# Phonological Knowledge beyond the Lexicon in Taiwanese Double Reduplication

**Jie Zhang and Yuwen Lai**
**The University of Kansas**

## 1. Introduction

In a recent trend, phonologists are becoming increasingly aware of the fact that many phonological processes are variable (Labov 1972, 1994), gradient (Hayes 2000, Frisch et al. 2004), and full of exceptions (Zimmer 1969, Zuraw 2000). This is not only because variability, gradience, and exceptions have been shown to be par for the course for many lexical patterns in phonology, but also because speakers have been shown to possess *knowledge* of variability, gradience, and patterns of exceptions (Albright and Hayes 2003, Frisch and Zawaydeh 2001, Hayes and Londe 2006, Pierrehumbert 2006, Zuraw 2000). For example, Frisch and Zawaydeh (2001) showed that Arabic speakers' wordlikeness judgments of nonce words corresponded with the gradient consonant cooccurrence restrictions in the Arabic lexicon.

Recent research has also shown that it does not suffice to study only the lexically manifested sound patterns observable from elicited data, as speakers' phonological knowledge does not always match the lexical patterns. There are areas in which speakers seem to know more than the lexicon (Berent et al. 2007, Davidson et al. 2004, Zuraw 2007). For example, Berent et al. (2007) showed that English speakers have knowledge of onset sonority sequencing that their lexicon does not inform them of, as evidenced by syllable counting, identity judgment, and identity priming experiments. There are also areas where speakers seem to know less than the lexicon, particularly for phonologically opaque patterns (Sanders 2001). The opaque Taiwanese "tone circle" (see details in §2), for instance, is repeatedly shown to be largely unproductive in "wug" tests (Berko 1958) despite its exceptionlessness in the language itself (Hsieh 1970, 1975, 1976, Wang 1993, Zhang et al. 2006, 2007)

The goal of this paper is two-fold. We first show that Taiwanese speakers' knowledge of tone sandhi in double reduplication, as reflected in a wug test, is a combination of more than, less than, and exactly what their lexicon informs them of. Second, we argue that the "dual-listing/generation" theory of Zuraw's (2000) can model the phonological knowledge of Taiwanese speakers.

## 2. Tonal Patterns in Taiwanese Double Reduplication

### *2.1. Taiwanese Tone Sandhi*

Taiwanese tone sandhi is a tonal alternation pattern that affects tones on nonfinal syllables of a syntactic phrase XP (Chen 1987, Lin 1994). As shown in (1), it is characterized by an opaque "tone circle" that involves four of the five tones in the inventory and a phonotactically transparent sandhi 24 → 33 based on the true generalization that the rising tone 24 cannot occur in nonfinal positions.

(1) Taiwanese "tone circle":

```
51 → 55 → 33 ← 24
  ↖      ↙
     21
```

* We are grateful to Jane Tsay and James Myers for providing us with a Taiwanese frequency corpus, Paul Boersma and Mietta Lennes for helping us with Praat scripts, and Craig Sailor for his help with data processing. All remaining errors are strictly our own. This 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.

 Proceedings of the 27th West Coast Conference on Formal Linguistics, ed. Natasha Abner and Jason Bishop, 458-466. Somerville, MA: Cascadilla Proceedings Project.

### *2.2. Taiwanese Double Reduplication*

Reduplication is a productive morphological process in Taiwanese, and the meaning of reduplication, depending on the part of speech being reduplicated, ranges from intensification, tentativeness, to repetition and continuity (Wu 1996). Monosyllabic adjectives can be reduplicated in two different ways in Taiwanese: single reduplication diminishes the meaning of the adjective, while double reduplication intensifies the meaning; e.g., *kin51* 'fast,' *kin55-kin51* 'somewhat fast,' *kin55-kin55-kin51* 'very fast.' Since the tone sandhi occurs in nonfinal positions, we assume that in both types of reduplication, the final syllable is the base and the nonfinal syllables are the reduplicants.

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 tonal patterns are summarized in (2).

(2) 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 |

The tone on the first syllable of double reduplication has a straightforward analysis *provided that the sandhi tone is known*: it is a combined result of the sandhi tone and a floating High, inserted to the left edge of the syllable (vacuously) when the sandhi tone starts High (51 and 55), and to the right edge when the sandhi tone starts Mid (33 and 21) to preserve the tonal identity of the left edges of the reduplicant syllables (Yip 1990, Lin 2004). This pattern of floating High docking is transparent and easily analyzable in either rule-based or OT phonology.

The derivation of the opaque sandhi tones, however, is problematic for standard Optimality Theory (Moreton 2004). Moreover, earlier works on the productivity of Taiwanese tone sandhi indicate that speakers may exhibit both underlearning and overlearning from the lexicon: the opaque nature of the tone circle may cause the speakers to underlearn the pattern from the lexicon; but the different phonetic bases among the different sandhi processes may also provide the speakers with knowledge that they cannot glean from lexical statistics. 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 nonfinal syllables, which are known to be shorter than final syllables due to the lack of final lengthening (Wightman et al. 1992, etc.), 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. The effects of duration on the productivity Taiwanese tone sandhi have been shown in Zhang et al. (2006, 2007).

The upshot, then, is that Taiwanese tone sandhi in double reduplication is a good test case for both the interaction between lexical and nonlexical factors on phonological productivity and a theory that allows underlearned, overlearned, and properly learned patterns from the lexicon to coexist in phonological grammar.

## 3. A Wug Test Experiment

### *3.1. Methods*

To this end, we designed a wug test experiment in which the subjects' basic task was to produce the single and double reduplications of monosyllabic words. The stimuli included three types of monosyllabic words: real Taiwanese words with actual occurring reduplications (AO), real Taiwanese words with no actual occurring reduplications (*AO), and accidental gaps in the Taiwanese syllabary

(AG). Each word type included all five tones, and for each word type and tone combination, eight different words were used, which made 120 test words.

The test was conducted with SuperLab. The subjects first heard an introduction which explained in both prose and examples that Taiwanese can mark the lesser degree of an adjective by saying it twice and its intensification by saying it three times, and their task was to repeat the monosyllabic stimuli they heard from a headphone, and then form the diminutive and intensive forms of the words in turn. 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. The 120 stimuli were randomized for each subject.

Our data from 14 subjects (4 male, 10 female) were collected by the second author during two field trips to Taiwan in the summers of 2006 and 2007. The subjects had an average age of 50.2 at the time of the experiment.

### *3.2. Hypotheses*

We hypothesized that the wug test results would reflect underlearning, overlearning, as well as proper learning from the lexicon.

Underlearning was expected in the subjects' responses for sandhi tones in the tone circle, as reflected in the low correct response rates for the sandhi tones on the first syllable of single reduplication and the second syllable of double reduplication in wug words.

Overlearning was also expected for the opaque sandhis in that the durational property of a sandhi would have an effect on its productivity.

Proper learning was expected in three areas. First, the phonotactically driven sandhi 24 → 33 should have a higher productivity. Second, the floating High docking on the first syllable of double reduplication should be highly productive if we assume that the tone on the second syllable was what the subject considered to be the sandhi tone. For example, for a base tone 21, if the subjects produced the correct sandhi tone 51 on the second syllable, then we would expect them to produce also 51 on the first syllable by docking the floating High to the left; but if they produced the wrong sandhi tone 21 on 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 — thus, the first tone would be 35, as the floating High would now have to dock onto the right edge of the syllable. Lastly, we also expected lexical frequencies, in particular, the token frequencies of base tones and the disyllabic reduplicative melodies as shown in (3) (based on Tsay and Myers 2005), to have an effect on productivity. Our hypothesis on the relevance of token frequencies instead of type frequencies was based on our earlier works on Taiwanese tone sandhi (Zhang et al. 2007).

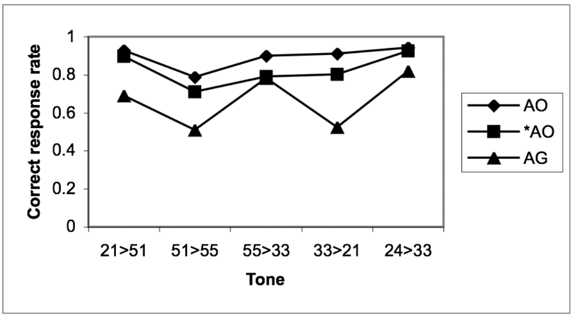
(3) Token frequencies of tones and tonal melodies:
  a. Tones: 55 > 24 > 33 > 51 > 21
  b. Tonal melodies in reduplication: 33-55 > 55-51 > 33-24 > 51-21 > 21-33

### *3.3. Results*

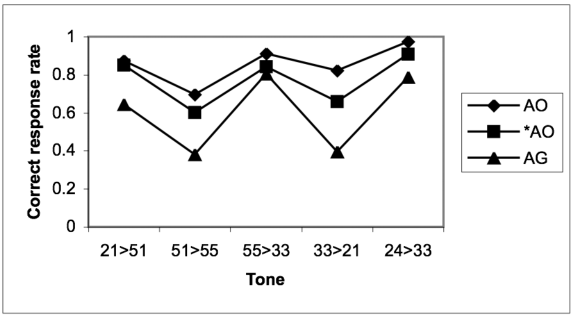
The correct response rates for the sandhi tones in single reduplication are given in (4). A Two-Way Repeated-Measures ANOVA showed that the effect of word type is significant ($F(1.877, 24.407) = 37.599$, $p<0.001$), so is the effect of tone ($F(3.759, 48.862) = 3.599$, $p<0.05$). The interaction between 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 more accurately in AO than in *AO ($p<0.05$), and in *AO than in AG ($p<0.001$). Posthoc tests also showed that there is a significant difference between 24 → 33 and 51 → 55 ($p<0.05$), but not other pairs. The 55 → 33 sandhi has a higher correct response rate than 51 → 55 in numerical values, but the difference does not reach statistical significance.

The correct response rates for the sandhi tones in double reduplication ($\sigma_2$ of $\sigma_1\sigma_2\sigma_3$), which show similar patterns to those in single reduplication, are given in(5). 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 ($F(5.579, 72.528) = 2.347$, $p<0.05$). Subjects performed better in AO than in *AO, and in *AO than in AG (both at $p<0.001$). 24 → 33 has a higher correct response rate than 51 → 55 and 33 → 21 (both at $p<0.05$). 55 → 33 has a higher correct response rate than 51 → 55 and 33 → 21 in numerical values, but the differences do not reach statistical significance.

(4) Correct response rates for sandhi tones in single reduplication ($\sigma_1$):

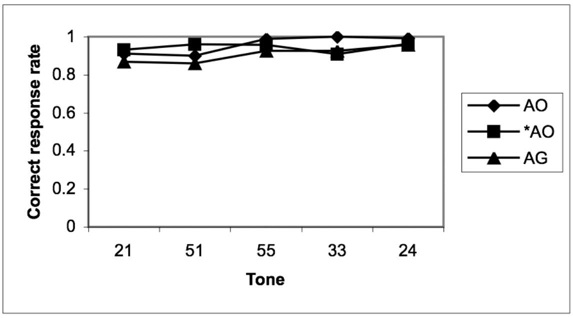



(5) Correct response rates for sandhi tones in double reduplication ($\sigma_2$):




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. This result is given in(6). 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. The interaction between the two factors is not significant ($F(6.702, 87.128) = 1.576, p>0.05$).

(6) Correct response rates for $\sigma_1$ tone, given $\sigma_2$ tone as sandhi tone:




### *3.4. Discussion*

Our hypotheses for the experiment are supported to various degrees by the results. There is indeed a certain degree of underlearning from the lexicon, in that the sandhis in the tone circle are less productive than 24 → 33 and floating High docking. Riding on this overall underlearned pattern is also a certain degree of overlearning, as reflected in the low productivity of the duration increasing sandhi 51 → 55. Proper learning is also reflected in the results: the transparent sandhi 24 → 33 and floating High docking in double reduplication have generally been properly learned from the lexicon, and the high productivity of 55 → 33 and the low productivity of 33 → 21 are most likely due to frequency effects, as 55 is by far the most frequent base tone, and 33 and 21 are among the rarer tones; 21-33 is also the rarest reduplicative melody.

## 4. A Theoretical Model of the Speakers' Wug-Test Behavior

We now turn to the modeling of our wug-test results. Our goals are to model the coexistence of opaque but exceptionless sandhis in the lexicon and the variable sandhi behavior in the wug test (underlearned pattern), the duration effects on sandhi productivity (overlearned pattern), the high productivity of transparent and exceptionless patterns in the lexicon, and the effects of lexical statistics (properly learned pattern). We assume a version of stochastic OT by Boersma and Hayes (2001). We also assume Zuraw's (2000) dual listing/generation theory, in which existing forms are lexically listed and protected by high-ranking faithfulness constraints and constraints that require the use of listed forms, but lower ranked constraints encode lexical statistic patterns and phonetically biased generalizations.

### *4.1. USELISTED Constraints*

The gradation in sandhi productivity from AO to *AO to AG indicates that the theory needs to encode three different levels of listedness. First, the higher correct response rates in AO than in *AO indicate that real reduplicated forms are listed in the lexicon; i.e., for an existing syllable with an existing reduplication like /kin51/, its reduplicated form has a listed lexical entry /kin55-kin51/. Second, the higher correct response rates in *AO than in AG indicate that sandhi allomorphs of existing syllables are also listed; for example, for /kin51/, its nonfinal allomorph is listed as /kin55/. Finally, to account for the moderate productivity of opaque sandhis in AG words, we also list the sandhi patterns independent of segmental contents directly in the grammar; for instance, the nonfinal allomorph of /51/ is listed as /55/.

To implement the three different types of listedness in the grammar, we define three classes of USELISTED constraints as in (7). The constraints in (7a) require that real words with existing reduplications use the listed reduplicative forms; the constraints in (7b) require an existing syllable to use its listed allomorph in nonfinal positions; and the constraints in (7c) require appropriate tonal allomorphs to be used for existing tones in nonfinal positions regardless of the segmental contents.

(7) USELISTED constraints:

a. USELISTED(σ51-σ21): Use the listed /σ51-σ21/ for /RED-σ21/.
USELISTED(σ55-σ51), USELISTED(σ33-σ55), USELISTED(σ21-σ33), USELISTED(σ33-σ24), *mutatis mutandis*.

b. USELISTED(σ21): Use the listed allomorph /σ51/ for /σ21/ non-XP-finally.
USELISTED(σ51), USELISTED(σ55), USELISTED(σ33), USELISTED(σ24), *mutatis mutandis*.

c. USELISTED(21): Use the listed tonal allomorph /51/ for /21/ non-XP-finally.
USELISTED(51), USELISTED(55), USELISTED(33), USELISTED(24), *mutatis mutandis*.

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 (8). The ranking in (8a) is based on the token frequencies of tonal melodies in single reduplication in (3b), and the rankings in (8b) and (8c) are based on the token frequencies of base tones in (3a).

(8) A priori rankings:

a. USELISTED(σ33-σ55) » USELISTED(σ55-σ51) » USELISTED(σ33-σ24) » USELISTED(σ51-σ21) » USELISTED(σ21-σ33)

b. USELISTED(σ55) » USELISTED(σ24) » USELISTED(σ33) » USELISTED(σ51) » USELISTED(σ21)

c. USELISTED(55) » USELISTED(24) » USELISTED(33) » USELISTED(51) » USELISTED(21)

### 4.2. Durationally Based Tonal Markedness Constraints

To capture the effect of duration on the phonotactics of tones on nonfinal syllables, we propose the tonal markedness constraints in (9a) and their a priori ranking in (9b) that penalizes a longer tone more severely than a shorter tone on nonfinal syllables.

(9) Tonal markedness constraints:

a. *55-NONFINAL (*55-NF): Tone 55 cannot occur on non-XP-final syllables.
*33-NF, *24-NF, *21-NF, *51-NF, *mutatis mutandis*.

b. *55-NF, *33-NF, *24-NF » *21-NF, *51-NF

### 4.3. Constraints Governing Floating High Docking

Finally, the behavior of the floating High can be captured by properly ranking the constraints in (10). We crucially recognize an RR-correspondence constraint (10c) (Urbanczyk 2001, Lin 2004), which needs to outrank the constraint that requires the left docking of the floating High (10b).

(10) Constraints governing floating High docking:

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

b. ALIGN(Float, Left, Word, Left) (ALIGN-L): The left edge of a floating tone must be aligned with the left edge of a word.

c. IDENT-RR(Tone, Left): The left edges of the tones of two reduplicants derived from the same base must be identical.

### 4.4. Constraint Ranking in a Stochastic OT Grammar

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 and 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 nonfinal syllables and assumed that the tone on the last syllable remained unchanged. Second, we only attempted to model the correct responses and non-applications of the sandhis in both single and double reduplications. Third, we attempted to model an ideal situation for AO words, in which all responses had the correct application of sandhis. Although the speakers did not perform perfectly for these words in the experiment, we assumed that they were able to use the reduplications correctly in real life, and their mistakes were due to the difficulty posed by the experimental setting.

With these provisos, the Graduate Learning Algorithm in *OTSoft* was used to find a grammar, shown in (11), that best matched the wug test result. This grammar has the following properties in its constraint ranking. First, the USELISTED constraints for existing reduplications are ranked high, which accounts for the relative stable behavior of AO words. Second, the USELISTED constraints for the allomorphs of existing syllables are ranked lower, but they generally outrank the corresponding USELISTED constraints for tonal allomorphs. This accounts for the higher sandhi productivity in *AO words than AG words. Third, among the constraints that govern floating High docking, REALIZE(Float) and IDENT-RR(Tone, Left) are highly ranked, while ALIGN-L is lowly ranked, which accounts for the productive application of floating High docking. Fourth, among the tonal markedness constraints, *24-NF is the most highly ranked, as it is the only constraint that reflects a true phonotactic generalization. This accounts for the higher productivity of the 24 → 33 sandhi. *21-NF and *51-NF are ranked the lowest among tonal markedness, which on the one hand discourages base 21 and 51 to undergo sandhi in nonfinal positions, on the other hand encourages 21 and 51 to appear as the sandhi tones. Finally, the relative high ranking of USELISTED(σ55) 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.

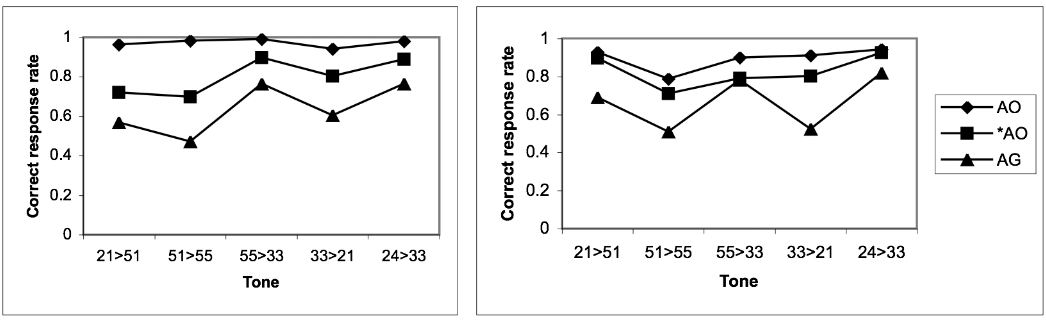
The tonal patterns generated by the stochastic grammar in (11), estimated by running the grammar 2000 times on each input, are juxtaposed with the wug test results in (12)-(14). The mismatch between this grammar's predictions and the data pattern that it set out to model is 2.8%.

(11) Ranking values for constraints in a stochastic OT grammar:

| Ranking value | Constraint | Ranking value | Constraint |
|---|---|---|---|
| 121.117 | USELISTED(σ33-σ55) | 115.814 | USELISTED(σ51) |
| 120.617 | USELISTED(σ55-σ51) | 115.778 | USELISTED(33) |
| 120.487 | REALIZE(Float) | 115.314 | USELISTED(σ21) |
| 120.117 | USELISTED(σ33-σ24) | 115.278 | USELISTED(51) |
| 119.604 | USELISTED(σ51-σ21) | 114.819 | IDENT-BR(Tone) |
| 118.290 | USELISTED(σ21-σ33) | 114.778 | USELISTED(21) |
| 117.637 | IDENT-RR(Tone, Left) | 113.304 | *55-NONFINAL |
| 117.317 | USELISTED(σ55) | 113.302 | *33-NONFINAL |
| 117.313 | USELISTED(55) | 112.802 | *21-NONFINAL |
| 116.814 | USELISTED(σ24) | 110.841 | ALIGN-L |
| 116.798 | USELISTED(24) | 89.463 | IDENT-RR(Tone) |
| 116.314 | USELISTED(σ33) | 80.880 | *51-NONFINAL |
| 116.210 | *24-NONFINAL | | |

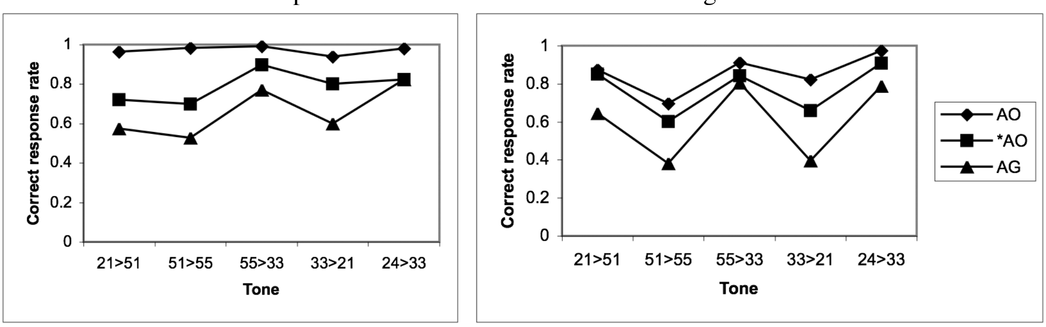
(12) Correct response rates for sandhi tones in single reduplication ($\sigma_1$):

Grammar output: Wug test result:



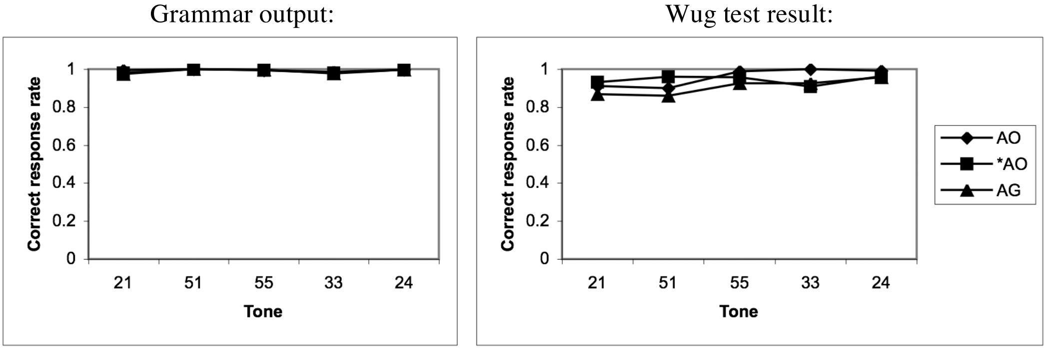

(13) Correct response rates for sandhi tones in double reduplication ($\sigma_2$):

Grammar output: Wug test result:




(14) Correct response rates for $\sigma_1$ tone, given $\sigma_2$ tone as sandhi tone:

Grammar output: Wug test result:



## 5. Remaining Issues

A number of issues remain outstanding in our model, the most pressing of which is the one on learning. The model requires the USELISTED constraints for allomorphs, especially the ones for abstract tonal allomorphs, to be ranked relatively low; otherwise the tonal allomorphs will apply entirely productively to novel words. But these constraints are never violated in real Taiwanese and will inevitably be promoted to high ranking during learning given the current implementations of constraint demotion. We tentatively suggest 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.

Abstract allomorph listing also potentially poses a duplication problem for the theory of phonology: there are now two different mechanisms through which allomorphy can be derived — the M » F ranking and the high ranking of USELISTED constraints for abstract allomorphs. Our allomorph listing 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 in the lexicon. Transparent allomorphy, then, still derives from the M » F ranking, with the role of USELISTED 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.

## 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 opaque tone circle is only variably productive in novel words, indicating that the speakers have underlearned this exceptionless pattern in the lexicon. The productivity of the opaque sandhis, however, also shows signs of both overlearning and proper learning from the lexicon: speakers prefer shorter tones as sandhi tones on nonfinal syllables, which could not have been deduced from lexical statistics; but 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 in novel words.

With the stochastic OT of Boersma and Hayes (2001) as the backdrop, we have shown that the dual listing/generation theory of Zuraw (2000) 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 speakers' phonological knowledge.

## References


Albright, Adam and Bruce Hayes (2003). Rules vs. analogy in English past tense: a computational/experimental study. *Cognition* 90: 119-161.

Berent, Iris, Donca Steriade, Tracy Lennertz, and Vered Vaknin (2007). What we know about what we have never heard: evidence from perceptual illusions. *Cognition* 104: 591-630.

Berko, Jean (1958). The child's learning of English morphology. *Word* 14: 150-177.

Boersma, Paul and Bruce Hayes (2001). Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32: 45-86.

Davidson, Lisa, Paul Smolensky, and Peter Jusczyk (2004). The initial and final states — theoretical implications and experimental explorations of Richness of the Base. In Rene Kager, Joe Pater, and Wim Zonneveld (eds.), *Fixing priorities: constraints in phonological acquisition*. Cambridge University Press, Cambridge, UK. 321-368.

Frisch, Stefan A., Janet B. Pierrehumbert, and Michael B. Broe (2004). Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* 22: 179-228.

Frisch, Stefan A. and Bushra Zawaydeh (2001). The psychological reality of OCP-place in Arabic. *Language* 77: 91-106.

Hayes, Bruce (2000). Gradient wellformedness in Optimality Theory. In Joost Dekkers, Frank van der Leeuw, and Jeroen van de Weijer (eds.), *Optimality Theory: phonology, syntax, and acquisition*. Oxford University Press, Oxford. 88-120.

Hayes, Bruce and Zsuzsa Cziráky Londe (2006). Stochastic phonological knowledge: the case of Hungarian vowel harmony. *Phonology* 23: 59-104.

Hsieh, Hsin-I (1970). The psychological reality of tone sandhi rules in Taiwanese. In *Papers from the 6th Meeting of the Chicago Linguistic Society*. 489-503.

Hsieh, Hsin-I (1975). How generative is phonology. In E. F. Koerner (ed.), *The transformational-generative paradigm and modern linguistic theory*. John Benjamins, Amsterdam, The Netherlands. 109-144.

Hsieh, Hsin-I (1976). On the unreality of some phonological rules. *Lingua* 38: 1-19.

Labov, William (1972). *Sociolinguistic patterns*. University of Pennsylvania Press, Philadelphia.

Labov, William (1994). *Principles of linguistic change*. Blackwell, Oxford, UK and Cambridge, USA.

Lin, Chienjer (2004). *Beyond output correspondence: tone sandhi in Southern Min reduplication*. Talk presented at the 78th annual meeting of the Linguistic Society of America, Boston.

Lin, Hwei-Bing (1988). *Contextual stability of Taiwanese tones*. Ph.D. dissertation, University of Connecticut.

Lubowicz, Ania (2003). *Contrast preservation in phonological mappings*. Ph.D. dissertation, University of Massachusetts, Amherst.

McCarthy, John (1999). Sympathy and phonological opacity. *Phonology* 16: 331-399.

McCarthy, John (2007). *Hidden generalizations: phonological opacity in Optimality Theory*. Equinox, London.

Moreton, Elliot (2004). Non-computable functions in Optimality Theory. In: John J. McCarthy (ed.), *Optimality Theory in phonology*. Blackwell Publishing, Malden, MA. 141-164.

Peng, Shu-hui (1997). Production and perception of Taiwanese tones in different tonal and prosodic contexts. *Journal of Phonetics* 25: 371-400.

Pierrehumbert, Janet B. (2006). The statistical basis of an unnatural alternation. In Louis Goldstein, D. H. Whalen, and Catherine T. Best (eds.), *Laboratory phonology VIII, varieties of phonological competence*. Mouton de Gruyter, Berlin. 81-107.

Sanders, Nathan (2001). Preserving synchronic parallelism: diachrony and opacity in Polish. In *Papers from the 37th Meeting of the Chicago Linguistic Society*. 501-515.

Tsay, Jane and James Myers (2005). *Taiwanese spoken corpus*. National Chung Cheng University, Taiwan.

Urbanczyk, Suzanne (2001). *Patterns of reduplication in Lushootseed*. Garland, New York.

Wang, H. Samuel (1993). Taiyu biandiao de xinli texing (On the psychological status of Taiwanese tone sandhi). *Tsinghua Xuebao* (*Tsinghua Journal of Chinese Studies*) 23: 175-192.

Wightman, Collin W., Stefanie Shattuck-Hufnagel, Mari Ostendorf, and Patti J. Price (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 91: 1707-1717.

Wu, Tzuling (1996). *A morphological analysis of reduplication in Taiwan Southern Min*. MA thesis, National Tsinghua University, Taiwan.

Yip, Moira (1990). *The tonal phonology of Chinese*. Garland, New York.

Zhang, Jie, Yuwen Lai, and Craig Turnbull-Sailor (2006). Wug-testing the "tone circle" in Taiwanese. In *Proceedings of the 25th West Coast Conference on Formal Linguistics*. Cascadilla Proceedings Project, Somerville, MA. 453-461.

Zhang, Jie, Yuwen Lai , and Craig Turnbull-Sailor (2007). *Effects of phonetics and frequency on the productivity of Taiwanese tone sandhi*. Paper presented at the 43rd Meeting of the Chicago Linguistic Society, Chicago.

Zimmer, Karl E. (1969). Psychological correlates of some Turkish morpheme structure conditions. *Language* 45: 309-321.

Zuraw, Kie Ross (2000). *Patterned exceptions in phonology*. Ph.D. dissertation, UCLA.

Zuraw, Kie (2007). The role of phonetic knowledge in phonological patterning: corpus and survey evidence from Tagalog infixation. *Language* 83: 277-316.

# Proceedings of the 27th West Coast Conference on Formal Linguistics

## edited by Natasha Abner and Jason Bishop

Cascadilla Proceedings Project Somerville, MA 2008

### Copyright information



ISBN 978-1-57473-428-7 library binding



### Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, e-mail: sales@cascadilla.com

### Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Zhang, Jie and Yuwen Lai. 2008. Phonological Knowledge beyond the Lexicon in Taiwanese Double Reduplication. In *Proceedings of the 27th West Coast Conference on Formal Linguistics*, ed. Natasha Abner and Jason Bishop, 458-466. Somerville, MA: Cascadilla Proceedings Project.

or:

Zhang, Jie and Yuwen Lai. 2008. Phonological Knowledge beyond the Lexicon in Taiwanese Double Reduplication. In *Proceedings of the 27th West Coast Conference on Formal Linguistics*, ed. Natasha Abner and Jason Bishop, 458-466. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #1862.