

# UNDERPHONOLOGIZATION AND MODULARITY BIAS<sup>1</sup>

*Elliott Moreton*

*University of North Carolina, Chapel Hill*

## **Abstract**

The most straightforward theory of how phonologization interacts with Universal Grammar to determine typology is that UG defines the cognitively possible grammars ("hard" typology), while phonologization determines how frequent they are ("soft" typology). This paper argues instead that some soft typology has a cognitive source, and proposes a formal explanation. Phonological patterns relating tone to tone are shown to be more common than those relating tone to voicing and aspiration (20 families on 5 continents versus 8 families on 4 continents). This soft typological fact cannot be derived from differential robustness of the phonetic precursors, which have similar magnitude (survey of 26 studies of 17 languages). A learning algorithm is proposed in which the learner chooses between Optimality-Theoretic constraint sets based on how probable they make the training data ("Bayesian Constraint Addition"). This biases the learner towards phonologizing processes driven by "modular" markedness constraints, i.e., ones that interact with few other constraints. Its application to the tone case is illustrated by simulation, and compared with alternatives.

## **1. Introduction**

Why are some phonological patterns common, while others are rare or nonexistent? As languages are continually changing and mutating into new languages, it must be the case that some patterns are likelier than others to be innovated or retained in the face of language change. Discussion has centered on a single important typological fact: "phonetic naturalness", the tendency for phonological patterns to look like exaggerated versions of subtle phonetic interactions. Two main factors have been identified as favoring the innovation and retention of "natural" patterns.

One factor is *analytic bias* (Steriade, 2002; Wilson, 2003), cognitive propensities which are hypothesized to make some patterns difficult or impossible to acquire, even from perfect training data. Optimality Theory, and other work in the generative tradition of Universal Grammar, has focused almost exclusively on analytic bias as an explanation for typology in general, and for the "naturalness" bias in particular (e.g., Chomsky and Halle, 1968: 251, 296--297; McCarthy, 1988; Prince and Smolensky, 1993/2004; Archangeli and Pulleyblank, 1994: 391--395; Hayes et al., 2004).

The other factor is *phonetic precursor robustness*. The hypothesis is that phonological patterns are innovated when phonetic precursors, such as coarticulation, are mis- or re-interpreted as phonological ("phonologized", Hyman, 1976). If some precursors are less subtle or more frequent than others, we will see analogous biases in phonological typology, even if the cognitive system supporting phonology is relatively unrestricted as to the patterns it can acquire (e.g., Ohala 1990, 2005; Hale and Reiss 2000; Hume and Johnson 2001; Blevins 2004:19--21, 41, 281--285).

An adequate theory of typology will have to recognize diachronic filtering caused by asymmetries in both analytic bias and precursor robustness. The simplest theory of their interaction is that the analytic bias delimits the cognitively possible grammars ("hard" typology), while precursor robustness determines their frequency ("soft" typology) (Hyman, 2001; Myers, 2002). This paper argues that some soft typological

facts are actually due to analytic bias. There are two sides to the argument, one empirical, one theoretical.

Empirically, it is shown that the typological frequency of a phonological pattern cannot in general be predicted from precursor robustness -- that there exist cases of what I will call "underphonologization", i.e., situations in which two phonetic patterns of similar magnitude correspond to phonological patterns of very different frequency (a phenomenon first noted by Hombert et al., 1979). Specifically, phonological interaction is more common between two tones than between a tone and consonant voicing, even though the phonetic precursors have acoustical effects of the same size. Hence, some soft phonological typology is not just the imprint of phonetic typology.

The theoretical goal is to show how analytic bias can account for the tone generalization. Since the precursors are equally robust, learners must tend to notice tone-tone covariation and overlook consonant-tone covariation. The hypothesis is that the noticing is done by an Optimality-Theoretic grammar, and that the grammar is better at noticing a pattern that is "modular", in the sense that the markedness constraint driving it interacts with few other constraints. Modularity bias emerges when a learner chooses between possible constraint sets based on how probable they make the observed data.

The paper is organized as follows: §2 shows that phonological interactions between two tones outnumber those between a tone and the voicing, aspiration, or fortis-lenis status of a preceding consonant. §3 shows that their phonetic precursors are equally robust. §4 presents the learning algorithm and illustrates its application. Discussion is in §5.

## **2. Tone-tone patterns outnumber Voice-tone patterns**

The cases studied here are the phonologization of tone patterns from phonetic  $F_0$  coarticulation, and that of consonant-tone patterns from the phonetic interaction between  $F_0$  and the obstruent features of voicing, aspiration, and fortis/lenis status—collectively referred to here as "Voicing". These two cases were identified from a wider preliminary search for potential instances of underphonologization. A focused survey was then undertaken to test the hypothesis that dependencies between tone height in adjacent syllables (tone-tone patterns or "TTP") occurred more frequently than dependencies between tone height and the Voicing of a preceding obstruent ("VTP").

Cases of TTP and VTP were located by searching (1) the collection of language-description books held by the University of North Carolina at Chapel Hill and written in Western European languages, (2) print and on-line journals focused on language description, such as *Oceanic Linguistics*, (3) general works on tonal phonology, such as Bradshaw (1999) and the XTONE Project (<http://xtone.linguistics.berkeley.edu>), and (4) the World Wide Web, using the Google search engine to search on the string consisting of *tones* plus the name of each language family listed in Ethnologue (Gordon, 2005). The resulting sample was therefore unsystematic, but broad, and there is no *a priori* reason to expect it to be biased as between TTP and VTP. (It is certainly not exhaustive, but that is in the nature of samples.)

The following selection criteria were applied. (1) The search was restricted to patterns which resembled stylized versions of the phonetic effects of tone-tone and consonant-tone coarticulation, i.e., dependencies between the height of adjacent tones, and between the Voicing of an obstruent and the height of a following tone. This excluded, for example, the devoicing that occurs *after* a falling tone in Kiowa (Kiowa-Tanoan family; Watkins, 1984: 40--41). (2) The sample was limited to

languages in which both TTP and VTP had the opportunity to occur, i.e., tone languages described as having a voicing, aspiration, or fortis-lenis contrast in obstruents. This eliminated a number of languages, mostly in South America and Oceania, which have tone-tone dependencies but only one series of oral stops or fricatives, such as Telefol (Trans-New Guinea; Healey, 1964). The same criterion excluded Yabem (Austronesian; Dempwolff, 1939 [2005]), in which voicing is entirely predictable from tone and hence is non-contrastive. (3) The Voicing contrast had to be free of confounding phonetic factors. For example, Kewa (Trans-New Guinea; Franklin, 1971) and Iau (Geelvink Bay; Bateman, 1990) have stop voicing which is inseparable from nasalization, while Wuming Zhuang (Tai-Kadai; Snyder and Lu, 1997) confounds voicing with preglottalization. (4) As a way of insuring that patterns were phonological rather than phonetic, the pattern was required to neutralize a contrast found elsewhere in the language. Thus, static phonotactic patterns and morphophonemic alternations qualified, but allophonic alternations did not. The falling tone of the Yeneseian language Ket, for instance, begins perceptibly higher when the syllable onset contains a voiceless consonant (Werner, 1997: 21--23), but the pattern does not qualify because the difference is subphonemic. (5) Alternations limited to specific morphemes did not qualify; e.g., the tone assimilations and dissimilations in the noun-class suffixes of Heiltsuk (Wakashan; Kortlandt, 1975). (6) Languages in the survey must have been described from work with living speakers, rather than reconstructed, as reconstructions can be contaminated by theoretical bias (Maddieson, 1976).

Finally, the survey counted language families, defined as top-level categories in Ethnologue (Gordon, 2005), rather than individual languages, to prevent cases of common inheritance from being counted twice. The rationale is that, by counting the different language families in which living languages exhibit TTP and VTP, we are counting *surviving independent innovations* of the two pattern types, and thereby approximating an answer to the question of whether one of them is more likely to be innovated or retained. The results are given in (1) and (2). Italics indicate non-primary sources.

(1) Tone-tone patterns: 20 Ethnologue families, 5 continents.

(a) Africa

- (i) Afro-Asiatic: Gashua Bade: H→L in L\_lH, where l is a clitic- or PPh-boundary (Schuh, 2002). Voicing contrast.
- (ii) Khoisan: lAni: Verbs have HH, HL, LH, but not LL (Vossen, 1997: 97--100). Voicing and aspiration contrasts.
- (iii) Nilo-Saharan: Zarma: In NP, VP, and between subject and verb, L→H in H\_l. HLL is unattested in lexical items (Tersis, 1972: 25--27, 80--81). Voicing contrast.
- (iv) Niger-Congo: Tsonga: When an H-toned prefix is added to a word with only L tones, all tones but the last become H (Baumbach, 1987: 46--47). Voicing and aspiration contrast.

(b) Asia

- (i) Austro-Asiatic: Bagan: Progressive tone assimilation across word boundaries (Li, 1996). Voicing and aspiration contrast.
- (ii) Indo-European: Chakma: A suffix with H tone causes an H-toned monosyllabic root to become L-toned (Huziwara, 2003; p.c., 2005). Voicing contrast.

- (iii) Sino-Tibetan: Many examples in Chinese languages (*Chen, 2000*).  
Voicing contrast.
  - (iv) Hmong-Mien: Hmong Daw: High-level and high-falling tones trigger neutralizing process in other tones. Many lexical exceptions (*Downer, 1967*). Aspiration contrast.
  - (v) Tai-Kadai: Lue: Low-falling tone becomes mid-level after mid-level tone (*Hudak, 1996: xx--xxiii*). Voicing and aspiration contrasts.
- (c) Central and North America
- (i) Caddoan: Caddo: High tone spreads left across an intervening sonorant (*Chafe, 1976: 62; Melnar, 2004: 207*). Voicing contrast.
  - (ii) Huavean: San Mateo Huave: Rightward spreading of H tone within phrases (*Stairs Kreger and Scharfe de Stairs, 1981; Noyer, 1991*). Voicing contrast.
  - (iii) Iroquoian: Oklahoma Cherokee: Except at the right edge of a word, H tones occur in pairs (H on V:, or LH on V: followed by HL on V:, or LH on V: followed by H on V), but odd numbers of L tones are possible (*Wright, 1996*). Voicing contrast.
  - (iv) Kiowa-Tanoan: Kiowa: H and HL lowered to L following HL (*Watkins, 1984: 30*). Voicing and aspiration contrast.
  - (v) Na-Dene: Dakelh/Carrier: Disyllabic nouns can have LH, HL, or HH, but not \*LL. (*Gessner, 2003: 111--127*). Aspiration contrast in stops, voicing contrast in fricatives.
  - (vi) Oto-Manguean: Zapotec: Three contrastive level tones, but disyllabic morphemes don't have final high tone. Some low tones become mid before mid and high tones, between and within words; some mid tones become high in a more complicated tonal context (*Pike, 1948*). Fortis-lenis contrast.
- (d) South America
- (i) Andoke: Andoke: High tone becomes low between two high tones. This is explicitly characterized as neutralizing (*Landaburu, 1979: 50*). Voicing contrast.
  - (ii) Creole (English-based): Saramaccan: In certain syntactic contexts, a series of Ls between two Hs becomes H, neutralizing the contrast between, e.g., /H LHL H/ and /H HLL H/. Some lexically marked Ls resist sandhi (*Ham, 1999*). Voicing contrast.
  - (iii) Tukanoan: Barasana: Bimoraic morphemes can have H or HL tone pattern. An HL root or suffix suppresses H tone on a following suffix (*Gomez-Imbert and Kenstowicz, 2000*). Voicing contrast.
  - (iv) Witotoan: Bora: Successive H tones are allowed, but adjacent L tones are possible only at the end of a tonal domain. The L tones of some suffixes can cause the deletion of root L tones (*Weber and Thiesen, 2001*). Fortis-lenis contrast.
- (e) Oceania
- (i) Skou: Skou: A word-tone system with three tones, H, L, and HL. Overwriting in compounds: second element wins unless it is L (*Donohue, 2003: 339--342*). Voicing contrast restricted to p/b.
- (2) Tone-voice and tone-aspiration patterns: 8 Ethnologue families, 4 continents.
- (a) Africa

- (i) Afro-Asiatic: Lamang: Syllables beginning with voiced obstruents have L tones; other syllables contrast L and H (Wolff, 1983: 66--69).
  - (ii) Niger-Congo: Ewe: H-tone nominal stem (CV) has voiceless obstruent or sonorant C. Non-H-tone nominal stem (CV) has voiced obstruent. This restriction does not apply to CV verbals (Ansre, 1961: 26--32, 36).
- (b) Asia
- (i) Austro-Asiatic: Bolyu: High and low tone registers contrast after voiceless stops, and voiceless and aspirated stops contrast before low-register tones, but high-register tones do not occur after aspirated stops (Edmondson and Gregerson, 1996).
  - (ii) Hmong-Mien: Highland Yao: Aspirated initials occur only with higher tones, while unaspirated ones occur with all tones (Downer, 1961).
  - (iii) Sino-Tibetan: Wuyi: Spreading of high-register tones causes devoicing of intervening voiced obstruents (*Yip, 1995: 485--487*).
  - (iv) Tai-Kadai: Mulao: Aspirated initial stops occur only with lower tones, while unaspirated ones occur with all tones (Wang and Zheng, 1993: 14).
- (c) North America
- (i) Na-Dene: Dakelh/Carrier. Disyllabic nouns have three tone patterns: LH, HL, and (less frequently) HH. Some trigger tone sandhi (lowering of surface H tone on a following word in certain syntactic contexts); these are analyzed as having an underlying H tone. HL sandhi triggers can have either a fortis or a lenis onset in the first syllable. LH sandhi triggers only have lenis onsets (Gessner, 2003: 202--217).
- (d) Oceania
- (i) Skou: Skou: H/L contrast neutralized to phonetic mid tone in syllables with voiced-obstruent onsets. Voicing contrast before falling tone (Donohue, 2003: 350--352).

The survey found TTP in 20 families, and VTP in 8. Since the number of relevant Ethnologue families is large (114 top-level categories for oral natural language, plus 30 language isolates and 78 "unclassified" languages), we can model the survey process as a detector counting events emitted at random by two different Poisson processes, and test whether their parameters differ by using Poisson regression.<sup>2</sup> Assuming independence between families, TTP occur significantly more often in this sample than VTP ( $p = .029$ ), indicating that TTP are either innovated more often or lost less often. This finding is surprising in view of two other facts. First, phonologization can create TTP only in a tone language, whereas VTP can also arise in non-tonal languages undergoing tonogenesis (Svantesson, 1989). Second, VTP relate (phonetically-) adjacent elements in the same syllable, while TTP relate (phonetically-) distant elements in different syllables. This could work against TTP, since distant dependencies are in general less salient than close ones (Moreton and Amano, 1999; Newport and Aslin, 2004; Creel et al., 2004). The next section investigates whether the preponderance of TTP over VTP can be explained by differing precursor robustness.

### 3. Tone-tone and Voice-tone precursors

The less phonetic interaction there is between X and Y, the fewer opportunities there are for the listener to misinterpret phonetic covariation as phonological, and hence the less often the X-Y pattern should become phonologized (Ohalo, 1994; Kavitskaya, 2002: 123--133; Barnes, 2002: 151--159; Myers, 2002; Blevins, 2004:

108--109). If this explanation applies to the typology of tone patterns, it must be true that, across a wide range of languages, the phonetic precursor of tone-tone patterns must be substantially larger than that of the consonant-tone patterns. It is assumed here that the TTP precursor is  $F_0$  coarticulation between tones, and that the VTP precursor is the perturbation of  $F_0$  by the laryngeal features of the preceding consonant--a widespread but still incompletely understood process (for reviews, see Kingston and Diehl, 1994; Jansen, 2004: 52--53)

For the TTP precursor, the literature was searched for studies where vowel  $F_0$  was measured in the context of neighboring tones and reported in physical units such as Hertz. Only tonal languages could be used, since no others could provide a neighboring tone. For each study, the context deemed likeliest to raise the target tone was designated the "Raising" context. E.g., if the study measured a [33] target tone in the contexts [44\_], [35\_], and [21\_], the [35\_] context was the Raising context, because the author's description implied that this context had the highest  $F_0$  adjacent to the target tone. A "Lowering" context was likewise designated. The effect of context was defined to be the target  $F_0$  in the Raising context, divided by that in the Lowering context. This procedure automatically normalizes for inter-speaker differences in  $F_0$  range. If the study measured target  $F_0$  at more than one point, the point closest to the context was used. If the study used multiple target tones, effects were computed for each target, then averaged together within each speaker, and then averaged across speakers. Studies which did not allow these computations were not used: those which did not provide comparable contexts (e.g., Odé, 2002), used non-physical units (e.g., Abramson, 1971), or reported only the difference between the two contexts (e.g., Gandour et al., 1994).

A similar procedure was followed for interaction between  $F_0$  and Voicing of the preceding consonant. This phonetic effect can occur in both tonal and non-tonal languages, and can serve as a phonetic precursor of tone rules in both (Svantesson, 1989; Svantesson and House, 2006), so the survey included both tonal and non-tonal languages. The voiceless, aspirated, or fortis consonant was deemed the Raising context (Hombert et al., 1979). Two measurement points in the target tone were collected, when available: the onset of voicing, and a second point at least 40 ms later.

Figure 1 shows the results. Each plotting code represents one study. The vertical axis shows the effect. A value of 1.0 means the context had no effect; larger values mean that  $F_0$  was higher in the Raising context.

[FIGURE 1 NEAR HERE]

*Figure 1.* Effects of context on  $F_0$ . Each plotting code represents one study. "H" and "L" represent higher and lower tone contexts; "ph", "p", and "b" represent voiceless aspirated, voiceless, and voiced consonantal contexts.

(3) Key to Figure 1.

(a) Effect of tonal context (in tone letters; 1 is lowest, 5 highest):

(M1) Mandarin 2 speakers. /pa pa pa/ with all possible tonal combinations, including neutral tone in positions 2 and 3, in frame sentence. Carryover: average of all 5 tones in context 35\_ vs. 0\_, measured at onset: 1.17. Lookahead: average of all 5 tones in context \_35 vs. \_0, measured at offset: 1.05 (Shen, 1990, Figure 1).

(M2) Mandarin 8 speakers. /mama/ with all possible tonal combinations, in frame sentence. Carryover: average effect on tone of second "ma",

- preceding tone ends high (55\_ or 35\_) vs. low (21\_ or 51\_): 1.22.  
 Lookahead: average effect on F0 maximum of first “ma” tone, (\_55 or \_51) vs (\_35 or \_21): .98 (Xu, 1997, Figures 4 and 7).
- (T1) Taiwanese 1 speaker. 238 disyllables in frame sentence, measured at offset of first syllable and onset of second. Carryover: average of all 5 non-checked tones, \_55 vs. \_21: 1.05. Lookahead, average of all 5 non-checked tones, 55\_ vs. 21\_: .98 (Chang, 1988, Tables 6a--e and 7a--e).
- (T2) Taiwanese 6 speakers; /do/ in carrier sentence after /si/. All possible tonal combinations on /si do/. Average of all /do/ tones in 55\_ vs. 21\_: 1.05 (Lin, 1988, Table 2.7).
- (T3) Taiwanese 4 speakers. /kau/ with 55, 33, 24, 21, or 51 tone, in \_55 or \_21, frame sentence, measured at /kau/ offset. Effect of \_55 vs. \_21: 1.05 (Peng, 1997, Figures 2--5 ).
- (V) Vietnamese 1 speaker. 559 disyllables in frame sentence. Only most extreme contextual variants were reported. Carryover: average for 4 non-glottalized tones, measured at onset, after the context that made it highest vs. the context that made it lowest: 1.34. Lookahead: average for 4 non-glottalized tones, before context that made it highest vs. context that made it lowest, measured at offset: 1.08 (Han and Kim, 1974, Figure 2).
- (b) Effects of preceding consonant voicing:
- (Dh) Dakelh (Carrier) 2 speakers; monosyllabic nouns (all H tone) in 4 different morphological and prosodic contexts.; 204 items for Speaker A, 98 for Speaker C. Measured average F<sub>0</sub> for entire vowel. Initial voiceless fricatives vs. voiced fricatives: 1.17 (Gessner, 2003: 175--177).
- (E2) English 5 speakers, /sp st sk/ vs. /b d g/ in “symmetrical CVC syllables” (e.g., /spip/, /bib/); frame sentence. Average across 5 vowels, measured at first glottal pulse after release: 1.36; measured at 5th glottal pulse (= about 40 ms): 1.16 (Ohde, 1984).
- (F1) French. 1 speaker. 4 CVC minimal pairs in frame sentence, /p t k/ vs. /b d g/. Average across three vowels (/i a u/) measured at voicing onset, ratio = 1.13 (Fischer-Jørgensen, 1972: 173--174).
- (F2) French 5 speakers, 18 CV syllables, /p t/ vs. /b d/. Average across 3 vowels (/i a u/), measured at voicing onset; ratio = 1.19 (Serniclaes, 1992, Table 2.5); measured 40 ms after release, 1.06 (Serniclaes, 1992, Figure 2.4).
- (H) Hindi 1 speaker; trisyllabic nonsense phrases /thikiCi/ in isolation, where C is one of /p t/ vs. /b d/; likewise /CiCi/, /Ci/, /iCi/. Measured 3 pitch periods at release for /b d/, at voicing onset for /p t/: 1.22 (Kagaya and Hirose, 1975, Table II).
- (J) Japanese 3 speakers. Disyllables /kVCV/, where C was /p t k/ vs. /b d g/, identical vowels, initial accent; frame sentence. Average across /e a o/, measured at voicing onset: 1.08; at vowel steady-state: 1.02 (Kawahara, 2005, Figures 20 and 21).
- (Tb) Lhasa Tibetan 1 speaker, /pa/ vs. /ba/, frame sentence. Measured at oral release; ratio = 1.36 (Kjellin, 1977). (Possibly phonological.)
- (Th) Thai 1 speaker; CVV syllables, C was /p t/ vs. /b d/. Average across 4 tones (HIGH excluded because of gap in table) and 3 vowels, measured at onset: 1.17 (Gandour, 1974, Table III).

- (Y) Yoruba 2 speakers, /k/ vs. /g/. Average across 3 tones (H/M/L), measured at voicing onset; ratio = 1.27; 40 ms later: 1.03 (Hombert et al., 1979, Figure 3).
- (c) Effects of preceding consonant aspiration:
- (C) Cantonese 3 speakers, /phei55/ vs. /pei55/, frame sentence. Measured at voicing onset: 1.08 (Zee, 1980, Table I).
- (Dn) Danish 6 speakers; 2-syllable words in frame sentence; initial /ph th kh/ vs. /p t k/. Measured at start of following vowel: 1.09 (Jeel, 1975, Table I).
- (Dh) Dakelh (Carrier) See (b) for circumstances. Initial voiceless aspirated vs. voiceless unaspirated: 1.02 (Gessner, 2003: 175--177).
- (E1) English 5 speakers, /ph th kh/ vs. /b d g/ before /i/; frame sentence. Measured at voicing onset: 1.14; 40 ms later: 1.08 (Hombert et al., 1979, Figure 1).
- (E2) English 5 speakers, /ph th kh/ vs. /b d g/ in “symmetrical CVC syllables” (e.g., /phip/, /bib/); frame sentence. Average across 5 vowels, measured at first glottal pulse after release: 1.23; at 5th glottal pulse (= about 40 ms): 1.15 (Ohde, 1984, Figure 5).
- (Gm) German 1 speaker; /t/ vs. /d/ between stressed /ai/ and an unstressed schwa or syllabic /n/. Words read in isolation, in a monotone. Measured at release: 1.11; “later”: 1.00 (Kohler, 1982, Table I).
- (H) Hindi See (b) for circumstances. Measured 3 pitch periods at release for /b d/, at voicing onset for /ph th/: 1.16 (Kagaya and Hirose, 1975, Table II).
- (K1) Korean 2 speakers; 1400 tokens. /ph th kh/ vs. /p t k/. Measured at voice onset: 1.27 (Han and Weitzman, 1970, Table II).
- (K2) Korean 2 speakers, nonsense words /CV/ and /VCV/. /ph th kh/ vs. /p t k/. Measured near voice onset: 1.02 (Kagaya, 1974, Table II).
- (M) Mandarin 7 speakers, /tha/ vs. /ta/, with all four tones; /pha35/ substituted for meaningless /tha35/. Real disyllabic words, with target in first or second syllable; all possible surface tone combinations. Measured at 1st glottal pulse, overall average: .91; 40 ms later: .99 (Xu and Xu, 2003, Figure 7).
- (Si) SiSwati 4 speakers. 8 real-word CVX stimuli, /ph/ vs. /\_/. Measured at vowel onset: 1.09; 40 ms later: 1.08 (Wright and Shyrock, 1993, Figure 1).
- (Sw) Swedish 3 speakers; th\_t vs. d\_d, long vs. short low vowel; second consonant short iff vowel long; frame sentence. Mean of short and long vowels; measured at onset of vowel: 1.23 (Löfqvist, 1975, Table X).
- (Th) Thai See (b) for circumstances. /ph th/ vs. /b d/, measured at onset: 1.08 (Gandour, 1974, Table III)

Pairwise comparisons between each tone condition and each of the consonant conditions were carried out using a linear mixed-effects model with Language as a random effect. Only two of the 8 comparisons reached significance at the 5% level (uncorrected for multiple comparisons). The effect of following tone was significantly smaller than that of the voiced/unvoiced status of a preceding obstruent measured at voicing onset (means of 1.20 and 1.03, respectively;  $p = .0091$ ), whereas the effect of preceding tone was significantly greater than that of the aspirated/unaspirated status of a preceding obstruent measured at voicing onset (means of 1.24 and 1.03;  $p = .0159$ ).

The survey does not support the hypothesis that phonetic tone-tone interaction is greater than Voicing-tone interaction; rather, the two effects are of about the same size. Thus, the different frequency of TTP and VTP cannot be explained by differences in precursor robustness. When vowel height is substituted for tone height, a similar picture emerges: height-height patterns are more common than height-Voicing patterns, but vowel F1 is less affected by other vowels than by consonant laryngeal features (Moreton 2007).

#### 4. Modularity bias via Bayesian Constraint Addition

With precursor robustness eliminated as an explanation, it is the turn of analytic bias. Why are Voicing-tone and Voicing-height patterns underphonologized relative to tone-tone and height-height patterns? It will not do to say that the latter are more “salient” (e.g., because of their featural symmetry) and stop there, for that merely restates the problem. The formal challenge lies in *deriving* the greater salience from constraint interaction.

This section presents an explicit model of how a Speaker's phonetic precursor becomes phonologized as an optional (probabilistic) phonological process by a misperceiving Learner. The key assumptions are that markedness constraints have to be added to the ranking before they can be ranked, and that the Learner decides whether to add or not based on which choice makes the observed training data more probable (“Bayesian Constraint Addition”, BCA). It is shown that BCA automatically disfavors adding markedness constraints that interact with many other constraints. A consequence is that it discourages interaction between phonological subsystems, and thus favors tone-tone interactions over voice-tone interactions.

Background assumptions are stated in §4.1. The learning component is presented in §4.2, and a simulation of the tone facts in §4.3. Discussion of nonstandard assumptions, and predictions, is in §4.4. The model is meant only to show that “soft” analytic bias, in which some patterns are discouraged but not forbidden, can be derived from constraint interaction, and so it has been made as simple as possible. Embedding the concept in a realistic model of acquisition and phonologization is left for future research.

##### 4.1. Background assumptions

Underlying and surface representations are restricted to words of the form *maCaC*, where each vowel bears either H or L tone, each C is [b] or [p], and [m] is an irrelevant sonorant. Thus, every word has equal numbers of tones and obstruents, one tone-tone sequence, and one obstruent-tone sequence. The Speaker's lexicon contains all 16 possibilities, which are produced with equal frequency:

- (4)
- mápáp, mápàp, màpáp, màpàp,*  
*mápáb, mápàb, màpáb, màpàb,*  
*mábáp, mábàp, màbáp, màbàp,*  
*mábáb, mábàb, màbáb, màbàb,*

The Speaker's grammar is fully faithful, so that the 16 forms in (4) occur with equal frequency as phonological surface representations. These undergo some phonetically-biased coarticulatory distortion in transmission to the Learner. The Learner is able to compensate for (i.e., undo) most of it, but some coarticulated tokens are misperceived as different from the Speaker's phonological surface representation.

As in Prince and Tesar (1999), the Learner is assumed to be acquiring only surface distributions, not yet a lexicon. Ignorant of the actual contents of the lexicon, this Learner (unlike Prince and Tesar's) assumes correctly that all 16 possible inputs are equally likely, and that any observed inequality among surface forms is due to unfaithful mapping caused by ranked constraints.

We will compare two different cases of phonologization. In the "LH condition", tonal coarticulation becomes phonologized as rightward L-tone spreading. In the "bH condition", the phonetic lowering effect of a voiced obstruent becomes a process lowering /H/ after /b/. Both H tone and [+voice] are assumed to be privative, implying markedness constraints \*H and \*[+voice] which the learner is assumed to have already ranked very low. The analysis of vowel harmony proposed by Pulleyblank (2004) is adapted to tone spreading:

(5) Constraints for tone spreading (after Pulleyblank, 2004).

(a) MAX-H: "Don't delete an H tone." Give one mark for each underlyingly H-toned vowel with no H-toned surface correspondent.

(b) \*LH: "No LH tone sequences." Give one mark for each H tone in the next syllable following an L tone.

These constraints allow only two grammars, distinguished by their effect on /LH/ (the tie between [LL], [HL], and [HL] is broken by the low-ranked \*H):

(6)

(a) Faithful realization

/LH/	MAX-H	*LH
☞ LH		*
LL	*!	

(7) Rightward L-tone spreading

b. /LH/	*LH	MAX-H
LH	*!	
☞ LL		*

Analogous constraints are assumed to be involved in lowering H after /b/:

(8) Additional constraints for post-/b/ lowering

(a) MAX-VOICE: "Don't delete a [voice] feature". \* = underlying voiced segment without a voiced surface correspondent.

(b) \*bH: "No bH sequences." \* = a H tone following a voiced obstruent

Since both MAX-VOICE and MAX-H are relevant, there are three possible grammars, with different effects on /bH/:

(9)

(a) Faithful realization

/bH/	MAX-VOICE	MAX-H	*bH
☞ bH			*
bL		*!	
pH	*!		

(b) Post-/b/ lowering

/bH/	*bH	MAX-VOICE	MAX-H
bH	*!		
$\text{☞}$ bL			*
pH		*!	

(c) Pre-/H/ devoicing

/bH/	*bH	MAX-H	MAX-VOICE
bH	*!		
bL		*!	
$\text{☞}$ pH			*

Finally, in order to allow a probabilistic phonetic precursor to be phonologized as a probabilistic phonological process, constraint rankings are continuous. The rank of a constraint specifies the mean position where it will be observed when the grammar is consulted in any particular case (Nagy and Reynolds, 1997; Zubritskaya, 1997; Boersma, 1998: 269--273; Boersma and Hayes, 2001).

#### 4.2. Bayesian Constraint Addition

Crucially, \*LH and \*bH are off-stage in the Learner's initial state, and must be added to the set of ranked constraints in response to the corpus of perceived training data  $D$ . The Learner compares two hypotheses.  $H_0$  is that the constraint set is what the Learner previously thought it was, while  $H_1$  is that it also contains the new constraint. To decide between the two hypotheses, the Learner compares how probable each one is given  $D$ . Bayes's Rule prescribes how to do this (MacKay, 2003: 48--57):

$$(10) \quad \frac{P(H_1 | D)}{P(H_0 | D)} = \frac{P(D | H_1)P(H_1) / P(D)}{P(D | H_0)P(H_0) / P(D)} \\ = \frac{P(D | H_1)P(H_1)}{P(D | H_0)P(H_0)}$$

Before hearing the data, the Learner estimates that  $H_0$  is true with probability  $P(H_0)$ , and  $H_1$  with probability  $P(H_1)$ . The ratio of these probabilities,  $P(H_1)/P(H_0)$ , reflects the Learner's prior bias towards one or the other hypothesis. To keep things simple, the Learner is assumed to be initially unbiased, so that the ratio is 1. When the data arrives, the original estimate is multiplied by  $P(D | H_1)/P(D | H_0)$ , the ratio of how likely  $D$  is if  $H_1$  is true to how likely  $D$  is if  $H_0$  is true. The result is the Learner's new estimate of the relative probability of  $H_0$  and  $H_1$ . ( $P(D)$ , the Learner's prior estimate of how probable the data itself is, cancels out in the numerator and denominator.)

Now suppose  $D$  exhibits a pattern that is inconsistent with any ranking under  $H_0$ , but consistent with some rankings under  $H_1$ . The Learner has to choose between two improbable coincidences. Is the Speaker's constraint set the one described by  $H_0$ , and the "pattern" merely a statistical fluke? Or is the constraint set the one described by  $H_1$ , whose constraints just happen to be ranked exactly right? The key point is: The plausibility of  $H_1$  as an explanation for the pattern depends on how many *other* patterns  $H_1$  allows. If the observed pattern is a one-in-ten coincidence under  $H_1$ , then  $H_1$  is a better alternative to  $H_0$  than if the pattern is only a one-in-ten-thousand coincidence. The more distinct patterns  $H_1$  allows, the smaller  $P(D | H_1)$ , and hence,

by (10), the smaller  $P(H_j | D)$ , i.e., the less credence the Learner places in  $H_j$ . This effect has been termed the "Bayesian Occam's Razor", since it penalizes less-restrictive hypotheses (MacKay, 2003: 343ff.).

The connection to modularity is that modular constraint sets make for more-restrictive hypotheses. Suppose a (discretely-ranked) constraint set contains two types of constraint,  $m$  involving only tone, and  $n$  involving only segments. The set is modular: Rankings within each subsystem matter (they potentially change the underlying-to-surface mapping), but rankings of tonal constraints with respect to segmental constraints do not. There are  $(m+n)!$  ways to rank the entire set, but actually there are at most  $m!n!$  distinct grammars. If a new tonal constraint is added, the number of potentially distinct grammars increases slightly, to at most  $(m+1)!n!$ . But if a new tone-segment constraint is added, the wall of separation between the modules collapses, and the number of potentially distinct grammars can soar above  $(m+n)!$ .<sup>3</sup> Adding a non-modular constraint causes a steeper increase in the number of different grammars, reduces  $P(D | H_j)$  more, and so incurs a stronger penalty from the Razor.

Continuous ranking amplifies this effect. For the Learner to recognize whether some H tones are turning into L, what matters is the perceived proportion of four word types. In the LH condition, those types are HH, HL, LH, and LL; in the bH condition, they are pH, pL, bH, and bL. If  $r$  is the rate at which the Speaker's intended productions are misperceived, then the frequencies of these types in  $D$  are as shown in the "Perceived" column of Table 1:

Table 1. Corpus frequencies, actual and predicted. See text below for explanation of  $p$ ,  $q$ , and  $s$ .

Word type	Intended by Speaker	$D$ Perceived by Learner	Predicted by Learner's hypotheses	
			$H_0$ : No change	$H_j$ : New constraint needed
			LH condition	bH condition
HH, pH	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}(1 + s_{\alpha,\beta})$
HL, pL	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
LH, bH	$\frac{1}{4}$	$\frac{1}{4}(1-r)$	$\frac{1}{4}$	$\frac{1}{4}(1-p_\alpha)$ $\frac{1}{4}(1-q_{\alpha,\beta} - s_{\alpha,\beta})$
LL, bL	$\frac{1}{4}$	$\frac{1}{4}(1+r)$	$\frac{1}{4}$	$\frac{1}{4}(1+p_\alpha)$ $\frac{1}{4}(1+q_{\alpha,\beta})$

In either condition, the Learner has two hypotheses.  $H_0$  is that the grammar contains only MAX-VOICE and MAX-H. It predicts that the four categories occur with frequency 1/4. The more the observed deviation from equal proportions in  $D$ , the more improbable is  $H_0$ .  $H_j$  differs depending on condition.

In the LH condition,  $H_j$  says that the constraint set is {MAX-VOICE, MAX-H, \*LH}. Under  $H_j(\text{LH})$  the grammar has one genuinely adjustable parameter,  $\alpha$ , the ranking distance between the \*LH and MAX-H. Changing  $\alpha$  changes  $p_\alpha = P(*\text{LH} \gg$

MAX-H |  $H_I(\text{LH})$ ,  $\alpha$ ). The distance  $\beta$  between \*LH and MAX-VOICE can also be adjusted, but that has no effect on the output of the grammar, because the two constraints do not interact. By setting  $\alpha$  so that  $p_\alpha = r$ , the  $H_I(\text{LH})$  grammar can be made to match  $D$ , while the  $H_0(\text{LH})$  grammar cannot. This is shown in Figure 2.

[FIGURE 2 NEAR HERE]

Figure 2.  $H_I(\text{LH})$  and  $H_I(\text{bH})$  nominally have two adjustable ranking parameters. However, adjusting  $\beta$  in  $H_I(\text{LH})$  has no effect on the input-output map.

In the bH condition,  $H_I$  says that the constraint set is {MAX-VOICE, MAX-H, \*bH}. Unlike  $H_I(\text{LH})$ ,  $H_I(\text{bH})$  gives the grammar two genuinely adjustable parameters. Both  $\alpha$  and  $\beta$  matter, since \*bH interacts with both MAX-H and MAX-VOICE. The extra degree of freedom allows for more distributions among the four categories than in the LH condition, as shown in the last column of Table 1. Matching the Speaker's data requires setting  $\alpha$  and  $\beta$  so that  $q_{\alpha,\beta} = r$  and  $s_{\alpha,\beta} = 0$ . Again,  $H_0(\text{bH})$  cannot be adjusted to fit the data.

In both conditions, the Learner is confronted with the same distribution. The data slightly mismatches  $H_0$ , to the same degree in the LH and bH conditions.  $H_I(\text{LH})$  and  $H_I(\text{bH})$  can accommodate the data, but  $H_I(\text{bH})$  needs finer tuning to do so, as two parameters have to be set properly rather than just one. The data is therefore deemed less likely (more of a coincidence) by  $H_I(\text{bH})$  than by  $H_I(\text{LH})$ . As a result,  $H_I(\text{LH})$  is a better alternative to  $H_0(\text{LH})$  than  $H_I(\text{bH})$  is to  $H_0(\text{bH})$ .

### 4.3. Simulation

To make the discussion entirely concrete, let  $r$  be set arbitrarily to 1/15. We need to specify how the probabilities  $p$ ,  $q$ , and  $s$  are related to the ranking distances  $\alpha$  and  $\beta$ . Following Boersma (1998; Boersma and Hayes, 2001), each time the grammar is consulted, normally-distributed noise is added to each constraint's fixed position to determine its observed position. The observed distance between two constraints is thus the difference between two independent normal distributions with equal variance, and hence is itself normally distributed. For  $H_I$ , then, we can adopt a scale on which the markedness constraint is always observed at 0, while the observed positions of MAX-H and MAX-VOICE are normally distributed with means of  $\alpha$  and  $\beta$  and standard deviation of 1.<sup>4</sup> (For  $H_0$ , the rankings are irrelevant.) The Learner initially assumes that all possible rankings are equally probable (i.e., a uniform prior distribution for  $\alpha$  and  $\beta$ , reflecting the Learner's complete ignorance).

For any corpus  $D$ , the probability of  $D$  under a hypothesis is just the product of the probabilities assigned to each word of  $D$  by that hypothesis. Table 1 shows that, for  $H_0(\text{LH})$  or  $H_0(\text{bH})$ , the probability is always 1/4. Hence, if there are  $N$  tokens in  $D$ , then

$$(11) \quad P(D | H_0(\text{LH})) = P(D | H_0(\text{bH})) = \left(\frac{1}{4}\right)^N$$

In the LH condition, let  $n_{HH}$ ,  $n_{HL}$ ,  $n_{LH}$ , and  $n_{LL}$  be the number of words in each category in  $D$ . Then for a given  $\alpha$ ,

(12)

$$P(D|H_1(LH),\alpha) = \left(\frac{1}{4}\right)^{n_{HH}} \left(\frac{1}{4}\right)^{n_{HL}} \left(\frac{1}{4}(1-p_\alpha)\right)^{n_{LH}} \left(\frac{1}{4}(1+p_\alpha)\right)^{n_{LL}}$$

The probability of  $D$  under  $H_1(LH)$  is obtained by integrating (12), times the probability density at each  $\alpha$ , over all  $\alpha$ . Likewise, in the bH condition, for given  $\alpha$  and  $\beta$ ,

(13)

$$P(D|H_1(bH),\alpha,\beta) = \left(\frac{1}{4}(1+s_{\alpha,\beta})\right)^{n_{pH}} \left(\frac{1}{4}\right)^{n_{pL}} \left(\frac{1}{4}(1-q_{\alpha,\beta}-s_{\alpha,\beta})\right)^{n_{bH}} \left(\frac{1}{4}(1+q_{\alpha,\beta})\right)^{n_{bL}}$$

with  $P(D|H_0(bH))$  obtained by integrating (13), times the probability density at each  $(\alpha, \beta)$ , over all  $\alpha$  and  $\beta$ . For practical reasons, the integrals were approximated by discretizing  $\alpha$  and  $\beta$  as a grid of points spaced 0.1 units apart in the region  $[-4, 4] \times [-4, 4]$ . These limits allow  $p_\alpha$ ,  $q_{\alpha,\beta}$ , and  $s_{\alpha,\beta}$  to vary from  $< .0001$  to  $> .9999$ . Random corpora  $D$  of size  $N$  ranging from 0 to 3000 were simulated using R (R Development Core Team, 2005). The resulting likelihood ratios were logit-transformed and plotted to produce Figure 3.

[FIGURE 3 NEAR HERE]

*Figure 3.* Simulation results. Solid curve: LH. Dashed curve: bH. Log-odds of 0 means that the Learner assigns equal probability to  $H_0$  and  $H_1$ . Positive values favor  $H_1$ . Bars show 95%  $t$  confidence intervals around mean of 1000 runs of the simulation.

The solid curve shows the LH condition. Initially, the log-odds is near 0, because with little data the Learner can't tell whether  $D$  is more consistent with either hypothesis. (We can't tell whether a coin is unfair by tossing it three times.) As more data arrives, it becomes clear that only a very specific tuning of  $H_1(LH)$  can match it, and the Bayesian Occam's Razor begins to cut against  $H_1(LH)$ . The log-odds drops rapidly. But as even more data arrives, and the Learner gets better estimates of the frequencies of the four word categories, it becomes clear that, while the data is consistent only with a very specific tuning of  $H_1(LH)$ , it is not consistent at all with  $H_0(LH)$ . The log-odds begins to rise. After about 1250 words, the Learner is again equiposed between  $H_0(LH)$  and  $H_1(LH)$ . Thereafter, the preference for  $H_1(LH)$  increases without bound.

In the bH condition (dashed curve),  $H_0$  is the same. However,  $H_1(bH)$  requires an even more specific tuning to match the data, since both  $\alpha$  and  $\beta$  have to be just right. This incurs a greater penalty from the Razor. Any criterion set by the Learner for accepting a new constraint will therefore be met sooner in the LH than the bH condition.

This concludes the demonstration that BCA can delay phonologization of a non-modular pattern. In a human learner, the delay would increase the chance that the end of acquisition would intercept phonologization, and so would reduce the typological frequency of the pattern.

#### 4.4. Comments on the model

Despite the term “constraint addition”, it is enough that certain constraints initially be unable to dominate any other constraint, and remain so until the learner explicitly grants permission. These constraints may begin in a separate stratum at the bottom. Alternatively, they may be created by local conjunction (Smolensky, 1996), inductive grounding (Hayes, 1999), or constraint schemata (Smith, 2004), and literally added to the constraint set.

Continuous ranking is not in principle necessary, but it provides two crucial services which are otherwise hard to come by. (1) For BCA to be effective, there has to be a *large* difference between the number of grammars available under  $H_i(\text{LH})$  and  $H_i(\text{bH})$ . Their discrete versions afford just 2 and 3 grammars respectively, but this difference is greatly amplified by gradience. Larger constraint sets may have a similar effect; however, early trials with four and five discretely-ranked constraints have not been encouraging. (2) A subtle phonetic effect that turns *some* LH’s into LL’s cannot fool a discrete learner into thinking that the correct grammar bans *all* LH’s. This is a problem that has to be faced by any theory of misperceptive sound change: Misperception affects one utterance at a time, while sound change affects the whole grammar.

Continuous rankings solve the problem by allowing the first generation to innovate an optional phonological process rather than leap directly to a categorical one. If we suppose that this generation also retains the phonetic precursor, then their speech contains even fewer surface LH’s than their parents’, since some are changed to LL by the phonology and others by the phonetics. The second generation, learning from this input, will add \*LH earlier and rank it higher than the first. Repeated cycles will lead to a near-categorical grammar with high-ranked \*LH.<sup>5</sup>

Modularity bias overlaps with the *SPE* evaluation metric (bias against multiple features) and Feature Geometry (bias against cross-tier interaction), but differs empirically from both. *SPE* counts features without regard to their content, so that a rule involving [+nasal] and [--nasal] is just as costly as one involving [+nasal] and [--high] (Chomsky and Halle, 1968: 334--335). Feature Geometry treats all single operations alike, so that a rule spreading [+anterior] is just as complex as one spreading [C-place] (Clements and Hume, 1995: 250). Modularity bias would favor the first of each pair of examples.

The BCA account of modularity bias predicts, uniquely, that the *second* constraint linking domains X and Y is easier to acquire than the first, because the damage is already done. It also predicts rapid acquisition of constraints linking phonology with morphology, since there are no purely morphological constraints for them to interact with. Indeed, morphological conditioning seems common, especially in the complex or unnatural processes most plausibly attributed to language-particular induced constraints (see, e.g., the discussion of Lardil FREE-V in Prince and Smolensky, 1993/2004). BCA may also apply to constraints phonologized from the learner’s own phonetics (Hayes, 1999, Smith, 2004). Macken (1995: 690--691) notes that child phonology abounds in consonant- and vowel-harmony processes, but that “none of the primary rules of acquisition and few of the other attested rules in the first year or two (ages one to three) show interactions between consonants and vowels.”

#### 5. General discussion

The hypothesis that causes and effects line up in a simple way, pairing hard typology with Universal Grammar and soft typology with other factors affecting language change (Hyman, 2001; see also Myers, 2002) is probably too strong. We

have seen that at least one soft generalization, the preponderance of tone-tone over consonant-tone patterns, cannot be explained by the likeliest diachronic factor, precursor robustness. The same holds for height-height and height-voice interactions. These generalizations, and others involving modularity bias, can be derived from constraint interaction if the learner chooses among constraint sets based on how probable they make the observed data. What are the alternatives?

Anttila (1995; Anttila and Cho, 2004) has used *combinatorial bias*, which occurs when the mapping /A/→[B] is generated by more rankings than /A/→[C]. If all rankings are equally probable, then /A/→[B] is predicted to occur more often than /A/→[C]. Originally applied to within-language variation, this idea has been extended to typology by Coetzee (2002). Here, though, it leads in the wrong direction. A language lacking [LH] must rank \*LH over MAX-H. A language lacking [bH] needs only to have \*bH dominate *at least one* of MAX-H and MAX-VOICE. Hence [bH]-less languages should outnumber [LH]-less ones. Generally, the more features appear in a markedness constraint, the more ways there are to satisfy it. Combinatorial bias thus favors processes triggered by featurally *complex* markedness constraints.

Another possibility is *initial-state bias*. Many OT learning algorithms require that markedness initially dominate faithfulness (Gnanadesikan, 1995; Smolensky, 1996). This creates a bias towards grammars which allow fewer surface forms, but not towards modular processes. The problem is the same as before: To block the effect of \*LH, \*LH must be demoted below MAX-H. To block that of \*bH, \*bH must be demoted below *both* MAX-H and MAX-VOICE. Less-modular markedness constraints are harder to deactivate, since deactivating them requires more changes relative to the initial state. If distance from the initial state is what determines typological frequency, then the initial M » F hypothesis is wrong, and some constraints are initially either bottom-ranked or outside the ranking entirely (just as in BCA).

BCA is thus the only proposal on offer that derives modularity bias (or featural-simplicity bias) from constraint interaction. A still-viable alternative explanation of the tone and height facts is that non-grammatical perceptual effects may skew the training data before it reaches the learner's pattern-finding mechanisms. For example, compensation for coarticulation may affect a shorter time window than coarticulation itself does, so that compensation fails more often for TT than CT interactions. This may be true *instead of* or *in addition to* analytic bias. The issue will only be settled in the lab. For instance, the existence of an analytic bias in favor of modularity is supported by the finding that English-speaking adults trained on non-coarticulated C<sub>1</sub>V<sub>1</sub>C<sub>2</sub>V<sub>2</sub> stimuli in an artificial-language paradigm learned to recognize height agreement between V<sub>1</sub> and V<sub>2</sub>, and voice agreement between C<sub>1</sub> and C<sub>2</sub>, better than height-voice agreement between V<sub>1</sub> and C<sub>2</sub> (Moreton, 2007).

Cognition and phonetics interact to determine typology in ways more complicated (and interesting) than has been generally acknowledged. Further progress will require a better quantitative understanding of the typology of phonetic precursors, and of the differential receptiveness of learners to different patterns.

## Notes

1. Several people have contributed to this paper, including Andries Coetzee, Paul de Lacy, Shigeto Kawahara, Paul Kiparsky, John McCarthy, Steve Parker, Joe Pater, Jennifer L. Smith, Paul Smolensky, David Teeple, and Anne-Michelle Tessier. It also owes much to Adam Albright and Josh Tenenbaum's 2005 Linguistic Institute class. Thanks are due to Chris Wiesen for statistical advice. Statistics, graphs, and simulations used R 2.2.1 (R Development Core Team 2005). Errors or omissions are solely the author's. This version is current as of June 5, 2008.

2. I am indebted for this suggestion to Chris Wiesen of the Odum Institute for Research in Social Science at the University of North Carolina at Chapel Hill.

3. These worst-case scenarios require artificial construction. Linguistically-plausible constraints would probably lead to much smaller numbers in all three instances.

4. This is an approximation, as it ignores the covariance between the two ranking distances MAX-H to \*bH, and MAX-VOICE to \*bH. For example, if all three constraints are ranked equally, the true  $P(\text{MAX-H} \gg \text{MAX-VOICE} \gg *bH)$  is  $1/6$ , as all 6 rankings are equally probable, but the approximation makes it  $1/8$  ( $= P(\text{MAX-H} \gg *bH) P(\text{MAX-VOICE} \gg *bH) P(\text{MAX-H} \gg \text{MAX-VOICE} \mid \text{MAX-H}, \text{MAX-VOICE} \gg *bH)$ ). Without this approximation, however, the simulation would run 81 times slower.

5. The prediction is that a new phonological process should grow gradually out of its phonetic precursor, rather than appearing all at once. Such a pattern has been observed (Moreton and Thomas in press), but could also be due to growth in the precursor itself.

## References

- Abramson, A. (1971) The coarticulation of tones: an acoustic study of Thai. In T. L. Thongkum, V. Panupong, P. Kullavanijaya, and M. R. K. Tingsabath (eds.), *Studies in Tai and Mon Khmer Phonetics and Phonology, in Honour of Eugénie J. A. Henderson* 1-9. Bangkok: Chulalongkorn University Press.
- Ansre, G. (1961) *The Tonal Structure of Ewe*. M.A. thesis, Kennedy School of Missions of the Hartford Seminary Foundation. Hartford Studies in Linguistics, Number 1.
- Anttila, A. (1995) Deriving variation from grammar: a study of Finnish genitives. Retrieved on 25 August 2005 from <http://roa.rutgers.edu>.
- Anttila, A., and Y.-M. Cho (2004) Variation and change in Optimality Theory. In: John J. McCarthy (ed.), *Optimality Theory in Phonology* 569--580. Malden: Blackwell.
- Archangeli, D., and D. Pulleyblank (1994) *Grounded Phonology*. Cambridge: MIT Press.
- Barnes, J. (2002) *Positional Neutralization: A Phonologization Approach to Typological Patterns*. Ph.D. dissertation, University of California, Berkeley.
- Bateman, J. (1990) Iau segmental and tone phonology. *Nusa* 32: 29--42.
- Baumbach, E. J. M. (1987) *Analytical Tsonga Grammar*. Pretoria: University of South Africa.
- Blevins, J. (2004) *Evolutionary Phonology*. Cambridge: Cambridge University Press.
- Boersma, P. (1998) *Functional Phonology: Formalizing the Interactions between Articulatory and Perceptual Drives*. Ph. D. dissertation, University of Amsterdam.
- Boersma, P., and B. Hayes (2001) Empirical test of the Gradual Learning Algorithm. *Linguistic Inquiry* 32(1): 45--86.
- Bradshaw, M. M. (1999) *A Crosslinguistic Study of Consonant-Tone Interaction*. Ph. D. dissertation, Ohio State University.
- Chafe, W. L. (1976) *The Caddoan, Iroquoian, and Siouan languages*. The Hague: Mouton.
- Chang, Y.-C. (1988) Sandhi tonal des syntagmes dissyllabiques du Min-Nan parlé à Taiwan. *Cahier de Linguistique Asie Orientale* 17(2): 193--234.
- Chen, M. (2000) *Tone Sandhi: Patterns Across Chinese Dialects*. Cambridge: Cambridge University Press.
- Chomsky, N., and M. A. Halle (1968) *The Sound Pattern of English*. Cambridge: MIT Press.
- Clements, G. N., and E. V. Hume (1995). The internal organization of speech sounds. In John A. Goldsmith (ed.), *The handbook of phonological theory* 245--306. Cambridge: Blackwell.
- Coetzee, A. (2002) Between-language frequency effects in phonological theory. Unpublished ms, Department of Linguistics, University of Massachusetts, Amherst.
- Creel, S. C., E. L. Newport, and R. N. Aslin (2004) Distant melodies: Statistical learning of non-adjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 1119--1130.
- Dempwolff, O. (1939 [2005]) *Grammar of the Jabêm Language in New Guinea*. Translated by Joel Bradshaw and Francisc Czobor. Oceanic Linguistics Special Publication No. 32. Honolulu: University of Hawai'i Press.
- Donohue, M. (2003) The tonal system of Skou, New Guinea. In Shigeki Kaji (ed.), *Proceedings of the Symposium Cross-linguistic Studies of Tonal Phenomena: Historical Development, Phonetics of Tone, and Descriptive Studies* 329--365. Tokyo: Tokyo University of Foreign Studies, Research Institute for Language and Cultures of Asia and Africa.

- Downer, G. (1967) Tone-change and tone-shift in White Miao. *Bulletin of the School of Oriental and African Studies* 30(3): 589--599.
- Downer, G. B (1961) Phonology of the word in Highland Yao. *Bulletin of the School of Oriental and African Studies* 24(3): 531--541.
- Edmondson, J., and K. J. Gregerson (1996) Bolyu tone in a Vietic perspective. *Mon-Khmer Studies* 26: 117--133.
- Egerod, S. (1956) *The Lungtu dialect: A Descriptive and Historical Study of a South Chinese Idiom*. Copenhagen: Ejnar Munksgaard.
- Fischer-Jørgensen, E. (1972) PTK et BDG français en position intervocalique accentuée. In Albert Valdman (ed.), *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre* 143--200. The Hague: Mouton.
- Franklin, K. J. (1971) *A Grammar of Kewa, New Guinea*. Pacific Linguistics Series C, No. 16. Canberra: Australian National University.
- Gandour, J. (1974) Consonant types and tone in Siamese. *Journal of Phonetics* 2: 337--350.
- Gandour, J., S. Potisuk, and S. Dechongkit (1994) Tonal coarticulation in Thai. *Journal of Phonetics* 22:477-492.
- Gessner, S. (2003) *The Prosodic System of the Dakelh (Carrier) Language*. Ph. D. dissertation, University of British Columbia.
- Gnanadesikan, A. (1995) Markedness and faithfulness constraints in child phonology. Unpublished ms, Department of Linguistics, University of Massachusetts, Amherst.
- Gomez-Imbert, E., and M. Kenstowicz. (2000) Barasana tone and accent. *International Journal of American Linguistics* 66(4): 4--19.
- Gordon, R. G., Jr. (ed.) (2005) *Ethnologue: Languages of the World*, 15th ed. Dallas: SIL International. <http://www.ethnologue.com>.
- Hale, M., and C. Reiss. (2000) "Substance abuse" and "dysfunctionalism": current trends in phonology. *Linguistic Inquiry* 31(1): 157--169.
- Ham, W. H. (1999) Tone sandhi in Saramaccan: a case of substrate transfer? *Journal of Pidgin and Creole Languages* 14(1): 45--91.
- Han, M. S., and K.-O. Kim (1974) Phonetic variation of Vietnamese tones in disyllabic utterances. *Journal of Phonetics* 2: 223--232.
- Han, M. S., and R. S. Weitzman (1970) Acoustic features of Korean /P T K/, /p t k/, and /ph th kh/. *Phonetica* 22: 112--128.
- Hayes, B. (1999) Phonetically driven phonology: the role of Optimality Theory and inductive grounding. In M. Darnell, E. A. Moravcsik, F. Newmeyer, M. Noonan, and K. M. Wheatley (eds.), *Formalism and Functionalism in Linguistics*, Volume I, 243--285. Amsterdam: Benjamins.
- Hayes, B., R. Kirchner, and D. Steriade (eds.). (2004) *Phonetically-Based Phonology*. Cambridge: Cambridge University Press.
- Healey, A. (1964) *Telefol Phonology*. Linguistic Circle of Canberra Publications, Series B, No. 3. Canberra: Australian National University.
- Hombert, J.-M., J. J. Ohala, and W. G. Ewan (1979) Phonetic explanations for the development of tones. *Language* 55(1). 37--58.
- Hudak, T. J. (1996) William J. Gedney's *The Lue Language: Glossary, Texts, and Translations*. Michigan Papers in South and Southeast Asia, #44. Center for South and Southeast Asian Studies, University of Michigan.
- Hume, E., and K. Johnson (2001) A model of the interplay of speech perception and phonology. In E. Hume and K. Johnson, *The role of speech perception in phonology* 3--26. San Diego: Academic Press.
- Huziwara, K. (2003) Tone sandhi in Chakma and Cak. Handout, 36th International

- Conference on Sino-Tibetan Languages and Linguistics, La Trobe University, Melbourne, November 30, 2003.
- Hyman, L. M. (1976) Phonologization. In Alphonse Juillard (ed.), *Linguistic Studies Offered to Joseph Greenberg: Second Volume: Phonology*. Studia linguistica et philologica 4. Saratoga, California: Anma Libri.
- Hyman, L. M. (2001) The limits of phonetic determinism in phonology: \*NC revisited. In E. Hume and K. Johnson, *The role of speech perception in phonology* 141--186. San Diego: Academic Press
- Jansen, W. (2004). *Laryngeal Contrast and Phonetic Voicing: a Laboratory Phonology Approach to English, Hungarian, and Dutch*. Ph.D. dissertation, Rijksuniversiteit Groningen.
- Jeel, V. (1975) An investigation of the fundamental frequency of vowels after various Danish consonants, in particular stop consonants. *Annual Report of the Institute of Phonetics of the University of Copenhagen* 9: 191--211.
- Kagaya, R. (1974). A fiberscopic and acoustic study of the Korean stops, affricates, and fricatives. *Journal of Phonetics* 2: 161--180.
- Kagaya, R., and H. Hirose (1975) Fiberoptic electromyographic and acoustic analyses of Hindi stop consonants. *Annual Bulletin of the Research Institute for Logopedics and Phoniatrics*. 9. 27--46.
- Kavitskaya, D. (2002) *Compensatory lengthening: phonetics, phonology, diachrony*. New York: Routledge.
- Kawahara, S. (2005) Voicing and geminacy in Japanese: an acoustic and perceptual study. In K. Flack and S. Kawahara (eds.) *University of Massachusetts Occasional Papers in Linguistics* 31, 87--120.
- Kingston, J., and R. L. Diehl (1994) Phonetic knowledge. *Language* 70 (3):419--454.
- Kjellin, O. (1977) Observations on consonant types and "tone" in Tibetan. *Journal of Phonetics* 5: 317--338.
- Kohler, K. J (1982) F<sub>0</sub> in the production of lenis and fortis plosives. *Phonetica* 39(4--5): 199--218.
- Kortlandt, F. H. H (1975) Tones in Wakashan. *Linguistics* 143: 31--34
- Landaburu, J. (1979) La langue des andoke (Amazonie colombienne): grammaire. *Langues et civilisations à tradition orale* 36. Centre National de la Recherche Scientifique.
- Li J. (1996) Bugan, a new Mon-Khmer language of Yunnan Province, China. *Mon-Khmer Studies* 26: 135--160.
- Lin, H.-B. (1988) *Contextual Stability of Taiwanese Tones*. Ph. D. dissertation, University of Connecticut.
- Löfqvist, A (1975) Intrinsic and extrinsic F<sub>0</sub> variations in Swedish tonal accents. *Phonetica* 31: 228--247.
- MacKay, D. J. C. (2003) *Information Theory, Inference, and Learning Algorithms*. Cambridge: Cambridge University Press.
- Macken, M. A. (1995) Phonological acquisition. In J. A. Goldsmith (ed.), *The Handbook of Phonological Theory* 671--696. Cambridge: Blackwell.
- Maddieson, Ian. (1976) A further note on tone and consonants. *UCLA Working Paper in Phonetics* 33: 131--159.
- Malone, T. (2006) Tone and syllable structure in Chimila. *International Journal of American Linguistics* 72(1): 1--58.
- Manuel, S. Y. (1990) The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Journal of the Acoustical Society of America*. 88(3). 1286--1298.
- McCarthy, J. J (1988) Feature Geometry and dependency: a review. *Phonetica* 45: 84--

- Melnar, L. R. (2004) *Caddo Verb Morphology*. University of Nebraska Press.
- Moreton, E. (2007) Phonotactic learning and phonological typology. Unpublished ms, Department of Linguistics, University of North Carolina, Chapel Hill.
- Moreton, E., and S. Amano (1999) Phonotactics in the perception of Japanese vowel length: Evidence for long-distance dependencies. *Proceedings of the 6th European Conference on Speech Communication and Technology, Budapest*.
- Moreton, E., and E. R. Thomas (in press) The origins of Canadian Raising in voiceless-coda effects: a case study in phonologization. To appear In J. S. Cole and J. I. Hualde (eds.), *Papers in Laboratory Phonology 9*.
- Myers, S. (2002) Gaps in factorial typology: The case of voicing in consonant clusters. Unpublished ms, University of Texas at Austin. Retrieved on 19 March 2004 from [uts.cc.utexas.edu/~smyers/voicing.pdf](http://uts.cc.utexas.edu/~smyers/voicing.pdf)
- Nagy, N., and W. T. Reynolds (1996) Accounting for variable word-final deletion within Optimality Theory. In J. Arnold, R. Blake, B. Davidson, S. Schwenter, and J. Solomon (eds.), *Sociolinguistic Variation: Data, Theory, and Analysis* 151--160. Stanford: Center for the Study of Language and Information.
- Newport, E. L., and R. N. Aslin (2004) Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology* 48: 127--162
- Noyer, R. (1991) Tone and stress in the San Mateo dialect of Huave. *Proceedings of the Eastern States Conference on Linguistics (ESCOL)* 277--288.
- Odden, D. (1995) Tone: African languages. In J. A. Goldsmith (ed.), *The Handbook of Phonological Theory* 444--475. Cambridge: Blackwell.
- Odé, C. (2002) *Mpur Prosody: an Experimental-Phonetic Analysis with Examples from Two Versions of the Fenora Myth*. Osaka: ELPR [Endangered Languages of the Pacific Rim] Publications, Series A1-003.
- Ohala, J. J. (1990). The phonetics and phonology of aspects of assimilation. In J. Kingston and M. E. Beckman (eds.), *Papers in laboratory phonology I: between the grammar and physics of speech* 258--275.
- Ohala, J. J. (1994) Hierarchies of environments for sound variation; plus implications for 'neutral' vowels in vowel harmony. *Acta Linguistica Hafniensia* 27: 371--382.
- Ohala, J. J. (2005) Phonetic explanations for sound patterns: implications for grammars of competence. In W. J. Hardcastle and J. M. Beck (eds.), *A Figure of Speech: A Festschrift for John Laver* 23--38. London: Erlbaum.
- Ohde, R. N (1984) Fundamental frequency as an acoustic correlate of stop-consonant voicing. *Journal of the Acoustical Society of America*. 75. 224--230.
- Peng, S.-H. (1997) Production and perception of Taiwanese tones in different tonal and prosodic contexts. *Journal of Phonetics* 25: 371--400.
- Pike, E. V. (1948) Problems in Zapotec tone analysis. *International Journal of American Linguistics* 14(3): 161--170.
- Prince, A. S., and B. Tesar (1999) Learning phonotactic distributions. Retrieved 6 May 2006 from <http://roa.rutgers.edu>.
- Prince, A. S., and P. Smolensky (1993/2004) *Optimality Theory: Constraint Interaction in Generative Grammar*. Unpublished ms, Department of Linguistics, Rutgers University.
- Pulleyblank, D. (2004) Harmony drivers: no disagreement allowed. In *Proceedings of the 28th Meeting of the Berkeley Linguistics Society* 249--267.
- R Development Core Team (2005) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.

- Schuh, R. G. (2002) Class handout for Linguistics 252, Bade/Ngizim Phonology and Morphology. Unpublished ms, Department of Linguistics, UCLA, retrieved 25 August 2005 from [http://www.humnet.ucla.edu/humnet/linguistics/people/schuh/Bade\\_Ngizim/Handout\\_04.pdf](http://www.humnet.ucla.edu/humnet/linguistics/people/schuh/Bade_Ngizim/Handout_04.pdf).
- Serniclaes, W. (1992) *Etude expérimentale de la perception du trait de voisement des occlusives du Français*. Doctoral dissertation, Université Libre de Bruxelles, Institut de Phonétique.
- Shen, X. S. (1990) Tonal coarticulation in Mandarin. *Journal of Phonetics* 18: 281--295.
- Smith, J. L. (2004) Making constraints positional: towards a compositional model of CON. *Lingua* 114(12): 1433--1464.
- Smolensky, P. (1996) The initial state and "richness of the base" in Optimality Theory. Technical Report JHU-CogSci-96-4, Cognitive Science Department, Johns Hopkins University.
- Snyder, W. C., and Lu T. (1997). Wuming Zhuang tone sandhi: a phonological, syntactic, and lexical investigation. In J. A. Edmondson and D. B. Solnit (eds.), *Comparative Kadai: the Tai Branch* 107--139. SIL International.
- Stairs Kreger, G. A., and E. F. Scharfe de Stairs (1981) *Diccionario Huave de San Mateo del Mar*. Serie de Vocabularios y Diccionarios Indígenas "Mariano Silva y Aceres", Num. 24. Instituto Lingüístico de Verano (SIL).
- Steriade, D. (2002) Emergent properties of phonological competence. Handout from a presentation given at the 8th Conference on Laboratory Phonology, Yale University, New Haven, Connecticut, 29 June 2002.
- Svantesson, J.-O. (1989) Tonogenetic mechanisms in northern Mon-Khmer. *Phonetica* 46: 60--79.
- Svantesson, J.-O., and D. House. (2006) Tone production, tone perception, and Kammu tonogenesis. *Phonology* 23(2): 309--333.
- Tersis, N. (1972) *Le Zarma (République du Niger): étude du parler djerma de Dosso*. Société d'études linguistiques et anthropologiques de France, 33--34.
- Vossen, R. (1997) *Die Khoe-Sprachen: ein Beitrag zur Erforschung der Sprachgeschichte Afrikas*. Köln: Rüdiger Köppe.
- Wang J., and Zheng G. (1993) *An Outline Grammar of Mulao*. National Thai Studies Center, Australian National University, Canberra.
- Watkins, L. J. (1984) *A Grammar of Kiowa*. University of Nebraska Press.
- Weber, D., and W. Thiesen. (2001) A synopsis of Bora tone. Work Papers of the Summer Institute of Linguistics, University of North Dakota Session, 45.
- Werner, H. (1997) Die ketische Sprache. *Tungos Sibirica*, Band 3. Wiesbaden: Harrassowitz.
- Wilson, C. (2003) Analytic bias in artificial language learning: consonant harmony vs. random alternation. Handout from a presentation given at the Workshop on Markedness and the Lexicon, Massachusetts Institute of Technology, 25 January 2003.
- Wolff, E. (1983) *A Grammar of the Lamang Language (Gwàz Làmàŋ)*. Glückstadt: Verlag J. J. Augustin GmbH.
- Wright, R. (1996) Tone and accent in Oklahoma Cherokee. In Pamela Munro (ed.), *Cherokee Papers from UCLA* 11--22. Los Angeles: UCLA Department of Linguistics.
- Wright, R., and A. Shyrock (1993) The effect of implosives on pitch in SiSwati. *Journal of the International Phonetic Association* 23(1): 16--23.
- Xu, C. X., and Y. Xu. (2003) Effects of consonant aspiration on Mandarin tones. *Journal*

- of the International Phonetic Association* 33(2): 165--181.
- Xu, Y. (1997) Contextual tonal variations in Mandarin. *Journal of Phonetics* 25: 61--83.
- Yip, M. (1995) Tone in East Asian languages. In J. A. Goldsmith (ed.), *The Handbook of Phonological Theory* 476--494. Cambridge: Blackwell.
- Zee, E. (1980) The effect of aspiration on the  $F_0$  of the following vowel in Cantonese. *UCLA Working Papers in Phonetics* 49: 90--97.
- Zubritskaya, Katya (1997). Mechanism of sound change in Optimality Theory. *Language Variation and Change* 9:121--148.