ
51	飽	pa	'full'	炒	ts ^h a	'to stir fry'	p ^h au
55	芳	p ^h aj	'fragrant'	區	k ^h u	'area'	tsan
33	鈍	tun	'blunt'	電	tian	'electricity'	k ^h ŋ
24	濕	tam	'wet'	頭	t ^h au	'head'	p ^h aĩ

3.1.2 Experimental set-up and data analyses

The experiment was conducted with SuperLab (Cedrus). Each subject first heard an instruction in Taiwanese through a headphone, which explained in both lay person's prose and examples that Taiwanese can mark the lesser degree of an adjective by single

⁴ The naming of the different groups follows Hsieh (1970, 1975, 1976).

reduplication and the intensification of an adjective by double reduplication, and their task was to form the diminutive and intensive forms of the words they would hear from the headphone. They were also instructed that some of the words are not real words in Taiwanese, but they should form the diminutive and intensive 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 produced the monosyllable, the single reduplication, and the double reduplication in turn. 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 called Paihe near Tainan, where the second author also recruited ten speakers and conducted the experiment in a quiet room. Among the twenty speakers, six did not produce usable data. Four of them did not have one or more of the tonal patterns in double reduplication laid out in (3), as indicated by incorrect production of the tonal patterns in all AO words. This could be due to the fact that the subjects spoke a different dialect of Taiwanese that had a different tonal pattern, or that they did not fully understand the instructions. One speaker did not provide single reduplication responses for any of the test stimuli. One speaker produced the contracted forms of double reduplication as her single reduplication responses.⁵ These six speakers' data were not used in subsequent data analyses. Among the fourteen speakers whose data we analyzed, four were male, ten were female, and they had an average age of 50.2 at the time of the experiment.

3.1.4 Data analyses

Due to the structure-preserving nature of Taiwanese tone sandhi (Tsay et al. 1999, Myers & Tsay 2001), the tonal responses to all the experimental stimuli were transcribed in Chao tone letters by the two authors of the paper—one native Taiwanese speaker (Lai) and one native Mandarin speaker (Zhang). For the first syllable, visual inspection of the pitch tracks in Praat (Boersma & Weenink 2005) indicated that correct responses to base tones 55, 33, and 24 all had a complex contour tone with an approximate value

⁵ For discussion on the tonal patterns on the contracted forms of double reduplication, see Chung (1996).

of 353. We simplified our transcription of the complex contour to 35, with the understanding that a rising tone in first syllable position necessarily entails a convex tone. Our agreement based on first-time auditory impression was about 95%. With help from pitch tracks in Praat, we agreed on virtually all tokens. For the handful of tokens on which we did not reach an agreement, we took the native Taiwanese speaker Lai's judgment.

In subsequent statistical analyses, we excluded a token if it fell under one of the following categories: (a) the subject mispronounced the tone of the monosyllable; (b) the subject did not provide a complete set of monosyllable/single reduplication/ double reduplication. Also, 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. The subjects' responses to these stimuli were subsequently discarded.

3.1.5 Hypotheses

The hypotheses for the experiment are as follows.

First, the subjects' correct response rates⁶ for sandhi tones, which occur initially in single reduplication and medially in double reduplication, should be low for novel reduplications (*AO and 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-reducing sandhi is more productive than a duration-increasing sandhi, indicating overlearning from the lexicon.

Third, the transparent floating High docking, derivable from the high ranking of the correspondence between the left edge of the two reduplicated tones, is productive: if we assume that the second syllable in the subjects' trisyllabic responses had what they considered to be the correct sandhi tone, their first syllable response *based on this second syllable* should be generally correct, even in wug words. For example, for a base tone 21, if the subjects produced the correct sandhi tone 51 in the second syllable, then we would expect them to produce also 51 in the first syllable, which is the correct tone; but if they produced the wrong sandhi tone 21 in the second syllable, we would still expect them to be able to calculate the first tone, only that it would be based on the wrong sandhi tone 21, so the first tone would be 35, as the floating High would now dock to the right to preserve the initial Mid. We would not expect the subjects to provide the

⁶ "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, of course, 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.

correct tone 51 for the first syllable if they had produced the wrong sandhi tone 21 in the second syllable.

Fourth, the transparent phonotactic generalization that no rising tone can appear on non-XP-final syllables (except on the first syllable of double reduplication) should be productive, and the speakers should avoid surface rising tones on the initial syllable of single reduplication and the medial syllable of double reduplication.

Fifth, 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 (Tsay & Myers 2005) is summarized in (5). In (5), 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 occurrence of the tone in the entire corpus. Frequency information on the five tonal melodies attested in single reduplication based on 163 reduplicated forms in the corpus is given in (6). It is crucial to note that none of these frequency scales correlates with the duration nature of the sandhi in that a more frequent tone undergoes a durationally more advantageous sandhi.

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

(5) 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

(6) Frequencies of tonal melodies in single reduplication:

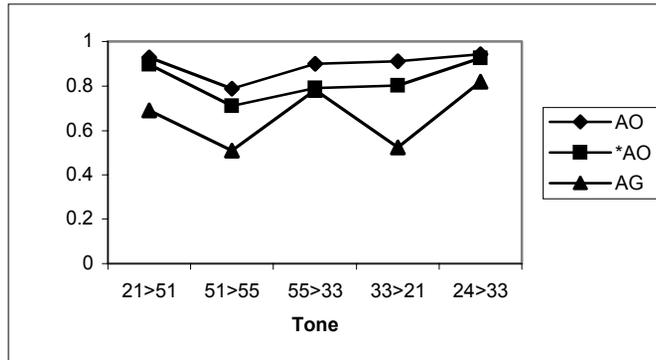
- 33-55 > 55-51 > 33-24 > 51-21 > 21-33

3.2 Results

The correct response rates for the sandhi tones in single reduplication, which are on the first syllable, are given in (7). A Two-Way Repeated-Measures ANOVA showed that the effect of word type is significant ($F(1.877, 24.407) = 37.599, p < 0.001$); so is the effect of tone ($F(3.759, 48.862) = 3.599, p < 0.05$). The interaction between word type and tone is not significant ($F(4.836, 62.868) = 2.146, p > 0.05$). Bonferroni posthoc tests indicated that the subjects performed the sandhis relatively accurately (i.e., according to the pattern in (1)) for real words (AO), less accurately for real syllables without existing reduplication (*AO), and the least accurately for novel words (AG); the difference between AO and *AO is significant at the $p < 0.05$ level, while the other two pairwise

comparisons are significant at the $p < 0.001$ level. Posthoc tests also showed that there is a significant difference between the $51 \rightarrow 55$ and the $24 \rightarrow 33$ sandhis ($p < 0.05$), but not other pairs. The $55 \rightarrow 33$ sandhi has a higher correct response rate than $51 \rightarrow 55$ in numerical values, but the difference does not reach statistical significance. The higher productivity for $24 \rightarrow 33$ indicates that it indeed benefits from its transparent phonotactic generalization.

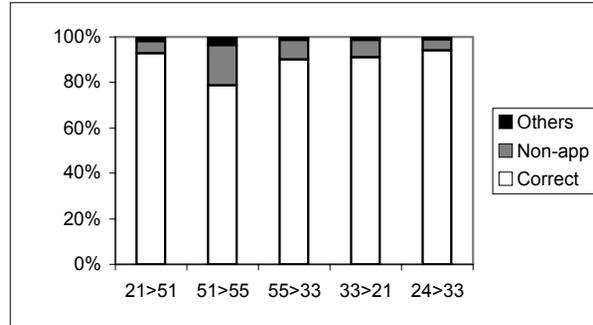
(7) Correct response rates for sandhi tones in single reduplication (σ_1):



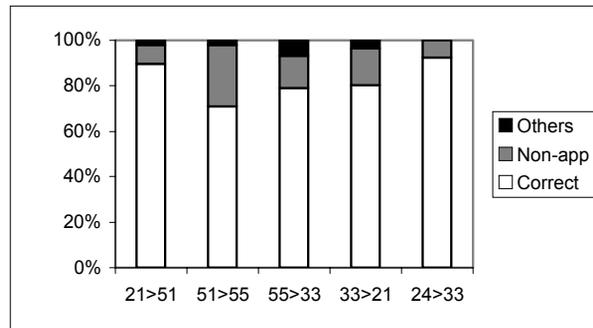
We also tabulated the numbers of responses in which the sandhi simply did not apply on the first syllable, i.e., the first syllable kept the base tone. Together with the cases of correct response, they account for the vast majority of all test items, as shown in the graphs in (8). It should be mentioned that both the correct response and non-application cases reported here include those in which the tone of the second syllable was changed, which account for around 6% of all the valid responses. For example, for a monosyllable 55, both 33-55 and 33-33 are counted as correct responses, as both have the correct sandhi tone on the first syllable; likewise, both 55-55 and 55-51 are counted as non-applications, as both have an initial syllable that is identical to the monosyllable in tone. Second syllable changes when the first syllable has the correct sandhi tone ($55 \rightarrow 33-33$), however, are extremely rare, and the vast majority of the second syllable changes occurred when the first syllable remained unchanged ($55 \rightarrow 55-51$). Therefore, the non-application rates should be seen as a reflection of the psychological reality of the phonotactic generalizations on legal tones on the first syllable, not of true non-application of the sandhis.

(8) Correct response and non-application rates for sandhis in single reduplication (σ_1):

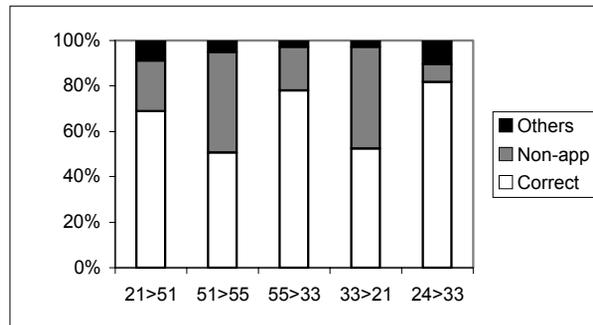
a. AO:



b. *AO:



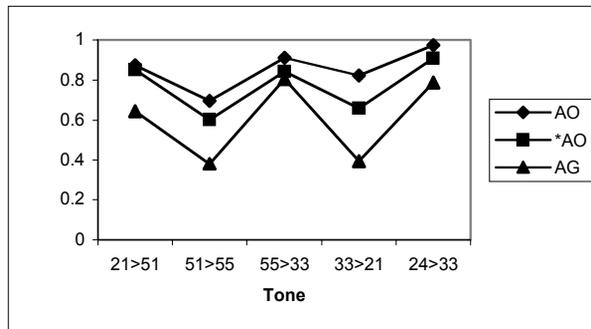
c. AG:



The correct response rates for the sandhi tones in double reduplication, which appear on the second syllable, are given in (9). This result is similar to that for single reduplication. The effect of word type is significant ($F(1.535, 19.954) = 67.486, p < 0.001$), so is the effect of tone ($F(3.267, 42.476) = 6.247, p < 0.01$). Their interaction is also significant ($F(5.579, 72.528) = 2.347, p < 0.05$). Posthoc analyses showed that the subjects'

performance conformed to the general sandhi pattern in AO more than *AO and AG, and in *AO more than AG (all at $p < 0.001$); and the 24 → 33 sandhi has a higher correct response rate than the 51 → 55 and 33 → 21 sandhis (all at $p < 0.05$). The 55 → 33 sandhi has a higher correct response rate than 51 → 55 and 33 → 21 in numerical values, but the differences do not reach statistical significance. The differences among different tones may be due to the combined effects of transparent phonotactic (high rate for 24 → 33), lexical frequency (low rate for 33 → 21; high rate for 55 → 33), and the durational property of the sandhi (low rate for 51 → 55).

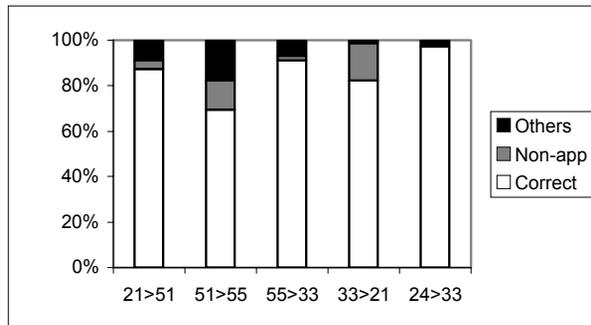
(9) Correct response rates for sandhi tones in double reduplication (σ_2):



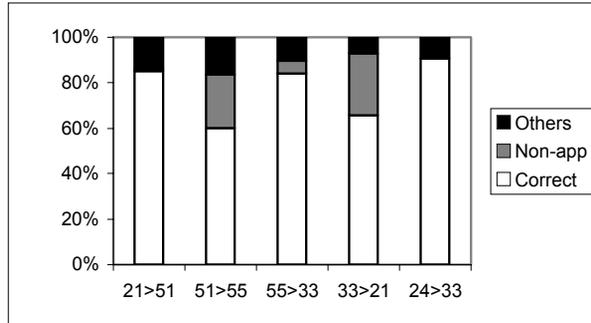
The numbers of correct response and non-application of the sandhis on the second syllable also add up to the vast majority of the test items, as shown in (10). Both the correct response and non-application cases again include those that had a changed tone on the last syllable, which also account for around 6% of the cases. The vast majority of the second syllable changes again occurred when the first syllable remained unchanged.

(10) Correct response and non-application rates for sandhis in double reduplication (σ_2):

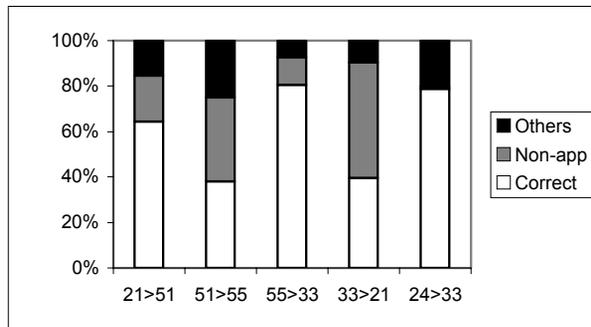
a. AO:



b. *AO:

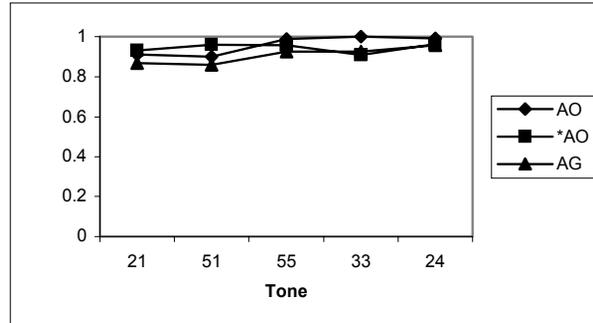


c. AG:



To evaluate the productivity of the floating High docking, we calculated the correct response rates for the tone on the first syllable, given the tone on the second syllable as the sandhi tone in double reduplication. For example, if the subject produced 21 on the second syllable, a 35 on the first syllable will be counted as a correct response for the first syllable, while other responses will be counted as incorrect, and this is regardless of whether or not 21 is the correct sandhi tone for the base tone. This result is given in (11): the correct response rates are close to 100% for all the word types and all the base tones, indicating that the speakers' knowledge of floating High docking is highly productive. The effects of word type and tone, however, do pass the $p < 0.05$ significance threshold (word type: $F(1.904, 24.746) = 3.519, p = 0.047$); tone: $F(3.036, 39.470) = 3.016, p = 0.041$). The interaction between the two factors is not significant ($F(6.702, 87.128) = 1.576, p > 0.05$).

(11) Correct response rates for σ_1 tone, given σ_2 tone as sandhi tone:



For each word type in double reduplication, we also compared the subjects' performances between the first two syllables. The comparison showed that for all word types, the subjects' correct response rates for the tone on the second syllable were significantly lower than those for the tone on the first syllable, provided that the tone on the second syllable was used as the base for floating High docking. AO: $F(1.000, 13.000) = 18.299$, $p < 0.001$; *AO: $F(1.000, 13.000) = 20.576$, $p < 0.001$; AG: $F(1.000, 13.000) = 71.776$, $p < 0.001$. This comparison indicates that there is a significant difference in productivity between the transparent pattern of floating High docking and the general tone sandhi pattern, which is largely opaque.

3.3 Discussion

Our hypotheses for the experiment are supported to various degrees by the results.

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 40 to 80%—there is a significant difference in productivity between many of the opaque sandhis and the phonotactically transparent sandhi 24 → 33, and between the sandhi patterns overall and the pattern of floating High docking. This indicates that there is indeed underlearning from the lexicon when the input-output mapping is opaque.

Second, among the opaque sandhis, we have observed that the durational property of the sandhi may have had a small effect: 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, since the sandhi occurs in a durationally disadvantaged position. This is an instance of overlearning from the lexicon.

Third, the high productivity of the transparent floating High docking finds strong

support from the results. The speakers seemed to have completely internalized both the presence of the floating High and how the docking site of this High is predicted, and were able to carry out this computation regardless of the sandhi tone that they selected for the reduplicants.

Fourth, as just stated, the transparent phonotactic generalization that no rising tone can appear on non-XP-final syllables (except on the first syllable of double reduplication) applied significantly more productively to novel words than the opaque base-sandhi mappings.

And lastly, we also 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 durational-reducing, but could potentially be related to the low lexical frequencies of the base tone 33 and the reduplicative melody 21-33. The higher numerical rate of the 55 → 33 application may also be due to the high lexical frequencies of the base tone 55 and the reduplicative melody 55-33.

Therefore, we contend that the results from our wug-test experiment show that the tone pattern of Taiwanese double reduplication indeed represents a case in which under-learning, overlearning, and proper learning from the lexicon coexist. In the next section, we turn to the task of modeling the speakers' variable behavior in the wug test with Zuraw's (2000) theory as a blueprint.

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

With the limited number of speakers and the difficulty in the experimental task (as evidenced by the relatively large percentage of speakers that did not produce useable data), we recognize that it is unlikely that our theoretical model can account for our subjects' behavior to in every detail. But we have set the following criteria for the model to be judged successful.

First, it must be able to account for both the exceptionless tone sandhi behavior in the Taiwanese lexicon and the variable tone sandhi behavior in the wug test (under-learning).

Second, it should be able to model the durational effect on productivity (over-learning).

Finally, it needs to have a mechanism that ensures that transparent and exceptionless generalizations can be completely productive and lexical statistics can have its effect felt in productivity (proper learning).

The basic theoretical framework we assume is a stochastic version of Optimality Theory developed by Boersma (1997, 1998) and Boersma & Hayes (2001). In this theory, constraints are ranked on a linear scale of strictness, with a higher strictness indicating

higher ranking. Moreover, a constraint occupies not just a single point on the scale, but a normally distributed range. At a particular instance of speaking, a ranking value is stochastically generated for each constraint according to its distribution, and the output is determined by the ranking of all constraints according to their ranking values. When the ranges of constraints overlap to a substantial degree, variation in the output is predicted, as the constraint ranking has a high likelihood of variation. Over a large number of trials, the frequency of occurrence of a particular output is proportional to the probability of the constraint rankings that generate it.

As mentioned earlier, we will also base our theory on the dual listing/generation model of Zuraw (2000). The 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 exceptionless patterns in the lexicon can be derived through highly ranked markedness constraints even without 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 must be listed in the lexicon. 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 to be completely productive in novel words.

4.1 USELISTED constraints

The gradation in sandhi productivity from AO to *AO to AG in the experiment indicates that the theory needs to encode three different levels of listedness. First, to account for the fact that the sandhis apply more productively in real words with real reduplication (AO) than real words without real reduplication (*AO), the tonal patterns of the singly reduplicated forms must have a certain degree of listedness in the lexicon. Simply listing the tonal allomorphs of existing syllables is not sufficient, as it can only predict a difference between real and wug words, not between the two types of real words. Second, to account for the fact that the sandhis apply more productively in real words without real reduplications (*AO) than wug words (AG), non-XP-final tonal allomorphs of existing syllables must also be listed to some extent. Simply listing the disyllabic forms can only predict a difference between words with and without reduplication and is hence insufficient. Finally, to account for the moderate degree of productivity of tone sandhis in wug words, the opaque sandhi patterns independent of segmental contents (i.e., 55 → 33; not pa51 → pa55) must also be directly stated in the grammar. Otherwise, the sandhis would be completely unproductive in AG words.

Therefore, for an existing syllable with existing reduplications, e.g., /pa51/, its single 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, e.g., /te21/, has a listed non-XP-final allomorph /ta51/, but no listed single 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.

It is also crucial for us to recognize that different lexical listings have different strengths; i.e., they are not equally available. Following Zuraw (2000), we assume that the listedness of lexical items is proportional to the number of times they have been heard, i.e., their token frequencies. We also assume that the listedness of the non-XP-final allomorphs of existing syllables and the listedness of non-XP-final tonal allomorphs of existing tones are also proportional to their token frequencies.

To implement the three different types of listedness in the grammar, we define three classes of USELISTED constraints as in (12). The constraints in (12a) require that real words with existing reduplications use the listed reduplicative forms; the constraints in (12b) require an existing syllable to use its listed allomorph in non-XP-final positions; and the constraints in (12c) require appropriate tonal allomorphs to be used for existing tones in non-XP-final positions regardless of the segmental contents.

(12) 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 (13) 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 (/σ51/), 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 non-final syllable maintains the base tone, for example, all three USELISTED constraints are violated. If the input syllable is an existing word without an existing reduplication (*AO: /σ21/), then it lacks the listed disyllabic reduplication. Therefore, when a sandhi tone that does not conform to the sandhi pattern is used, only USELISTEDσ21) and USELISTED(21) are violated. Finally, if the input syllable is a novel word, then it lacks both a listed disyllabic reduplication and a non-XP-final allomorph. Thus, an aberrant application of the sandhi on the reduplicant only incurs the violation of the USELISTED constraint that states the general lexical pattern —USELISTED(21). In these evaluation, we assume that IDENT-IO(Tone) is ranked high, and all candidates in the tableaux satisfy IDENT-IO(Tone). For example, the second candidate in the first tableau—[σ21-σ21]—is not derived from the listed /σ51-σ21/ via an IDENT-IO(Tone) violation.

(13) The evaluation of USELISTED constraints:

AO: /(RED)-RED-σ21/ Listed: /σ51-σ21/ Listed: /σ51-X/ Listed: /51-X/	USELISTED (σ51-σ21)	USELISTED (σ21)	USELISTED (21)
(σ)-σ51-σ21			
(σ)-σ21-σ21	*	*	*
*AO: /(RED)-RED-σ21/ Listed: /σ51-X/ Listed: /51-X/	USELISTED (σ51-σ21)	USELISTED (σ21)	USELISTED (21)
(σ)-σ51-σ21			
(σ)-σ21-σ21		*	*
AG: /(RED)-RED-σ21/ Listed: /51-X/	USELISTED (σ51-σ21)	USELISTED (σ21)	USELISTED (21)
(σ)-σ51-σ21			
(σ)-σ21-σ21			*

To encode the effect of frequency on the listedness of lexical items and lexical patterns, we posit a priori rankings among the USELISTED constraints according to the relevant token frequencies, as in (14). The ranking in (14a) is based on the token frequencies of tonal melodies in single reduplication given in (6), and the rankings in (14b) and (14c) are based on the token frequencies of base tones given in (5c). We recognize that (14a) and (14b) should likely be determined on a syllable-by-syllable basis; but given that our goal of the simulation here is only to model the overall pattern of the wug-test result with respect to tones without the details of each syllable, we consider the token frequencies

for the tone as a viable approximation for the different syllables used for the same tone in the experiment.⁷

(14) A priori rankings:

- a. USELISTED(σ_{33} - σ_{55}) » USELISTED(σ_{55} - σ_{51}) » USELISTED(σ_{33} - σ_{24}) » USELISTED(σ_{51} - σ_{21}) » USELISTED(σ_{21} - σ_{33})
- b. USELISTED(σ_{55}) » USELISTED(σ_{24}) » USELISTED(σ_{33}) » USELISTED(σ_{51}) » USELISTED(σ_{21})
- c. USELISTED(55) » USELISTED(24) » USELISTED(33) » USELISTED(51) » USELISTED(21)

4.2 Durationally based tonal markedness constraints

We have argued that the overlearned aspect of the sandhi pattern observed in the wug test results from the durational property of the sandhis: given that the sandhi occurs in non-final positions, which have shorter durations than the final position due to the lack of final lengthening (Oller 1973, Klatt 1975, Wightman et al. 1992, among others), sandhi tones that are intrinsically shorter in duration should be favored. Phonetic studies by Lin (1988) and Peng (1997) have shown that the two falling tones 51 and 21 are considerably shorter than the other three tones 55, 33, and 24. Therefore, we need a mechanism that favors 51 and 21 in non-final positions.⁸

⁷ 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 (p.50) as “The input portion of a candidate must be a single lexical entry.” We have adopted a simpler assumption that the candidate that is identical in form to the listed reduplication is necessarily derived from the listed form, as IDENT-IO is highly ranked and too costly to violate. Third, Zuraw uses only one USELISTED constraint, and encodes the strength of a lexical entry by equating the listedness of the lexical entry (from 0 to 1) to the availability of the lexical entry in the derivation of the output. We have adopted the “constraint family with a priori rankings” approach for its easier implementation in the Gradual Learning Algorithm, which we used to model our data.

⁸ It should be recognized that the sandhi domain in Taiwanese is not a prosodic domain, but a syntactic domain (Chen 1987, Lin 1994). This begs the question whether final lengthening can in fact be induced to motivate the sandhis phonetically. However, unlike non-XP-final syllables, which do not have opportunities to appear before a large prosodic boundary, an XP-final syllable does. Therefore, the average duration of an XP-final syllable in connected discourse will be greater than the same syllable in non-XP-final positions. A similar point was made by Zhang (to appear) to argue for the positive effect of word-final syllable on contour tone licensing despite its limited degree of lengthening as compared to phrase-final syllables.

To reflect the speakers' knowledge on the phonotactics of tones on non-XP-final syllables, we propose the tonal markedness constraints in (15a). Intuitively, *24-NON-FINAL should be quite highly ranked in the grammar of Taiwanese speakers, as the generalization is surface-true, while the other constraints should not be too highly ranked, as they are violated by many output data. We do not consider the first syllable of double reduplication to fall under the purview of these constraints, as it has different prosodic properties than the rest of non-XP-final syllables due to the docking of the floating High tone. For instance, our acoustic study of eight Taiwanese speakers has shown that this syllable has a considerably longer duration than the first syllable in single reduplication. (Zhang & Lai 2007 report the f0 result from this study, but do not discuss the duration result as it does not directly pertain to the gist of that paper.)

The effect of duration on tonal licensing is reflected in the a priori ranking in (15b), which penalizes a longer tone more severely than a shorter tone on non-XP-final syllables. These tonal markedness constraints and their a priori ranking are derivatives of the theory of contour tone licensing in Zhang (2002, 2004a, b). However, the proposal here uses the intrinsic durations of the tones as the predictor of their distribution rather than just the contour shapes. This is crucial in predicting that 51 → 55 is in fact a disfavored tone sandhi despite the fact that it is contour-reducing.

- (15) Tonal markedness constraints:
- a. *55-NONFINAL (*55-NF): Tone 55 cannot occur on non-XP-final syllables.
 - *33-NONFINAL (*33-NF): Tone 33 cannot occur on non-XP-final syllables.
 - *24-NONFINAL (*24-NF): Tone 24 cannot occur on non-XP-final syllables.
 - *21-NONFINAL (*21-NF): Tone 21 cannot occur on non-XP-final syllables.
 - *51-NONFINAL (*51-NF): Tone 51 cannot occur on non-XP-final syllables.
 - b. *55-NF, *33-NF, *24-NF » *21-NF, *51-NF

4.3 Constraints governing floating High docking

Finally, we need a set of constraints that governs the behavior of floating High docking in double reduplication. Together with Yip (1990), Wu (1996), Myers & Tsay (2001), and Lin (2004), we assume that intensification for monosyllabic adjectives is marked by both double reduplication and a floating High tone. The constraint in (16) requires the floating tone to be realized in the output.

- (16) REALIZE(Float): A floating tone must be realized in the output.

As we have argued, the location of the floating High insertion is determined as follows: the High is inserted to the left edge if it does not disrupt the identity between

the two reduplicated syllables at the left edge; otherwise the High is inserted on the right. To capture this intuition, we propose an ALIGNMENT constraint as in (17), which requires the floating tone to dock onto the left edge of the word, and a tonal correspondence constraint as in (18), which demands the identity between the initial tones of the two reduplicants. Plainly, in the grammar, ID-RR(Tone, Left) needs to outrank ALIGN(Float, Left, Word, Left).

- (17) ALIGN(Float, Left, Word, Left) (ALIGN-L): The left edge of a floating tone must be aligned with the left edge of a word.
- (18) IDENT-RR(Tone, Left) (ID-RR(T, L)): The left edges of the tones of two reduplicants derived from the same base must be identical.

The necessity of RR correspondence here echoes earlier proposals by Urbanczyk (2001) and particularly Lin (2004), who proposed an extended model of correspondence that includes RR correspondence to account for double reduplication in Taiwanese. A general RR correspondence constraint on tone can be stated as in (19).

- (19) IDENT-RR(Tone) (ID-RR(T)): The tones of two reduplicants derived from the same base must be identical.

4.4 Constraint ranking in stochastic OT

To find a stochastic OT grammar that produces outputs that match our wug test result, we used the Gradual Learning Algorithm (Boersma 1997, 1998, Boersma & Hayes 2001) in *OTSoft* (Hayes et al. 2003). We made the following simplifications in our attempts to model the wug test result.

First, we only attempted to model the tone changes in non-final syllables and assumed that the tone on the last syllable remained unchanged. As we have discussed, although this was the case for the vast majority of the responses in our wug test, a handful of the tokens (around 6%) did have a changed tone on the final syllable, even though the tone on the monosyllable was pronounced correctly. Many of these cases seem to simply be mistakes on the speakers' part. But in some of these tokens, the base tone remained on the first syllable, while the second syllable was changed to match one of the disyllabic tonal melodies used in reduplication; e.g., 51 was reduplicated as 51-21. We surmise that this has to do with the combination of the weakness of the phonotactic ban on 51 in non-final positions, the weakness of the listedness of the 51 → 55 mapping, and the relative strength of the listedness of disyllabic tonal melodies. But we did not attempt to formally model this effect.

Second, we only attempted to model the correct responses and non-applications of

the sandhis in both single and double reduplications. From the results in (8) and (10), we have seen that these cases account for the vast majority of the subjects' responses. We did not attempt to model any patterns that might have existed in the limited number of responses that did not represent either the correct or non-application of the sandhis.

Third, we attempted to model an ideal situation for AO words, where all responses had the correct application of sandhis. Although the speakers did not perform perfectly for these words in the experiment, we assume that they are able to use the reduplications correctly in real life, and their mistakes were due to the difficulty posed by the experimental setting. We recognized that this is a problem 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—the high productivity of transparent generalizations and the effects of both phonetics and lexical statistics—as the experimental difficulty should apply across the board to all tonal patterns in the reduplications.

With these provisos, the Graduate Learning Algorithm in *OTSoft* was used to find a grammar that best matched the wug test result. The initial ranking value of all constraints was set to 100, and the Gradual Learning Algorithm was provided with 100,000 input-output pairs, for which the inputs were randomly selected from the 30 possible inputs (AO: RED-21; AG: RED-RED-55, etc.), and the outputs matched the frequencies of occurrence of the forms in the wug test (with the simplifications mentioned above). The plasticity of the learning decreased from 0.1 to 0.001 in three steps during the 100,000 trials, and the a priori rankings were implemented such that two constraints that were intrinsically ranked must have a ranking value difference of at least 0.5. One such grammar as the result of this algorithm is given in (20).

(20) Ranking values for constraints in stochastic OT:

Ranking value	Constraint
121.117	USELISTED(σ 33- σ 55)
120.617	USELISTED(σ 55- σ 51)
120.487	REALIZE(Float)
120.117	USELISTED(σ 33- σ 24)
119.604	USELISTED(σ 51- σ 21)
118.290	USELISTED(σ 21- σ 33)
117.637	IDENT-RR(Tone, Left)
117.317	USELISTED(σ 55)
117.313	USELISTED(55)
116.814	USELISTED(σ 24)
116.798	USELISTED(24)
116.314	USELISTED(σ 33)
116.210	*24-NONFINAL
115.814	USELISTED(σ 51)

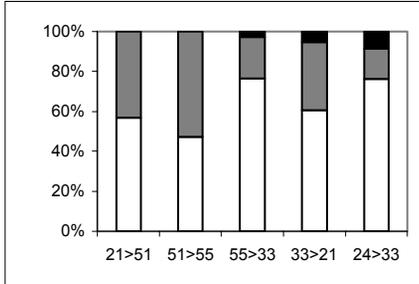
115.778	USELISTED(33)
115.314	USELISTED(σ 21)
115.278	USELISTED(51)
114.819	IDENT-BR(Tone)
114.778	USELISTED(21)
113.304	*55-NONFINAL
113.302	*33-NONFINAL
112.802	*21-NONFINAL
110.841	ALIGN(Float, Left, Word, Left)
89.463	IDENT-RR(Tone)
80.880	*51-NONFINAL

The grammar in (20) has the following properties in its constraint ranking. First, the USELISTED constraints requiring real words with existing reduplications to use the listed reduplications are ranked high. This accounts for the idealized stable behavior of real words with existing reduplications. Second, the USELISTED constraints that require an existing syllable to use its listed allomorph in non-XP-final positions are generally more highly ranked than the USELISTED constraints that require appropriate tonal allomorphs to be used for existing tones in non-XP-final positions regardless of the segmental contents. In particular, for any Tone_i, USELISTED(σ Tone_i) » USELISTED(Tone_i). This accounts for the higher sandhi productivity in *AO words than AG words. Third, among the constraints that govern the behavior of floating High docking, REALIZE(Float) and IDENT-RR(Tone, Left) are ranked very high, while ALIGN(Float, Left, Word, Left) and IDENT-RR(Tone) are ranked very low. This accounts for the speakers' productive application of the transparent floating High docking, even in wug words: the floating High must be realized, and it is inserted to the left of the first syllable if it does not disrupt the tonal identity between the left edges of the two reduplicants; otherwise it is inserted to the right of the first syllable. Fourth, among the tonal markedness constraints, *24-NONFINAL is ranked the highest, as it is the only one that reflects a true phonotactic generalization of Taiwanese tones. This accounts for the higher productivity of the 24 → 33 sandhi. *21-NONFINAL and *51-NONFINAL are ranked the lowest among tonal markedness, which on the one hand discourages base 21 and 51 to undergo sandhi in non-final positions, on the other hand encourages 21 and 51 to appear as the sandhi tones in non-final positions. Finally, the high ranking of USELISTED(σ 55) and USELISTED(55), partly due to the a priori rankings within their classes of USELISTED constraints, accounts for the relatively high productivity of the 55 → 33 sandhi despite its opacity.

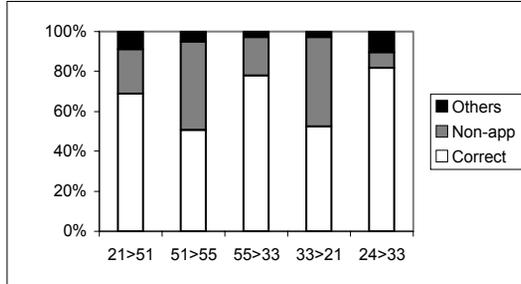
The tonal patterns generated by the stochastic grammar in (20), estimated by running the grammar 2000 times on each input, are juxtaposed with the wug test results in (21)-(25). The figures in (21) and (22) represent the patterns of tone sandhi in single reduplication, which occurs on the first syllable; the figures in (23) and (24) represent

c. AG:

Grammar output:

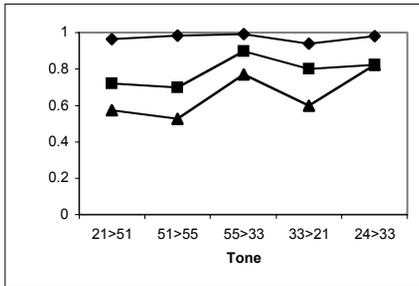


Wug test result:

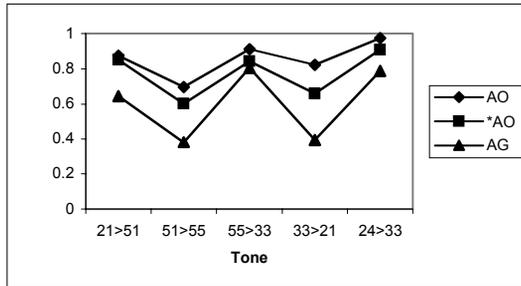


(23) Correct response rates for sandhi tones in double reduplication (σ_2):

Grammar output:



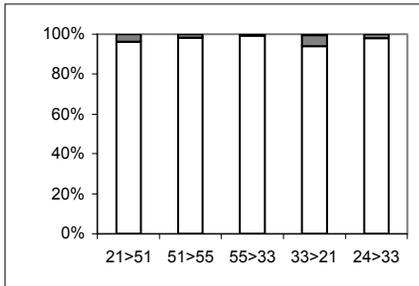
Wug test result:



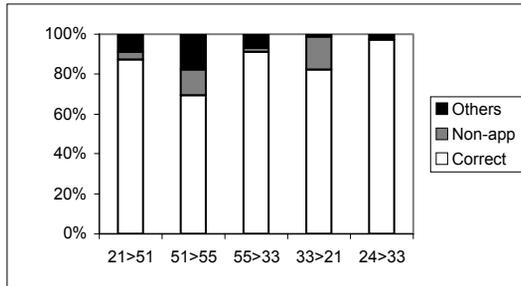
(24) Correct response and non-application rates for sandhis in double reduplication (σ_2):

a. AO:

Grammar output:

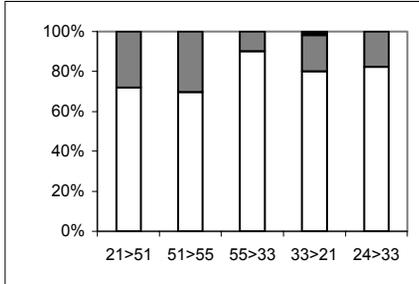


Wug test result:

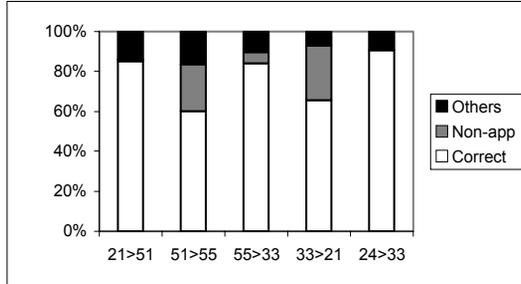


b. *AO:

Grammar output:

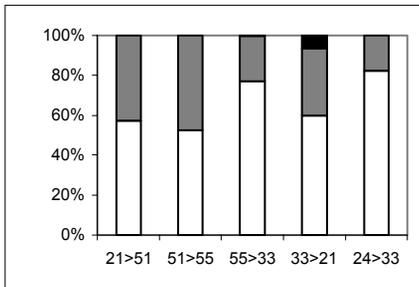


Wug test result:

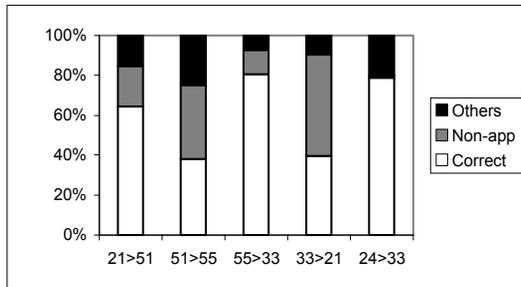


c. AG:

Grammar output:

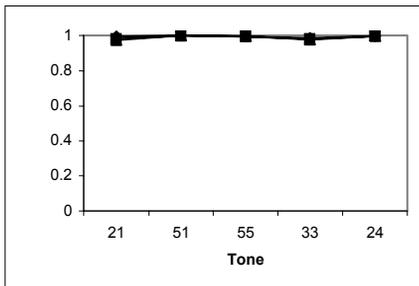


Wug test result:

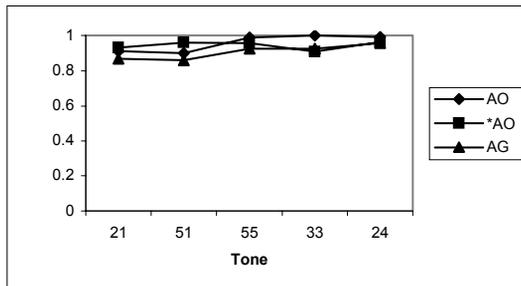


(25) Correct response rates for σ_1 tone, given σ_2 tone as sandhi tone:

Grammar output:



Wug test result:



As we can see from the comparison between the prediction of the stochastic grammar and the wug test results given in (21)-(25), our grammar succeeds in modeling all the effects that it sets out to model: it successfully models the behavior of the under-learned opaque tone circle in wug words despite its exceptionlessness in the lexicon; it captures the effect of phonetic duration on the productivity of the opaque sandhis despite the lack of such information in the lexicon; and it correctly predicts that transparent and

exceptionless generalizations from the lexicon are completely productive and lexical statistics can also have an effect on productivity. The grammar does predict more accurate applications of the sandhis in AO than the wug results entail, but this is because what the grammar sets out to model is the real life situation in which these real words are used, presumably with close to 100% accuracy, not the wug results, as discussed at the beginning of this subsection. The grammar also generally underrepresents any variations beyond the correct and non-application of the sandhis; this is also due to a simplifying assumption of the model discussed earlier—such variations in the wug test are few and far between, and many of them simply seem to represent mistakes on the speakers' part; we therefore made no attempt to model them.

The success of the model relies on a number of its crucial properties. The under-learned knowledge from the lexicon is due to the dual listing/generation nature of the model: opaque patterns are listed in the lexicon and high-ranking USELISTED constraints ensure the exceptionlessness of the lexical items; but the lack of high-ranking markedness constraints that motivate the opaque patterns prevent them from being productive in novel words. 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. Transparent and exceptionless generalizations from the lexicon are properly learned in the grammar as highly ranked markedness constraints. Finally, overlearning from the lexicon due to phonetic knowledge is encoded as phonetically based a priori constraint rankings.

4.5 Illustrative tableaux

We present a number of OT tableaux in this section to illustrate how the representative variants in both single and double reduplications can be derived from the stochastic grammar. We only illustrate with AG words here, for which the only relevant USELISTED constraints are USELISTED(T_i), as USELISTED(σT_i) and USELISTED(σT_j - σT_i) (σT_j = the non-XP-final allomorph of σT_i) are vacuously satisfied. For *AO, USELISTED(σT_i) constraints are also relevant; and for AO, USELISTED(σT_i) and USELISTED(σT_j - σT_i) are both relevant.

Let us first consider the single reduplication of an AG syllable with a 51 tone. The relevant constraints are USELISTED(51), *55-NONFINAL, *51-NONFINAL, and IDENT-BR(Tone), which have the ranking values 115.278, 113.304, 80.880, and 114.819, respectively. The ranking values of USELISTED(51), *55-NONFINAL, and IDENT-BR(Tone) are quite close to each other, which means that the ranking among the three constraints is relatively free; but they are all considerably greater than the ranking value of *51-NONFINAL, which means that the chances that *51-NONFINAL outranks any of these three constraints are close to none.

The tableau in (26) illustrates that both [55-51] and [51-51] may be the predicted output form. [55-51] will be the winner if USELISTED(51) outranks both IDENT-BR(Tone) and *55-NONFINAL at the instance of speaking; [51-51] will be the winner if either IDENT-BR(Tone) or *55-NONFINAL outranks USELISTED(51). Based on the ranking values, the probability that USELISTED(51) outranks both IDENT-BR(Tone) and *55-NONFINAL is 0.473. Therefore, [55-51] is predicted 47.3% of the time, while [51-51] is predicted 52.7% of the time. This matches relatively well with the wug test result, in which [55-51] surfaced 51% of the time, while [51-51] surfaced 44% of the time.

(26) AG: /RED-51/ → [55-51], [51-51]

AG: /RED-51/	USELISTED(51)	ID-BR(T)	*55-NF	*51-NF
☞ 55-51		*	*	
☞ 51-51	*			*

For double reduplication of AG words, we illustrate the prediction of the grammar for base tones 21 and 55. The AO patterns for these base tones are [51-51-21] and [35-21-33], respectively, with the floating High inserted to the left of the first syllable for the former, but to the right of the first syllable for the latter. The relevant constraints for floating High docking are REALIZE(Float), IDENT-RR(Tone, Left), ALIGN(Float, Left, Word, Left), and IDENT-RR(Tone), which have the ranking values 120.487, 117.637, 110.841, and 89.463, respectively.

For /RED-RED-21/, additional relevant constraints are USELISTED(21), *21-NONFINAL, *51-NONFINAL, and IDENT-BR(Tone), which have the ranking values 114.778, 112.802, 80.880, and 114.819, respectively. The eight relevant constraints roughly fall into three groups. REALIZE(Float) and IDENT-RR(Tone, Left) are highly ranked; IDENT-BR(Tone), USELISTED(21), *21-NONFINAL, and ALIGN(Float, Left, Word, Left) are clustered in the next stratum, and IDENT-RR(Tone) and *51-NONFINAL are lowly ranked.

The tableau in (27) illustrates that the grammar predicts that two forms are the most likely outputs for the double reduplication of a wug 21: [51-51-21] and [35-21-21]; the former is the expected output for real words, while the latter has non-application of the sandhi on the second syllable, but docks the floating High “correctly”—to the right edge of the first syllable in order to preserve the left-edge identity between the two reduplicant tones. The second candidate [55-51-21], with the floating High docking on the right of the first syllable, is harmonically bound by [51-51-21] and will never be the winner. The fourth candidate [51-21-21], with left-docking of the floating High despite the incorrect 21 sandhi tone on the second syllable, violates the highly ranked IDENT-RR(Tone, Left) and has very slim chances of winning the competition. The last candidate, which fails to realize the floating High, is also only predicted to win on very rare occasions due to its violation of the high ranking REALIZE(Float). Between the two

more likely winners, [51-51-21] is the predicted output if either *21-NONFINAL or ALIGN(Float, Left, Word, Left) outranks both IDENT-BR(Tone) and USELISTED(21); [35-21-21] is the predicted output if either IDENT-BR(Tone) or USELISTED(21) outranks both *21-NONFINAL and ALIGN(Float, Left, Word, Left). Based on the ranking values, [51-51-21] is predicted to win 57.4% of the time, while [35-21-21] should surface 40.1% of the time. The rest of the outputs is distributed between [21-21-21] (2.3%) and 51-21-21 (0.3%). This means that the speakers should produce the correct sandhi tone 51 on the second syllable 57.4% of the time, apply no sandhi to the second syllable 42.6% of the time, and apply floating High docking “correctly” 97.5% of the time. Except for underrepresenting the variations beyond the correct and non-application of the sandhi, the grammar’s prediction is again relatively close to the wug test result, which showed a 64.5% correct application rate and a 20.1% non-application of the sandhi on the second syllable, and an 86.8% of the “correct” application of floating High docking on the first syllable.

(27) AG: /RED-RED-21/ → [51-51-21], [35-21-21]

AG: /RED-RED-21/	REALIZE (Float)	ID-RR (T, L)	ID-BR (T)	USELISTED (21)	*21-NF	ALIGN-L	ID-RR (T)	*51-NF
☞ 51-51-21			**	**				*
55-51-21			**	**		*	*	*
☞ 35-21-21			*	*	*	*	*	
51-21-21		*	*	*	*		*	
21-21-21	*			*	*	*		

For /RED-RED-55/, the relevant constraints beyond those for floating High docking are USELISTED(55), *55-NONFINAL, *33-NONFINAL, and IDENT-BR(Tone), which have the ranking values 117.313, 113.304, 113.302, and 114.819, respectively. The eight relevant constraints roughly fall into four groups here. REALIZE(Float) is the highest ranked; IDENT-RR(Tone, Left) and USELISTED(55) are closely ranked in the next stratum, followed by the cluster IDENT-BR(Tone), *55-NONFINAL, *33-NONFINAL, and ALIGN(Float, Left, Word, Left); IDENT-RR(Tone) is clearly the lowest ranked.

The tableau in (28) evaluates the crucial candidates against this grammar. The most likely winner is clearly [35-33-55]—the expected output for real words, as it does not violate any constraints in the top two strata. The second candidate, which fails to register the floating High, violates the very highly ranked REALIZE(Float) and has practically no chance of winning. Between the last candidates that violate the two constraints in the second stratum respectively, the odds are in [55-55-55]’s favor: the constraint that it violates in the second stratum—USELISTED(55)—is the lower ranked of the two, and [53-33-55] has collectively many more violations of constraints in the

lower strata. Calculations based on the ranking values predict that [35-33-55] should win 76.7% of the time, and [55-55-55] 22.5% of the time. [53-33-55] has only a 0.5% chance of winning. This means that the speakers should produce the correct sandhi tone 33 on the second syllable 77.2% of the time, apply no sandhi to the second syllable 22.5% of the time, and apply floating High docking “correctly” 99.2% of the time. This again matches relatively well with the wug test result, which showed a 80.6% correct application rate and a 12.0% non-application of the sandhi on the second syllable, and a 92.4% of the “correct” application of floating High docking on the first syllable.

(28) AG: /RED-RED-55/ → [35-33-55], [55-55-55]

AG: /RED-RED-55/	REALIZE (Float)	ID-RR (T, L)	USELISTED (55)	ID-BR (T)	*55- NF	*33- NF	ALIGN- L	ID-RR (T)
☞ 35-33-55				**		*	*	*
33-33-55	*			**		*	*	
53-33-55		*		**		*		*
☞ 55-55-55			*		*			

5. Remaining issues

Although we have a proposed a model of phonological knowledge beyond the lexicon that can account for the patterns observed in the wug test of Taiwanese double reduplication, our model leaves a number of significant questions unanswered. 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 Learning

The most pressing of the unresolved issues is how such a grammar can be learned by speakers of Taiwanese, who are only exposed to real words. To ensure that the speakers can successfully learn the grammar proposed above or something similar to it, two issues must be explicitly addressed.

First, although our work is in line with works such as Wilson (2003, 2006), Davidson (2005), and Zuraw (2007) in showing that there must exist a substantive bias in phonological learning, possibly in the form of a priori rankings of markedness constraints based on phonetic knowledge—in our case, the durationally based ranking among *T_i-NONFINAL constraints—we have not provided a learning simulation that shows that the speakers can arrive at the *exact* interaction between these markedness constraints and other constraints, *even with an a priori ranking*. In other words, the a priori ranking

among the tonal markedness constraints is presumably purely phonetically determined, but the final ranking here does not seem to be: *21-NONFINAL (112.802) and *51-NONFINAL (80.880), for example, are ranked quite apart from each other, even though the phonetics posits no a priori ranking between the two, nor does the lexicon provide such information. Consequently, the ranking values of ALIGN(Float, Left, Word, Left) and IDENT-RR(Tone) fall in-between these two constraints. We so far do not have a concrete answer to this conundrum except for the conjecture that we perhaps need a better understanding of both phonetically based markedness and the way in which substantive biases are encoded in phonological learning.

Second, the learning procedure also needs to ensure that the USELISTED constraints that require appropriate tonal allomorphs for existing tones regardless of the segmental contents—USELISTED(T_i)—do not creep up too high in the ranking hierarchy during learning; otherwise the tonal allomorphs will apply entirely productively to novel words. But this is unavoidable given the current implementation of the Gradual Learning Algorithm: the use of the tonal allomorphs is exceptionless in Taiwanese; therefore, these USELISTED constraints will gradually ascend along the ranking scale during learning to high rankings. This problem is essentially of the same nature as the one identified in Albright & Hayes (2006): constraints that capture accidentally exceptionless generalizations in Navajo sibilant harmony must be suppressed in ranking during learning to ensure the unproductiveness of the generalizations. However, the solution that Albright & Hayes sketched out for Navajo is not applicable to the case here. The accidental generalizations in Navajo all have relatively low generality, in that they only apply to a few forms that illustrate a particular allomorph; therefore, a low initial ranking of these constraints will keep these constraints low in the learning process and let the more general constraints that correctly capture the allomorph distribution do the work. But the situation here is exactly the opposite: the USELISTED constraints that capture the tonal allomorphy on an abstract level are in fact more general. Therefore, the mechanism that suppresses the ranking of these constraints cannot be based on generality, but something else. We tentatively suggest that all constraints governing the listing of input-output mappings based on any level of abstraction should be suppressed in ranking in some fashion, but we so far do not have a concrete way of achieving this.

5.2 The listing of abstract input-output mappings

We now turn to the nature of the USELISTED constraints that govern the allomorphy on an abstract level. The gradient differences in productivity among AO, *AO, and AG words in the wug test prompted our proposal that allomorphs must be listed on different levels of abstractness: the listing of real disyllabic reduplications including their tonal

melodies accounts for the productivity difference between AO and *AO; the listing of non-XP-final allomorphs of existing syllables including their tones accounts for the difference between *AO and AG; and crucially, the abstract tonal allomorphs must also 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 in phonology—the listing of abstract input-output mappings and the M » F 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 *NC_o » 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 constraints on abstract allomorph listing low despite their exceptionlessness. Transparent allomorphy then still derives from the M » F ranking, and the role of the USELISTED constraints is invisible. But opaque allomorphy cannot be derived from any M » F rankings, and the ranking 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, 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: cross-linguistic variation is predicted by differences in constraint ranking.

5.3 How phonology interacts with morphology and syntax

Finally, we must acknowledge that our analysis of the productivity patterns in Taiwanese double reduplication rests on a number of assumptions on how phonology interacts with morphology and syntax. Our treatment of syntactically based allomorphy is consistent with the “precompiled phrasal phonology” of Hayes (1990), which considers

syntactically determined allomorphs to derive from lexical rules that refer to subcategorization frames with syntactic information within the lexicon. Our approach is also partly consistent with Paster (2006) and Bye (to appear) in considering phonologically conditioned suppletive allomorphy to be derived 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, to appear). 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.

6. Conclusion

We have argued in this paper that Taiwanese speakers' phonological knowledge on the tone pattern of double 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: the speakers prefer shorter tones as sandhi tones on non-final 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, transparent phonotactic generalizations on tones, such as *24-NONFINAL, and the transparent behavior of floating High docking are almost entirely productive, even in novel words.

With the stochastic OT of Boersma (1997, 1998) and Boersma & Hayes (2001) as the backdrop, we have used the dual listing/generation theory of Zuraw (2000) to model the speakers' knowledge as reflected in the wug test. We have shown that Zuraw's theory is not only able to model patterned exceptionality based on the lexicon as shown in her own work, but also able to model underlearned knowledge from the lexicon due to opacity. The opaque mappings are essentially treated as exceptions and listed in the lexicon with various degrees of abstractness. The exceptionless lexical behavior is due to highly ranked USELISTED constraints that govern existing words, while the variable behavior in the wug test is due to lower-ranked constraints that encode the phonetic and frequency biases of the patterns. The transparent patterns, however, can be derived productively through highly ranked markedness constraints in both real and novel words.

Although we have left a number of open issues, most notably the issues of learning, we hope that we have demonstrated that (a) we can arrive at a more precise estimation 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 the speakers' phonological knowledge.

Finally, we hope that the pattern and the analysis of Taiwanese double reduplication have illustrated that a comprehensive understanding of phonology involves the understanding of how it interfaces with almost every other branch of linguistics. Double reduplication is an interesting morphological phenomenon in and of itself; the domain of Taiwanese tone sandhi lies at the interface between phonology and syntax (Chen 1987, Hsiao 1991, Lin 1994); the exceptional behavior of the opaque patterns and frequency effects are properties of the lexicon; the preference for shorter tones on shorter syllables results from the effects of phonetics on phonology; and the methodology that we used to address these questions is undoubtedly psycholinguistic. By viewing phonology as a truly interface science, we provide ourselves with both a broader range of data and a broader range of analytical tools. We hope that the current volume will spawn more interface research that will shed light on the true nature of phonological patterning in both Chinese languages and elsewhere.

Appendix: Stimuli for the Wug-Test

Tone	AO		*AO			AG	
21	膨	p ^h oŋ	‘inflated’	進	tsin	‘to enter’	tsɔ
	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 ɔ	‘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 āi
	笑	ts ^h io	‘to laugh’	教	ka	‘to teach’	t ^h ua
	夠	kau	‘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	pɔ	‘dried’	表	piau	‘cousin’	k ^h ŋ
	少	tsio	‘few’	走	tsau	‘to run’	pai
	普	p ^h ɔ	‘ordinary’	手	ts ^h iu	‘hand’	p ^h au
	鬼	kui	‘impish’	請	ts ^h iã	‘to treat’	t ^h ŋ
歹	p ^h āi	‘bad’	統	t ^h oŋ	‘to unite’	tsam	
55	翹	k ^h iau	‘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’	pē
	甜	tī	‘sweet’	添	t ^h iam	‘to add’	p ^h ɔ
	芳	p ^h aŋ	‘fragrant’	車	ts ^h ia	‘car’	t ^h ɔ
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	pē	‘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	tsiŋ	‘quiet’	豆	tau	‘bean’	pa
	癢	tsiũ	‘itchy’	娶	ts ^h ua	‘to marry’	p ^h ai
	厚	kau	‘thick’	渡	tɔ	‘to ferry’	t ^h ɔ
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	tse	‘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 āi
	槌	t ^h ui	‘foolish’	奔	tsoŋ	‘to run quickly’	t ^h ua
著	tiau	‘sturdy’	茶	te	‘tea’	tsam	

References

- Albright, Adam. 2002. Islands of reliability for regular morphology: evidence from Italian. *Language* 78.4:684-709.
- Albright, Adam, and Bruce Hayes. 2003. Rules vs. analogy in English past tenses: a computational/experimental study. *Cognition* 90.2:119-161.
- Albright, Adam, and Bruce Hayes. 2006. Modeling productivity with the Gradual Learning Algorithm: the problem of accidentally exceptionless generalizations. *Gradience in Grammar: Generative Perspectives*, ed. by Gisbert Fanselow et al., 185-204. Oxford & New York: Oxford University Press.
- Albright, Adam, Argelia Andrade, and Bruce Hayes. 2001. Segmental environments of Spanish diphthongization. *UCLA Working Papers in Linguistics* 7:117-151.
- Berko, Jean. 1958. The child's learning of English morphology. *Word* 14:150-177.
- Boersma, Paul. 1997. How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* 21:43-58.
- Boersma, Paul. 1998. *Functional Phonology: Formalizing the Interactions between Articulatory and Perceptual Drives*. The Hague: Holland Academic Graphics.
- 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*. Retrieved November 2005, from <http://www.praat.org/>.
- Burzio, Luigi. 2002. Surface-to-surface morphology: when your representations turn into constraints. *Many Morphologies*, ed. by Paul Boucher, 142-177. Somerville: Cascadilla Press.
- Bye, Patrick. (to appear). Allomorphy—selection, not optimization. *Freedom of Analysis?*, ed. by Sylvia Blaho, Patrik Bye & Martin Krämer. Berlin: Mouton de Gruyter.
- Bybee, Joan. 2001. *Phonology and Language Use*. Cambridge & New York: Cambridge University Press.
- Bybee, Joan, and Elly Pardo. 1981. On lexical and morphological conditioning of alternations: a nonce-probe experiment with Spanish verbs. *Linguistics* 19.9-10: 937-968.
- Chen, Matthew Y. 1987. The syntax of Xiamen tone sandhi. *Phonology Yearbook* 4: 109-149.
- Chen, Matthew Y. 2000. *Tone Sandhi: Patterns across Chinese Dialects*. Cambridge & New York: Cambridge University Press.
- Cheng, Robert L. 1973. Some notes on tone sandhi in Taiwanese. *Linguistics* 100:5-25.
- Cheng, Robert L. 1997. *Taiwanese and Mandarin Structures and Their Developmental Trends in Taiwan: Taiwanese Phonology and Morphology*. Taipei: Yuanliu.

- Chung, Raung-fu. 1996. *The Segmental Phonology of Southern Min in Taiwan*. Taipei: Crane.
- Davidson, Lisa. 2005. Phonology, phonetics, or frequency: influences on the production of non-native sequences. *Journal of Phonetics* 34.1:104-137.
- Frisch, Stefan A., and Bushra Adnan Zawaydeh. 2001. The psychological reality of OCP-place in Arabic. *Language* 77.1:91-106.
- Hayes, Bruce. 1990. Precompiled phrasal phonology. *The Phonology-syntax Connection*, ed. by Sharon Inkelas & Draga Zec, 85-108. Chicago: University of Chicago Press.
- Hayes, Bruce, and Zsuzsa Cziráky Londe. 2006. Stochastic phonological knowledge: the case of Hungarian vowel harmony. *Phonology* 23:59-104.
- Hayes, Bruce, Bruce Tesar, and Kie Zuraw. 2003. *OTSoft 2.1*. Downloadable at <http://www.linguistics.ucla.edu/people/hayes/otsoft/>.
- Hsiao, Yuchau E. 1991. *Syntax, Rhythm and Tone: A Triangular Relationship*. Taipei: Crane.
- Hsieh, Hsin-I. 1970. The psychological reality of tone sandhi rules in Taiwanese. *CLS* 6:489-503. Chicago: Chicago Linguistic Society.
- Hsieh, Hsin-I. 1975. How generative is phonology. *The Transformational-generative Paradigm and Modern Linguistic Theory*, ed. by E. F. K. Koerner, 109-144. Amsterdam: John Benjamins.
- Hsieh, Hsin-I. 1976. On the unreality of some phonological rules. *Lingua* 38.1:1-19.
- Klatt, Dennis H. 1975. Vowel lengthening is syntactically determined in connected discourse. *Journal of Phonetics* 3.3:129-140.
- Lin, Chien-Jer. 2004. Beyond output correspondence: tone sandhi in Southern Min reduplication. Talk presented at the 78th LSA Annual Meeting, Boston.
- Lin, Hwei-Bing. 1988. *Contextual Stability of Taiwanese Tones*. Storrs: University of Connecticut dissertation.
- Lin, Jo-wang. 1994. Lexical government and tone group formation in Xiamen Chinese. *Phonology* 11:237-275.
- Lubowicz, Ania. 2003. *Contrast Preservation in Phonological Mappings*. Amherst: University of Massachusetts dissertation.
- MacBride, Alexander Ian. 2004. *A Constraint-based Approach to Morphology*. Los Angeles: University of California dissertation.
- McCarthy, John J. 1999. Sympathy and phonological opacity. *Phonology* 16:331-399.
- McCarthy, John J. 2007. *Hidden Generalizations: Phonological Opacity in Optimality Theory*. London & Oakville: Equinox.
- Myers, James, and Jane Tsay. 2001. Testing a production model of Taiwanese tone sandhi. *Proceedings of the Symposium on Selected NSC Projects in General Linguistics from 1998-2000*, 257-279.

- Oller, D. Kimbrough. 1973. The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America* 54.5:1235-1247.
- Paster, Mary Elizabeth. 2006. *Phonological Conditions on Affixation*. Berkeley: University of California dissertation.
- Peng, Shui-Hui. 1997. Production and perception of Taiwanese tones in different tonal and prosodic contexts. *Journal of Phonetics* 25.3:371-400.
- Pierrehumbert, Janet B. 2006. The statistical basis of an unnatural alternation. *Laboratory Phonology*, Vol. 8: *Varieties of Phonological Competence*, ed. by Louis Goldstein, Douglas H. Whalen & Catherine Best, 81-107. Berlin & New York: Mouton de Gruyter.
- Sanders, Nathan. 2001. Preserving synchronic parallelism: diachrony and opacity in Polish. *CLS* 37:501-515. Chicago: Chicago Linguistic Society.
- Sumner, Megan. 2003. *Testing the Abstractness of Phonological Representations in Modern Hebrew Weak Verbs*. Stony Brook: State University of New York dissertation.
- 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 (ICPhS'99)*, Vol. 3, 2407-2410. San Francisco, USA.
- Tsay, Jane, and James Myers. 2005. *Taiwanese Spoken Corpus*. Chiayi: National Chung Cheng University.
- Urbanczyk, Suzanne. 2001. *Patterns of Reduplication in Lushootseed*. New York: Garland.
- Wang, H. Samuel. 1993. Taiyu biandiao xianxiang de xinli texing [On the psychological status of Taiwanese tone sandhi]. *Tsing Hua Journal of Chinese Studies*, New Series 23.2:175-192.
- 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* 91.3:1707-1717.
- Wilson, Colin. 2003. Experimental investigation of phonological naturalness. *Proceedings of the 22nd West Coast Conference on Formal Linguistics (WCCFL-22)*, ed. by Gina Garding & Mimu Tsujimura, 533-546. Somerville: Cascadilla Press.
- Wilson, Colin. 2006. Learning phonology with substantive bias: an experimental and computational study of velar palatalization. *Cognitive Science* 30.5:945-982.
- Wu, Tzuling. 1996. *A Morphological Analysis of Reduplication in Taiwan Southern Min*. Hsinchu: National Tsing Hua University MA thesis.
- Yip, Moira. 1990. *The Tonal Phonology of Chinese*. New York: Garland.
- Zhang, Jie. 2002. *The Effects of Duration and Sonority on Contour Tone Distribution — A Typological Survey and Formal Analysis*. New York: Routledge.

- Zhang, Jie. 2004a. The role of contrast-specific and language-specific phonetics in contour tone distribution. *Phonetically Based Phonology*, ed. by Bruce Hayes, Robert Martin Kirchner & Donca Steriade, 157-190. Cambridge & New York: Cambridge University Press.
- Zhang, Jie. 2004b. Contour tone licensing and contour tone representation. *Language and Linguistics* 5.4:925-968.
- Zhang, Jie. (to appear). Contour tone distribution is not an artifact of tonal melody mapping. *Studies in the Linguistic Sciences* 33.1.
- Zhang, Jie, and Yuwen Lai. 2007. Two aspects of productivity in Taiwanese double reduplication. *Kansas Working Papers in Linguistics* 29:33-47.
- Zhang, Jie, Yuwen Lai, and Craig Turnbull-Sailor. 2006. Wug-testing the “tone circle” in Taiwanese. *Proceedings of the 25th West Coast Conference on Formal Linguistics (WCCFL-25)*, ed. by Donald Baumer, David Montero & Michael Scanlon, 453-461. Somerville: Cascadilla Press.
- Zhang, Jie, Yuwen Lai, and Craig Sailor. (to appear). Effects of phonetics and frequency on the productivity of Taiwanese tone sandhi. *CLS* 43. Chicago: Chicago Linguistic Society.
- Zuraw, Kie. 2000. *Patterned Exceptions in Phonology*. Los Angeles: University of California dissertation.
- Zuraw, Kie. 2007. The role of phonetic knowledge in phonological patterning: corpus and survey evidence from Tagalog infixation. *Language* 83.2:277-316.

Jie Zhang
Department of Linguistics
The University of Kansas
1541 Lilac Lane
Blake Hall, Rm. 427
Lawrence, KS 66044
USA
zhang@ku.edu