

# Morphologically-conditioned tonotactics in multilevel Maximum Entropy grammar

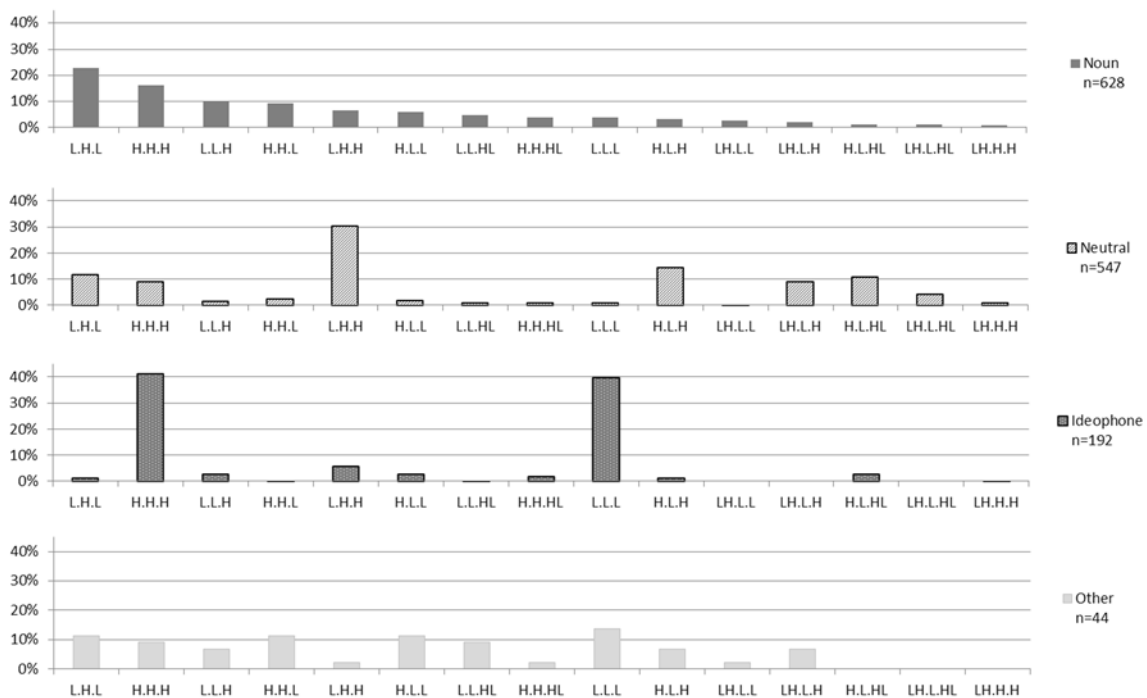
Stephanie S Shih<sup>1</sup> and Sharon Inkelas<sup>2</sup>

<sup>1</sup>University of California, Merced and <sup>2</sup>University of California, Berkeley

## 1 Introduction

This paper presents a case of lexically-conditioned tonotactic variation by part of speech in the Mende lexicon (Mande, Sierra Leone; data from Innes 1969). As illustrated by the plots in (1) of trisyllabic surface tone patterns, different parts of speech in Mende vary markedly in the tonal patterns they tend to exhibit:

(1) Trisyllabic surface tone patterns in Mende (type % out of all trisyllabic tone patterns)



Phonological sensitivity to lexical class has been addressed in the literature; at least two sets of hypotheses have been advanced regarding its codification in grammar. The first type of approach holds that lexical class differences are limited by restrictions in the design of grammar itself (or more specifically, Universal Grammar): for example, it has been hypothesized that only faithfulness, but not markedness, constraints can be ranked in a lexically class-specific way, or can be class-specific (e.g., Itô & Mester 1995a, 1999, 2009; Alderete 2001; Smith 2011); or that there certain preferential classes (e.g., nouns) demonstrate more contrasting phonological patterns than do other classes (e.g. verbs, with adjectives somewhere in between; Smith 2011). The second type of approach to lexical class specific phonology allows each lexical class to have its own completely independent phonological profile, without *a priori* restrictions or specifications on the structure of the grammar (Anttila 2002; Inkelas & Zoll 2007; Pater 2009; Becker & Gouskova, to appear; see also Itô & Mester 1995b).

\* For discussion on various portions of this work, we thank Laura MacPherson, Brian Smith, Alan Yu, Kie Zuraw, and audiences at UC Berkeley, UC San Diego, CLS 2015, and AMP 2015.

The case study of Mende we work up in this paper supports the latter approach. Lexical class-sensitive phonological differences in the Mende lexicon go beyond faithfulness and noun-adjective-verb privilege asymmetries; they more closely resemble the complexity that has been observed for morphophonological alternations (Inkelas 2014: ch. 2). Rather than trying a priori to constrain language-internal morphophonological variation, we focus instead on quantitatively modeling the space of lexically-conditioned variation, with the Mende lexicon as our case study. Using *varying slopes* in Maximum Entropy Harmonic Grammar (MaxEnt; Goldwater & Johnson 2003; Wilson 2006; Jäger 2007; Hayes & Wilson 2008), this approach gains traction on the larger, overarching problem in morphophonology of how to quantify the heterogeneity that is often the consequence of morphological conditioning in a phonological system.


The paper is organized as follows. Section 2 introduces Mende tonotactics and provides Optimality-theoretic constraints for capturing the observed surface tone patterns in the lexicon. The mechanics of the Maximum Entropy analysis are described in §3. Results of the MaxEnt grammar and the use of varying slopes to model lexical class-conditioned tonotactics are presented in §4. Section 5 discusses broader implications of the lexical-class sensitive grammatical model.

## 2 Mende tonotactics


**2.1 Data** Our data for this paper comes from a Mende dictionary ( $n=5,412$ ; Innes 1969). Morpheme breaks are not indicated in the Innes 1969 dictionary. Because a primary source of morphological complexity (in nouns, at least) is total reduplication that results in 4-syllable words, we limit our analysis to 1 to 3-syllable long words ( $n=4,989$ ). This helps to limit potential confounds that may arise from morphological complexity.

Following the broad outlines of Innes’s grammatical analysis, we sort the words we examine into four part-of-speech categories: nouns ( $n=2,494$ ), neutrals (i.e. verbs and adjectives) ( $n=1,442$ ), ideophones ( $n=762$ ), and a heterogeneous, grab bag category “other” that consists of all other parts-of-speech marked in Innes ( $n=291$ ), including interjections, adverbs, and a variety of function words.

**2.2 Early generative accounts** Early generative accounts of Mende tone focused on common, recurrent tone patterns among nouns (e.g., Leben 1978). In Autosegmental Phonology (AP; Goldsmith 1976; Leben 1973), these tone patterns were codified in the form of a small set of licensed *autosegmental tone melodies*, i.e. H, L, HL, LH, LHL. Underlyingly floating, the tones in these melodies were mapped to tone-bearing units (vowels or syllables) via universal principles of left-to-right, one-to-one mapping, rightward spreading, and leftover tone ‘docking’. These mapping principles, subsequently applied with success in other languages as well (e.g., Pulleyblank 1986; Hyman 1987) were intended to derive the observed surface tone melodies. The autosegmental analysis of a trisyllabic Mende noun with a LH tone melody is illustrated in (2):

(2) ndàvùlá ‘sling’  


Subsequent work noted, however, that many surface patterns in Mende deviate from the five underlying melodies and/or from the autosegmental association principles (Dwyer 1978; Conteh et al. 1983; Zoll 2003; Zhang 2007). Two problematic cases in (3) demonstrate these deviations.

(3) a. lèlémá ‘praying mantis’      b. gbágběmà ‘sensitive plant’  


Example (3a) violates the left-to-right, one-to-one association principle, which predicts \*lèlémá (L.H.H) rather than the attested lèlémá (L.L.H). Example (3b) not only violates association principles by having a word-medial contour (left-to-right mapping plus leftover tone docking predicts \*gbágběmà (H.L.HL)), but also exhibits a surface melodic pattern (HLHL) that is not present in the posited underlying inventory of tone melodies. Leben (1978) proposed lexical tone linking or alternative association principles to handle exceptions such as those in (3); Dwyer (1978) and Conteh et al. (1983) argued, however, that prelinking could just as well be used for all nouns in the system, obviating the autosegmental analysis.

**2.3 *A fresh look at tone: the theoretical underpinnings*** The autosegmental account of Mende targeted the most common patterns in the system by giving the grammar a role in producing them, consigning the less common patterns to lexical prespecification. This approach thus tacitly bifurcated surface tone patterns into the frequent and the infrequent, choosing only to account for the former.

When we examine the full set of data, however, we observe a cline of pattern frequency, rather than a categorical difference. This cline is problematic both for the autosegmental account (where to bifurcate?) and for the prespecification account (why, if all patterns are possible, are some more common than others?). We stake out a different strategy. Instead of rejecting a grammatical account of Mende tone in light of the exceptions to the autosegmental analysis, we instead propose a grammatical analysis whose stated goal is to characterize the more frequent melodies in terms of ‘harmony’, or grammatical goodness.

In part because it has different goals, our analysis of Mende tone patterns departs from some basic assumptions of Autosegmental Phonology. It is not operation-based, thus does not map underlying tone melodies to surface vowels. Rather, our analysis is constraint-based approach, relying on tonotactic constraints that evaluate surface tone patterns.

The constraints are formulated within a theory of surface correspondence and minimal unit representation, or Agreement by Correspondence + Q Theory (ABC+Q). Agreement by Correspondence Theory (Hansson 2001, 2010; Rose & Walker 2004; Bennett 2013; a.o.) is grounded in basic principles of similarity and proximity attraction; it models instability in syntagmatic phonological relationships (Wayment 2009; Inkelas & Shih 2014). The basic insight of ABC is that elements that are sufficiently similar and/or proximal will interact in processes such as assimilation and dissimilation. Our view is that tone patterns also follow these basic principles. For details of the ABC approach to tone for Mende, see Inkelas & Shih forthcoming. For simplicity in readability for this paper, however, the underlying ABC-style constraints have been reformulated into a more familiar phonotactic markedness format.

The representational portion of the analysis is based in Q Theory (e.g., Shih & Inkelas 2014; Shih & Inkelas 2015), which decomposes segments into strings of 2 or 3 smaller, temporally-sequenced, subsegments, each internally featurally uniform, all of which bear tone features. Q theory is designed to handle ‘contour segments’ which contain feature sequences; contour tone is an obvious application. This decomposition is schematized in (4a), where a segment  $Q$  is made up of a sequence of two subsegments  $q$ .

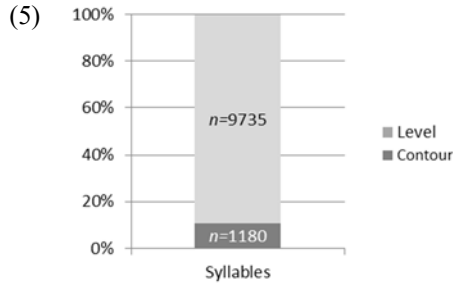
$$(4) \quad a. \quad Q \rightarrow (q^1 \ q^2) \quad b. \quad \check{a} \rightarrow (\grave{a} \acute{a}) \quad \widehat{LH} \rightarrow (L \ H) \quad c. \quad \grave{a} \rightarrow (\grave{a} \grave{a}) \quad L \rightarrow (L \ L)$$

Under Q Theory, a LH contour (e.g., (4b)) is represented as a single vowel (‘Q’) composed as a sequence of two subsegment ‘q’s, the first of which is L-toned and the second of which H-toned. A level-toned syllable, such as the all-L vowel in (4c), is represented as a sequence of two subsegments which have identical tone (in this case, L). Q theory provides a fine-grained point of reference for the grammar, thus logically separating the issue of what the minimal units that carry tone features are (‘q’) from the question of what grammatical units tonotactic constraints can refer to (‘q’, ‘Q’, the mora, the syllable, etc., as appropriate). The conflation of these issues has long created confusion in the theoretical analysis of tone (see, e.g., the discussion in Yip 2002:73ff.).

The key relevant difference, for purposes of this paper, between AP and ABC+Q is that our analysis does not employ floating tones, autosegmental ‘association line’ representations, or operations such as tone association. Rather, it constrains surface tones and tone sequences.

**2.4 *Observed patterns for Mende surface tones*** Surface tone patterns across the Mende lexicon exhibit five major tendencies that are encapsulated with the constraints posited in our analysis: avoidance of contour tones in general, a tendency for contour tones to be final, a tendency for tones to change at syllable boundaries, the avoidance of HLH ‘trough’, and a general preference for H tone. These observations are extended from Inkelas & Shih, forthcoming, and are codified below in the form of constraints.

**2.4.1 *Tone change and contour avoidance*** The tendency to avoid contour tones (HL, LH) is illustrated by the plot in (5): in Mende, level-toned syllables (HH, LL) are significantly more common than contour-toned syllables (HL, LH) ( $\chi^2=6705.270, p<0.0001$ ). We model contour tone avoidance using a phonotactic constraint \*CHANGE that penalises sequences of tonally non-identical adjacent subsegments, defined in (6).



(6) \* $[\alpha T]::[\beta T]$  (\*CHANGE) Penalize every sequence of adjacent *q*'s that are tonally non-identical.

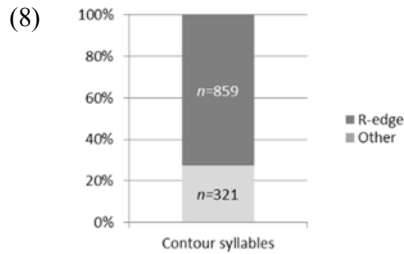
(7)

$\sigma.\sigma$	freq	*CHANGE * $[\alpha T]::[\beta T]$
☞ HH.HH	995	
a. LL.HH	701	W1
b. LL.HL	389	W2

The tableau in (7) demonstrates how the number of penalties assessed by \*CHANGE for changes between tones, as in candidates (b) (one change) and (c) (two changes), correlates inversely with the relative frequency of the tone pattern in disyllabic words in the Mende lexicon.

We have formulated \*CHANGE in the most general possible manner; as will be seen below, it competes with other more specific constraints, e.g. the preference (in certain lexical classes) for words to have all-level melodies, and the preference for syllable boundaries to align with tone changes (§2.4.2).

**2.4.2 Contour tone alignment** Though contours are dispreferred, when they do occur, contour tones are preferentially tolerated at the right edge. This observation has been noted previously for Mende (e.g., Conteh et al. 1983; Zoll 2003) as well as cross-linguistically (e.g., Gordon 2001; Zhang 2007). The plot in (8) shows that a significant majority—more than 70%—of contour-toned syllables (HL, LH) in Mende occur at the right edge of a word ( $\chi^2=353.407, p<0.0001$ ). We formalise contour tone alignment using the constraint in (9); \*WeakCONTOUR penalises contour tones—i.e., non-agreeing subsegments—in non-final (weak) syllables (cf. COINCIDE-CONTOUR; Zoll 2003).



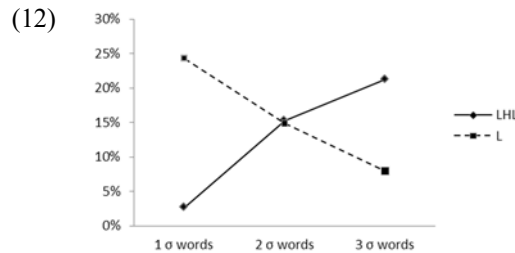
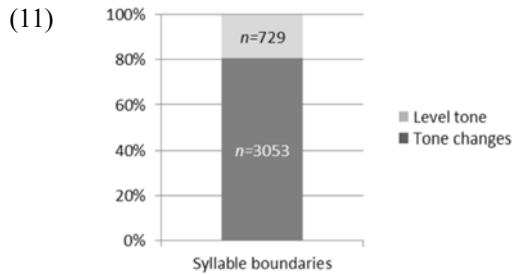
(9) \* $([\alpha T]::[\beta T])_{\sigma_w}$  (\*WeakCONTOUR) Penalise every adjacent *q* sequence of non-identical tones within a weak (i.e., non-final) syllable.

(10)

$\sigma.\sigma$	freq	*WeakCONTOUR	*CHANGE
☞ a. LL.HL	389		2
b. LH.LL	77	W1	2

Tableau (10) illustrates how \*WeakCONTOUR works. Candidate (b), with tonally non-agreeing subsegments (LH) in a non-final syllable, violates \*WeakCONTOUR and \*CHANGE. In contrast, candidate (a), with a final contour, incurs the same number of \*CHANGE violations but does not violate \*WeakCONTOUR.

**2.4.3 Tone change and syllable boundary alignment** A notable exception to the avoidance of tone changes in Mende occurs at syllable boundaries. As illustrated in (11) for polysyllabic words with non-level surface patterns, syllable boundaries coincide significantly with tone changes ( $\chi^2=1428.074, p<0.0001$ ).



The tendency for tones to change at syllable boundaries entails that surface tone melody complexity scales with increasing word length, illustrated by the plot in (12). As the number of syllables in a word increases

from one to three syllables, so does the attestation of complex tone melodies such as LHL. By contrast, the attestation of all-level melodies such as L decreases as the number of syllables increases.

We formalise the preference for tones to change at syllable boundaries with the constraint *CHANGE@\\$* (13), which penalises tonally-identical sequences separated by a syllable boundary. *CHANGE@\\$* is based on *xx-EDGE* constraints, proposed by Bennett 2013, which require that surface correspondence domains align with phonological or prosodic *EDGE* domain boundaries (see also *CRISPEGE*; Itô & Mester 1994, 1999b).

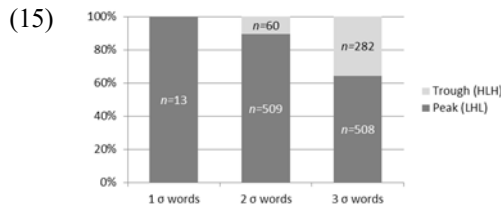
- (13) \* $[\alpha T]:\$:[\alpha T]$  Penalise every sequence of adjacent, tonally-identical *q*'s that crosses a  
(*CHANGE@\\$*) syllable boundary (*\\$*).

(14)

$\sigma.\sigma.\sigma$	<i>freq</i>	<i>CHANGE@\\$</i>	* <i>WeakCONTOUR</i>	* <i>CHANGE</i>
☞ a. LL.HH.LL	233			2
b. LL.HH.HH	218	W1		L1
c. LL.LL.LL	112	W2		L
d. LL.LL.HH	78	W1		L1
e. LH.LL.HH	64		W1	W3
f. LH.HH.LL	7	W1	W1	2

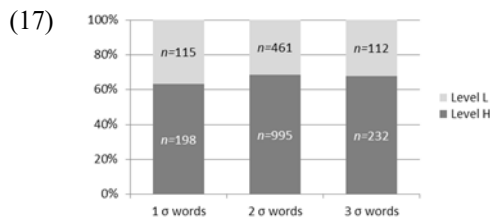
The tableau in (14) illustrates the effects of *CHANGE@\\$*. Here, the optimal candidate (a) is one where tones change at syllable boundaries but nowhere else. Level tones at syllable boundaries violate *CHANGE@\\$*, as in candidates (b), (c), (d), and (f). Changes within nonfinal syllables violate \**WeakCONTOUR*.

**2.4.4 Trough avoidance, H tone requirement** Like many languages, Mende avoids HLH sequences, or ‘troughs’ (Yip 2002; Cahill 2007; Hyman 2011) in surface melodies. These troughs do occur, especially as the length of a word increases, but they are still vastly underrepresented in comparison to LHL surface patterns, as shown in (15). Our version of \**TROUGH* penalises two H tones that are separated by any number of consecutive L tones (16).



- (16) \**TROUGH* Penalise any two H tones separated by any number of consecutive L tones.

The second general pattern that is observed in the Mende lexicon is the preference that words have at least one H tone. The chart in (17) illustrates this phenomenon, showing that level H tone patterns are much more common than level L tone patterns, across words of any syllable length. *HAVEH*, defined in (18), captures this preference in the grammar.



- (18) *HAVEH* Penalise any word with no H tone.

For lack of space we will not provide illustrative tableaux, but these two constraints function in the same manner as the preceding constraints discussed in this section: the relative number of constraint violations is inversely correlated with relative frequency of different surface tone patterns.

**2.5 A note on constraints** The goal in this paper is to examine and model the variance of part-of-speech-specific tonotactics. We have introduced a number of constraints to this end; however, we wish to

point out that any number of different constraint sets would doubtless work equally well for this purpose. Our focus is on the fact that there are lexically-conditioned differences in the relative frequency of tone patterns, (see e.g., (1)) rather than on the specifics of the general constraint set.

### 3 A varying slopes analysis of lexically-conditioned tonotactics patterns

We analyze the lexically-conditioned tonotactics patterns in Mende using Maximum Entropy Harmonic Grammar (MaxEnt; Goldwater & Johnson 2003; Wilson 2006; Jäger 2007; Hayes & Wilson 2008; a.o.), fitting models with the MaxEnt Grammar Tool (Hayes et al. 2009). MaxEnt grammars account for comparative grammaticality by ranking probabilities of outcome candidates in variable data. In our analysis, the output of the grammar is a probability distribution over all possible surface tone patterns, i.e. all possible combinations of level LL, level HH, and contour LH, HL syllables in 1-3 syllable words.<sup>1</sup>

As discussed earlier, morphologically-conditioned phonology has been modeled via a single grammar approach using indexed constraints (e.g., Itô & Mester 1995a; Alderete 2001) or a multiple grammar approach using cophonologies (e.g., Anttila 2002; Inkelas & Zoll 2005). As Inkelas & Zoll (2007) have argued, these approaches are essentially identical when morphological interleaving is not in play (as it is not in our analysis of static tone patterns in the Mende lexicon). These approaches can easily be translated into the model we use here, namely a multilevel, *varying slopes* analysis in which additive weight adjustments—i.e., ‘varying slopes’—are added for each constraint in the Base Grammar, per each lexical class (see also Albright 2008; Coetzee & Pater 2011; Moore-Cantwell & Pater, submitted). The varying slopes approach is schematized in (19):

$$(19) \quad \begin{array}{l} w_1 C_1 + w_2(C_1 \times \text{NOUN}) + w_3(C_1 \times \text{NEUT}) + w_4(C_1 \times \text{ID}) + \dots + w_k C_{iN} \\ \text{Base grammar} \qquad \qquad \qquad = \quad w_1 C_1 \\ \text{Adjusted noun grammar} \qquad \qquad = \quad w_1 C_1 + w_2(C_1 \times \text{NOUN}) \\ \text{Adjusted neutral grammar} \qquad \qquad = \quad w_1 C_1 + w_3(C_1 \times \text{NEUT}) \\ \text{Adjusted ideophone grammar} \qquad \qquad = \quad w_1 C_1 + w_4(C_1 \times \text{ID}) \end{array}$$

Under a varying slopes analysis, each constraint  $C$  in the Base Grammar receives a weight  $w$ . The Base Grammar predicts the overall tonotactics of the language. Lexical class-specific tonotactics are predicted by an additive combination of Base Grammar weights and lexical class-specific weights for the same constraints (e.g., for nouns:  $w_1 C_1 + w_2(C_1 \times \text{Noun})$ ). Because we are dealing with a finite set of lexical classes in Mende, varying slopes are formally executed here as interaction terms.<sup>2,3</sup>

**3.1 What is the base grammar?** The base grammar was trained on bootstrapped samples of possible tone melodies in Mende, drawn from the existing distribution over the lexicon. Each random sample was 2500 words. A representative result is reported here. The intention of using bootstrapped sampling to produce data for a base grammar was to mirror a Mende speaker knowing that a novel word has a tone pattern that is licit in Mende but not yet knowing the part of speech. Furthermore, we did not want to *a priori* assume that any one lexical class had privileged status (i.e., nouns, under e.g., Smith 2011) by using a particular lexical class as the base grammar.

For comparison, we also trained a separate grammar for another baseline on totally random tone patterns ( $n=5000$ ), generated from random combinations of level tone (HH, LL) and contour tone (LH, HL) syllables, for 1-to-3 syllable long words. The purpose of this random baseline was to test the difference between

<sup>1</sup> Mende does exhibit more tonally complex syllables. LHL contours occur primarily in monosyllables; when found in longer words, they are restricted to long vowels: e.g., *mēēndé*, ‘member of Humoi society’. The only potential instances of HLH-toned syllables in our corpus come from words with non-identical vowel-vowel sequences, which we counted as separate syllables: e.g., *wúā*, ‘Hello’. Ternary tone contours are expected in Q theory and thus our analysis could handle them. However, their rarity in Mende has allowed us to take the liberty of simplifying our representations (using 2  $q$ 's per syllable instead of 3) and analysis by excluding them from this paper.

<sup>2</sup> On the equivalency of interaction terms and random slopes in this situation, see e.g., Gelman & Hill 2007.

<sup>3</sup> In the MaxEnt grammar literature, interaction terms usually model weighted constraint conjunctions (e.g., Hayes et al. 2012; Pater & Moreton 2012; Green & Davis 2014; Shih 2016). In our Mende analysis, we are not treating the lexical classes themselves as constraints in the grammar, so are not using constraint conjunction (though this is certainly not outside the realm of possibility: see e.g., Pater et al. 2012).

language-specific tone patterns and completely random tone patterns, under the assumption that language-specific patterns will be governed by grammatical constraints.

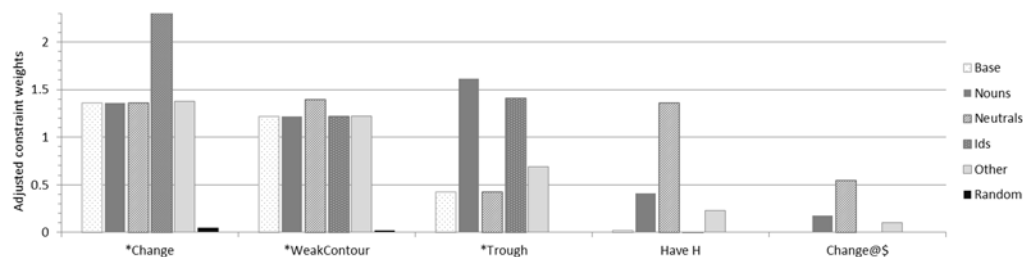
## 4 Results

The adjusted MaxEnt constraint weightings are given in (20) and represented graphically in (21). The table in (22) presents the rank orders of the constraints in each cophonology.

(20) Adjusted MaxEnt weights, in numbers

	Base	Adjusted Grammars				Random
		<i>Nouns</i>	<i>Neutrals</i>	<i>Ideophones</i>	<i>Other</i>	
*CHANGE	1.356	1.356	1.356	2.537	1.374	0.047
*WeakCONTOUR	1.220	1.220	1.391	1.220	1.220	0.018
*TROUGH	0.426	1.612	0.426	1.409	0.691	0
HAVE H	0.017	0.411	1.361	0.017	0.232	0
CHANGE@\$	0	0.175	0.543	0	0.099	0

(21) Adjusted MaxEnt weights



(22) Constraint weights by lexical class (top→bottom: highest weight→lowest weight)

Base	Cophonology Indexed Grammars			
	<i>Nouns</i>	<i>Neutrals</i>	<i>Ideophones</i>	<i>Other</i>
*CHANGE	*TROUGH	*WeakCONTOUR	*CHANGE	*CHANGE
*WeakCONTOUR	*CHANGE	HAVE H	*TROUGH	*WeakCONTOUR
*TROUGH	*WeakCONTOUR	*CHANGE	*WeakCONTOUR	*TROUGH
HAVE H	HAVE H	CHANGE@\$	CHANGE@\$	HAVE H
CHANGE@\$	CHANGE@\$	*TROUGH	HAVE H	CHANGE@\$

**4.1 Random baseline and Base Grammar results** As example (20) demonstrates, a MaxEnt grammar that was trained on completely random data yields constraint weights that are at or very near 0. This result shows that there is random tone combinations have no grammatical patterns, and provides a comparison point for constraints for the Mende grammar.

The mostly highly weighted constraints in the Base Grammar, as seen in (20)-(22), are \*CHANGE,  $w=1.356$ ), whose main effect is to limit contour tones, and \*WeakCONTOUR,  $w=1.22$ , which limits contour tones to the right edge. In the completely random model, the weights for these constraints are near 0. Both \*CHANGE and \*WeakCONTOUR are relatively stable and highly weighted across all of the lexical classes, consistent with the claim that avoiding contour tones and especially avoiding contour tones in non-final syllables are general, universal dispreferences in tone systems (e.g., Gordon 2001; Zhang 2004).

**4.2 Lexical class-conditioned results** Turning next to constraints whose weights vary considerably across lexical class, we note that the lexical class-specific adjusted weights for CHANGE@\$ show that only

nouns and neutrals exhibit an effect of syllable edge and tone transition alignment: CHANGE@ $\$$ ×NOUN,  $w=0.175$ ; CHANGE@ $\$$ ×NEUT,  $w=0.543$ . In comparison, CHANGE@ $\$$  does not have a positive weight for ideophones ( $w=0$ ), suggesting that ideophones (e.g., *vòvòlò* ‘creaking’) often involve much simpler, level surface tone patterns than the other lexical classes. Additionally, as part of this level preference, ideophones significantly increase the weight of \*CHANGE beyond the Base Grammar and other parts of speech (\*CHANGE×ID,  $w=2.537$ ). The tableau in (23) demonstrates these ideophone-specific weighted constraints.

(23)

$\sigma.\sigma.\sigma$ , ID	<i>freq</i>	*CHANGE×ID * $[\alpha T]::[\beta T]$ ×ID $w_{adj}=2.537$	CHANGE@ $\$$ ×ID * $[\alpha T]:\$:[\alpha T]$ ×ID $w_{adj}=0$	$\mathcal{H}$
a. LL.HH.LL	2	W2		<b>5.073</b>
☞ b. LL.LL.LL	76		L2	<b>0</b>

In (23), the LL.HH.LL candidate (a) has a high harmony penalty under ideophone-specific \*CHANGE×ID, and so is far less frequent than a candidate (b) that satisfies \*CHANGE but may violate zero-weighted CHANGE@ $\$$ ×ID.

Because ideophones prefer level surface tone patterns, they are the part of speech that tolerates all-level L surface patterns the best. In contrast, neutrals, in particular, show an affinity for having a requisite H tone in a word, which is captured by the result that the greatest adjusted weight for HAVEH is for neutrals: HAVEH×NEUT  $w=1.36$  (versus HAVEH×NOUN  $w=0.41$  and HAVEH×ID  $w=0.017$ ). As such, neutrals have the smallest likelihood of being all-level L-toned words. The tableaux in (24) show the comparison between adjusted weights for HAVEH across lexical classes: with the greater weight, HAVEH×NEUT imputes the highest penalised harmony score to an all-L word.

(24)

$\sigma.\sigma.\sigma$ , NEUT	<i>freq</i>	HAVEH×NEUT $w_{adj}=1.360656$	$\mathcal{H}$
a. LL.LL.LL	5	1	<b>1.36</b>

$\sigma.\sigma.\sigma$ , NOUN		HAVEH×NOUN $w_{adj}=0.410611$	
b. LL.LL.LL	25	1	<b>0.41</b>

$\sigma.\sigma.\sigma$ , ID		HAVEH×ID $w_{adj}=0.016843$	
c. LL.LL.LL	76	1	<b>.02</b>

Neutrals also differ from nouns and ideophones by showing a greater preference for tone transitions that align with syllable boundaries, even if the pattern is a HLH trough (e.g., *hívùndí* ‘be mouldy, mildewed’). The adjusted weight for \*TROUGH for neutrals reflects this tendency: \*TROUGH×NEUT  $w=0.426$  is much lower than the adjusted \*TROUGH weights for nouns ( $w=1.612$ ) and ideophones ( $w=1.409$ ). The result of these different weights is demonstrated in the set of tableaux in (25), where a low weight for neutral-specific \*TROUGH allows a candidate (a) that is HH.LL.HH and satisfies CHANGE@ $\$$  to win. In contrast, the higher weights for noun- and ideophone-specific \*TROUGH penalise troughs more strongly than CHANGE@ $\$$  penalizes all-level words.



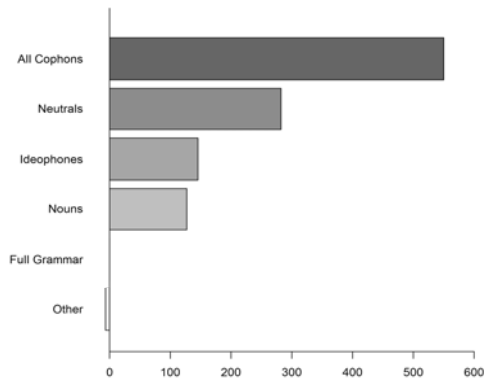
(25)

	$\sigma.\sigma.\sigma$ , NEUT	<i>freq</i>	*TROUGH×NEUT $w_{adj}=0.426$	CHANGE@\$ * $[\alpha T]:\$:[\alpha T]$ ×NEUT $w_{adj}=0.543$	$\mathcal{H}$
a.	i. HH.LL.HH	78	1		<b>0.43</b>
	ii. HH.HH.HH	48		<b>2</b>	<b>1.09</b>
	$\sigma.\sigma.\sigma$ , NOUN		*TROUGH×NOUN $w_{adj}=1.612$	* $[\alpha T]:\$:[\alpha T]$ ×NOUN $w_{adj}=0.175$	
b.	i. HH.LL.HH	21	1		<b>1.61</b>
	ii. HH.HH.HH	101		<b>2</b>	<b>0.35</b>
	$\sigma.\sigma.\sigma$ , ID		*TROUGH×ID $w_{adj}=1.409$	* $[\alpha T]:\$:[\alpha T]$ ×ID $w_{adj}=0$	
c.	i. HH.LL.HH	2	1		<b>1.41</b>
	ii. HH.HH.HH	79		<b>2</b>	<b>0.00</b>

In sum, permitting constraint weights to be adjusted for each lexical class allows a tone pattern that is preferred in one portion of the lexicon to be dispreferred in other portions of the lexicon.

**4.3 Importance of lexical conditioning** Lexically-specific constraint weighting adds complexity to the grammar. In this section we use information-theoretic model comparison to test whether this additional complexity is justified and whether having lexically-specific weights significantly improves our grammar in terms of modeling Mende surface tone patterns. The second-order Akaike Information Criterion ( $AIC_C$ ) penalises model complexity (i.e., increasing number of grammatical constraints) against the weight of evidence (for  $AIC_C$ -based model comparison, see Burnham & Anderson 2002, et seq.; for  $AIC_C$ -based model comparison for phonological grammars, see e.g., Shih 2016; cf., Wilson & Obdeyn 2009). In  $AIC_C$ -based comparison, each set of lexical-specific constraints (i.e., cophonologies) is removed in turn from the full model, and the difference in  $AIC_C$  ( $\Delta AIC_C$ ) is recorded.  $\Delta AIC_C$  values above 10 indicate that significant amounts of information are lost when that specific cophonology is removed. As seen in (26), the  $AIC_C$  comparison results show that the ‘Full Grammar’, with all three lexical class-specific cophonologies, produces a better fit for the Mende lexical data: removing all cophonologies and leaving only the base constraints, as shown by the topmost bar in (26), yields a very large  $AIC_C$  gain from the ‘Full Grammar’ bar.

(26)  $\Delta AIC_C$  comparison between the full model and models with cophonologies removed.



Interestingly, removing the cophonology for the ‘Other’ part-of-speech class produce an  $AIC_C$  dip, meaning that the grammar actually better models the Mende data without an Other-specific cophonology. This apparent exception is not surprising, since the Other lexical class is a heterogeneous grab-bag, which includes many different lexical groups: e.g., conjunctions, interjections, etc.

The Noun, Neutral and Ideophone cophonologies all contribute significantly to the grammar. The Neutral-specific cophonology contributes the most information to the full grammar—more than the ideophone and noun cophonologies do. In the case of the noun-specific cophonology, this may be because

nouns show a greater spread of tone patterns, consistent with previous cross-linguistic observations regarding noun-specific phonological behavior (Smith 2001, 2011). Interestingly, ideophones do not present in Mende as the most extreme lexical class: they contribute less information to the full grammar than neutrals and more than nouns. The patterning of ideophones suggests that they fall within the range of what Mende allows for surface tones; unlike their outlier behavior in some languages (see e.g. Voeltz & Kilian-Hatz 2001), ideophones do not wildly differ from the general tonotactics of Mende.

## 5 Implications and Conclusion

The lexical class-conditioned phonological patterns in Mende tone exhibit differences comparable in size to that observed in morphophonological alternations (e.g., Anttila 2002). Importantly, these differences involve differential markedness and seem unlikely to yield to an analysis of differential faithfulness, in which faithfulness to a specific lexical class outranks faithfulness to another lexical class (e.g., FAITH<sub>NOUN</sub> » FAITH<sub>VERB</sub>; Smith 2001, 2011). Though there is undoubtedly a role for faithfulness, we have not yet incorporated it into the current system, and the indexation of markedness constraints per lexical class used here for phonotactics already accounts for much of the data.

Mende clearly observes markedness reversals of the kind that ‘grammar dependence’ in faithfulness-based formulations of lexical class-specific phonology (Alderete 2001) were designed to rule out (though cf. Fukazawa et al. 1998; Inkelas & Zoll 2007; Pater 2009). For example, across nearly all of the base and lexical class-specific grammars in Mende, \*TROUGH is ranked fairly highly with respect to the other constraints in the system. For the class of neutrals, however, \*TROUGH is weighted at the bottom (see e.g., (22)). Thus, a structure (HLH) that is otherwise highly marked in the rest of the language is quite good and comparatively unmarked in the neutrals class.

Beyond modeling the categorical re-rankings of constraints, the varying slopes approach introduced here can quantify the degree to which lexical class-specific phonotactics differ from class to class and from the rest of the lexicon. One potential implication of this methodology is the ability to develop a flexible scale for how different ideophone phonology can be from the rest of the grammar, a long standing issue in the discussion of African languages (see Rose 2015 for a recent summary). Our results for Mende show that ideophones operate within fairly conservative parameters of the overall Mende tonotactics, especially when compared to other lexical classes such as neutrals.

One caveat of note is that because we *a priori* restrict Harmonic Grammar and Optimality-theoretic grammars to positive weights, so as to penalise rather than reward structures, the current approach does not allow us to fully examine the complete potential space of variation between cophonologies in a language. Because the varying slopes approach is couched in the domain-general statistics of Maximum Entropy models, however, it can in principle be deployed with unrestricted, unbounded constraint weights as well. Allowing the rewarding of structures via negative weights may reveal reversal effects even more pronounced than what we have seen with \*TROUGH in the neutral cophonology.

In the future, the varying slopes approach presented here has potential extensions to other types of lexically-sensitive morphology, such as affixation and dominant tone melodies, and the investigation of how statistical tonotactics in a lexicon can be grammaticalised into productive morphemes of a language.

## References

- ALBRIGHT, ADAM. 2008. How many grammars am I holding up? Discovering differences between word classes. *Proceedings of the 26th West Coast Conference on Formal Linguistics*, ed. by Charles B. Chang and Hannah J. Haynie, 1–20. Somerville, MA: Cascadilla Proceedings Project.
- ALDERETE, JOHN D. 2001. Dominance Effects as Transderivational Anti-Faithfulness. *Phonology* 18.201–253.
- ANTTILA, ARTO. 2002. Morphologically conditioned phonological alternations. *Natural Language and Linguistic Theory* 20.1–42.
- BECKER, MICHAEL.; and MARIA GOUSKOVA. to appear. Source-oriented generalizations as grammar inference in Russian vowel deletion. *Linguistic Inquiry*.
- BENNETT, WILLIAM G. 2013. Dissimilation, consonant harmony, and surface correspondence. New Brunswick, NJ: Rutgers, The State University of New Jersey Ph.D. dissertation.
- BURNHAM, KENNETH P.; and DAVID R. ANDERSON. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. 2nd edition. New York, NY: Springer-Verlag.
- CAHILL, MIKE. 2007. More universals of tone. Ms. SIL, ms.

- COETZEE, ANDRIES W.; and JOE PATER. 2011. The place of variation in phonological theory. *The Handbook of Phonological Theory*, ed. by John A Goldsmith, Jason Riggle, and Alan C. L. Yu, 401–434. John Wiley & Sons.
- CONTEH, P.; E. COWPER.; D. JAMES.; K. RICE.; and M. SZAMOSI. 1983. A reanalysis of tone in Mende. *Current Approaches to African Linguistics*, ed. by J. Kaye, H. Koopman, D. Sportiche, A. Dugas, J. Kaye, H. Koopman, D. Sportiche, and A. Dugas, 2:127–137. Dordrecht: Foris.
- DWYER, DAVID. 1978. What sort of tone language is Mende? *Studies in African Linguistics* 9.167–208.
- FUKAZAWA, HARUKA.; MAFUYU KITAHARA.; and MITSUHIKO OTA. 1998. Lexical stratification and ranking invariance in constraint-based grammars. *Proceedings of the Chicago Linguistic Society 34-2: The panels*, ed. by M. Catherine Gruber, Derrick Higgins, Kenneth S. Olson, and Tamra Wysocki. Chicago: University of Chicago. 260-0598.
- GELMAN, ANDREW.; and JENNIFER HILL. 2007. *Data analysis using regression and multilevel/hierarchical models*. Analytic methods for social research. New York, NY: Cambridge University Press.
- GOLDSMITH, JOHN. 1976. Autosegmental phonology. Cambridge, MA: Massachusetts Institute of Technology Ph.D. dissertation.
- GOLDWATER, SHARON.; and MARK JOHNSON. 2003. Learning OT Constraint Rankings Using a Maximum Entropy Model. *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, ed. by Jennifer Spenader, Anders Eriksson, and Östen Dahl, 111–120. Stockholm: Stockholm University.
- GORDON, MATTHEW. 2001. A typology of contour tone restrictions. *Studies in Language* 25.405–444.
- GREEN, CHRISTOPHER.; and STUART DAVIS. 2014. Superadditivity and Limitations on Syllable Complexity in Bambara Words. *Perspectives on Phonological Theory and Acquisition: Papers in Honor of Daniel Dinnsen*, ed. by Ashley W. Farris-Trimble and Jessica Barlow. Philadelphia/Amsterdam: John Benjamins Co.
- HANSSON, GUNNAR ÓLAFUR. 2001. Theoretical and Typological Issues in Consonant Harmony. University of California, Berkeley Ph.D. dissertation.
- HANSSON, GUNNAR ÓLAFUR. 2010. *Consonant Harmony: Long-Distance Interaction in Phonology*. UC Publications in Linguistics. Berkeley, CA: University of California Press.
- HAYES, BRUCE.; and COLIN WILSON. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39.379–440.
- HAYES, BRUCE.; COLIN WILSON.; and BEN GEORGE. 2009. *Maxent Grammar Tool*. <http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool/>.
- HAYES, BRUCE.; COLIN WILSON.; and ANNE SHISKO. 2012. Maxent grammars for the metrics of Shakespeare and Milton. *Language* 88.691–731.
- HYMAN, LARRY M. 1987. Prosodic Domains in Kukuya. *Natural Language & Linguistic Theory* 5.311–333.
- HYMAN, LARRY M. 2011. Tone: is it different? *The Handbook of Phonological Theory, 2nd Edition*, ed. by Alan Yu John Goldsmith, Jason Riggle, 197–239. Blackwell.
- INKELAS, SHARON. 2014. *The Interplay of Morphology and Phonology*. Oxford: Oxford University Press.
- INKELAS, SHARON.; and STEPHANIE S SHIH. forthcoming. Tone melodies in the age of surface correspondence. *Proceedings of the 51st Annual Meeting of the Chicago Linguistic Society*. University of Chicago.
- INKELAS, SHARON.; and STEPHANIE S SHIH. 2014. Unstable surface correspondence as the source of local conspiracies. *NELS 44: Proceedings of the 44th Meeting of the North East Linguistic Society*, ed. by Jyoti Iyer and Leland Kusmer, 1:191–204. University of Massachusetts, Amherst: GLSA Publications.
- INKELAS, SHARON.; and CHERYL ZOLL. 2005. *Reduplication: doubling in morphology*. Cambridge University Press.
- INKELAS, SHARON.; and CHERYL ZOLL. 2007. Is Grammar Dependence Real? A comparison between cophological and indexed constraint approaches to morphologically conditioned phonology. *Linguistics* 45.133–171.
- INNES, GORDON. 1969. *A Mende-English Dictionary*. London and New York: Cambridge University Press.
- ITÔ, JUNKO.; and ARMIN MESTER. 1994. Reflections on CodaCond and Alignment. *Phonology at Santa Cruz*, ed. by Jason Merchant, Jaye Padgett, and Rachel Walker, 3:27–46. Santa Cruz: Linguistics Research Center, UC Santa Cruz.
- ITÔ, JUNKO.; and ARMIN MESTER. 1995a. The core-periphery structure of the lexicon and constraints on reranking. *University of Massachusetts Occasional Papers in Linguistics* 18, 181–209. Amherst, MA: GLSA.
- ITÔ, JUNKO.; and ARMIN MESTER. 1995b. Japanese Phonology. *The Handbook of Phonological Theory*, ed. by John Goldsmith, 817–838. Cambridge, MA: Blackwell.
- ITÔ, JUNKO.; and ARMIN MESTER. 1999a. The structure of the phonological lexicon. *The handbook of Japanese linguistics*, ed. by N Tsujimura, 62–100. Malden, MA: Blackwell.
- ITÔ, JUNKO.; and ARMIN MESTER. 1999b. Realignment. *The prosody-morphology interface*, ed. by René Kager, Harry van der Hulst, and Wim Zonneveld. Cambridge: Cambridge University Press.
- ITÔ, JUNKO.; and ARMIN MESTER. 2009. Lexical classes in phonology. *Handbook of Japanese Linguistics*, ed. by Shigeru Miyagawa and Mamoru Saito, 84–106. Oxford: Oxford University Press.
- JÄGER, GERHARD. 2007. Maximum Entropy Models and Stochastic Optimality Theory. *Architectures, Rules, and Preferences: A Festschrift for Joan Bresnan*, ed. by Jane Grimshaw, Joan Maling, Chris Manning, Jane Simpson, and Annie Zaenen. Stanford, CA: CSLI.
- LEBEN, WILL. 1978. The representation of tone. *Tone: A Linguistic Survey*, ed. by Victoria Fromkin, 177–219. New York: Academic Press.
- LEBEN, WILLIAM R. 1973. Suprasegmental phonology. Cambridge, MA: Massachusetts Institute of Technology Ph.D. dissertation.

- MOORE-CANTWELL, CLAIRE.; and JOE PATER. submitted. Gradient Exceptionality in Maximum Entropy Grammar with Lexically Specific Constraints.
- PATER, JOE. 2009. Morpheme-specific phonology: constraint indexation and inconsistency resolution. *Phonological argumentation: essays on evidence and motivation*, ed. by Steven Parker. London: Equinox.
- PATER, JOE.; and ELLIOT MORETON. 2012. Structurally biased phonology: Complexity in learning and typology. *The English and Foreign Languages Journal* 3.1–44.
- PATER, JOE.; ROBERT STAUBS.; KAREN JESNEY.; and BRIAN SMITH. 2012. Learning probabilities over underlying representations. *Proceedings of the Twelfth Meeting of the ACL-SIGMORPHON: Computational Research in Phonetics, Phonology, and Morphology*, 62–71.
- PULLEYBLANK, DOUGLAS. 1986. *Tone in lexical phonology*. Dordrecht: D. Reidel.
- ROSE, SHARON. 2015. Phonology of Ideophones in African Languages. Paper. Paper presented at the WOCAL 8, Kyoto University, Japan.
- ROSE, SHARON.; and RACHEL WALKER. 2004. A Typology of Consonant Agreement by Correspondence. *Language* 80.475–531.
- SHIH, STEPHANIE S. 2016. Super additive similarity in Dioula tone harmony. *Proceedings of the 33rd West Coast Conference on Formal Linguistics*, ed. by Kyeong-min Kim, Pocholo Umbal, Trevor Block, Queenie Chan, Tanie Cheng, Mara Katz Finney, Sophie Nickel-Thompson, and Lisa Shorten. Cascadilla Proceedings Project.
- SHIH, STEPHANIE S.; and SHARON INKELAS. 2014. A subsegmental correspondence approach to contour tone (dis)harmony patterns. *Proceedings of Phonology 2013*. University of Massachusetts, Amherst. <http://journals.linguisticsociety.org/proceedings/index.php/amphonology/article/view/22>.
- SHIH, STEPHANIE S.; and SHARON INKELAS. 2015. Autosegmental Aims in Surface Optimizing Phonology. University of California, Merced and University of California, Berkeley, ms. lingbuzz/002520.
- SMITH, JENNIFER L. 2001. Lexical category and phonological contrast. *Papers in Experimental and Theoretical Linguistics 6: Proceedings of the Workshop on the Lexicon in Phonetics and Phonology*, ed. by Robert Kirchner, Joe Pater, and Wolf Wikely, 61–72. Edmonton: University of Alberta.
- SMITH, JENNIFER L. 2011. Category-specific effects. *The Blackwell Companion to Phonology*, ed. by Marc van Oostendorp, Colin Ewen, Elizabeth Hume, and Keren Rice, 2439–2463. Malden, MA: Wiley-Blackwell.
- VOELTZ, F. K. ERHARD, and CHRISTA KILIAN-HATZ eds. 2001. *Ideophones*. Amsterdam: John Benjamins.
- WAYMENT, ADAM. 2009. Assimilation as attraction: Computing distance, similarity, and locality in phonology. Baltimore, MD: Johns Hopkins University Ph.D. dissertation.
- WILSON, COLIN. 2006. Learning Phonology with Substantive Bias: An Experimental and Computational Study of Velar Palatalization. *Cognitive Science* 30.945–982.
- WILSON, COLIN.; and MARIEKE OBDEYN. 2009. Simplifying subsidiary theory: statistical evidence from Arabic, Muna, Shona, and Wargamay.
- YIP, MOIRA. 2002. *Tone*. Cambridge MA and Oxford: Cambridge University Press.
- ZHANG, JIE. 2004. Contour tone licensing and contour tone representation. *Language and Linguistics* 5.925–968.
- ZHANG, JIE. 2007. Contour tone distribution is not an artifact of tonal melody mapping. *Studies in the Linguistic Sciences* 33.
- ZOLL, CHERYL. 2003. Optimal Tone Mapping. *Linguistic Inquiry* 34.225–268.