UNDERPHONOLOGIZATION AND MODULARITY BIAS[1]

Elliott Moreton
University of North Carolina, Chapel Hill

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 (19 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 24 studies of 17 languages). A learning algorithm is proposed in which the learner chooses between 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 (ones that interact with few other constraints). Its application to the tone case is illustrated by simulation.

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. Discussion has centered on a single important typological fact: "phonetic naturalness", the tendency for phonological processes to look like exaggerated versions of subtle phonetic interactions. Two main factors have been identified as potential causes of this long-term bias.

One factor is *pattern selectivity* caused by cognitive biases 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 pattern selectivity as an explanation for typology in general, and for the "naturalness" bias in particular (e.g., Chomsky & Halle 1968:251, 296–297, McCarthy 1988, Prince & Smolensky 1993, Archangeli & Pulleyblank 1994:391–395, Hayes, Kirchner, & Steriade 2004).

The other factor is *phonetic precursor robustness.* The hypothesis is that phonological processes 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 1993, Hale & Reiss 2000, Kavitskaya 2002, Barnes 2002, Blevins 2004:140–151).

An adequate theory of typology will have to recognize diachronic filtering caused by biases in both pattern selectivity and precursor robustness. The simplest theory of their interaction is that the pattern selectivity of UG delimits the cognitively possible grammars ("hard" typology), while precursor robustness determines their frequency ("soft" typology) by skewing sound change (Kiparsky 2004). This paper argues that some soft typological facts are actually due to UG. There are two sides to the argument, one empirical, one theoretical.

---

[1] Several people have contributed to this draft, including Andries Coetzee, Paul de Lacy, Shigeto Kawahara, John McCarthy, Steve Parker, Joe Pater, Jennifer L. Smith, Paul Smolensky, and Anne-Michelle Tessier. It also owes much to Adam Albright and Josh Tenenbaum's 2005 LSA class. Thanks are due to Chris Wiesen for statistical advice. Errors or omissions are solely the author's. Address for correspondence: *moreton@unc.edu*. This draft is current as of May 27, 2006.

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 "under-phonologization", in which two phonetic patterns of similar magnitude correspond to phonological patterns of very different frequency (a point first raised 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 the tone generalization can be accounted for by assigning UG a more active role. 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 process 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 tone-tone interactions outnumber tone-voice interactions. §3 shows that their phonetic precursors are equally robust. §4 presents the learning algorithm and illustrates its application. Discussion is in §5.

2. Phonological typology: Tone-tone outnumbers tone-voice

This section considers two minimally different kinds of phonological process involving tone height: those relating the tone in one syllable to that in the next (phonological tone-tone processes or TTP), and those relating tone to the voicing or aspiration of a neighboring consonant (CTP).

Cases of TTP and CTP 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 (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 CTP. (It is certainly not exhaustive, but that is in the nature of samples.)

The following criteria were used to reduce the chance of gross error: (1) TTP can happen in any tone language, while CTP are possible only in tone languages which also have a voicing contrast. The sample was therefore restricted to tone languages described as having a voicing, aspiration, or fortis-lenis contrast in obstruents. (2) Static phonotactic patterns and morphophonemic alternations qualified, but allophonic alternations did not, since they might be phonetic. (3) Alternations limited to a single affix did not qualify. (4) The survey included only languages described while they were still living languages. (5) As a crude precaution against double-counting cases of shared inheritance or areal spread, I counted language families, defined as top-level categories in Ethnologue, and continents, rather than individual languages. The results are shown in (1) and (2). Italics indicate non-primary sources.

(1) Tone-tone processes: 19 Ethnologue families, 5 continents.
  (a) Africa
    (i) Afro-Asiatic: Gashua Bade: H→L in L_|H, where | is a clitic- or PPh-boundary (Schuh 2002). Voicing contrast.
    (ii) 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.

(iii) 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: Bugan: 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*). Lhasa Tibetan has H-tone spreading in compounds, neutralizing the tone contrast on the second member. (*Duanmu 1992*). 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: Wuming Zhuang: Low-register tones raise before another low-register tone (Snyder & Lu 1997). Voicing contrast; voiced stops preglottalized.

(c) Central and North America

(i) Caddoan: Caddo: High tone spreads left across an intervening sonorant (Chafe 1976:62). Voicing contrast.

(ii) Huavean: Rightward spreading of H tone within phrases (Stairs Kreger 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) Chibchan: Chimila: First of two H tones deletes (Malone 2006). 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) 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 & Thiesen 2001). Fortis-lenis contrast.

(iiv) 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 & Kenstowicz 2000). Voicing contrast.

(e) Oceania

(i) Sko: 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 interactions: 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) Hmong-Mien: Highland Yao: Aspirated initials occur only with higher tones, while unaspirated ones occur with all tones (Downer 1961)
   (ii) Sino-Tibetan: Wuyi: Spreading of high-register tones causes devoicing of intervening voiced obstruent (*Yip 1995:485–487*).
   (iii) Tai-Kadai: Mulao: Aspirated initial stops occur only with lower tones, while unaspirated ones occur with all tones (Wang & Zheng 1993:14).

(c) North America
   (i) Kiowa-Tanoan: Kiowa: Medial voiced stops become voiceless after falling tone (Watkins 1984:40). Blocked in imperfective stems.

(d) Oceania
   (i) Austronesian: Yabem: Voiceless stops are followed by H, voiced ones by L; H/L contrast possible after other consonants, and voice contrast possible word-finally (Capell 1949).
   (ii) Sko: Skou: H/L contrast neutralized to phonetic mid tone in syllables with voiced-obstruent onsets. Voicing contrast before falling tone (Donohue 2003:350–352).

Assuming independence between families, TTP occur significantly more often in this sample than CTP ($p = 0.040$ using Poisson regression), indicating that TTP are either phonologized more often or retained better. This finding is surprising in view of two other facts. First, phonologization can create TTP only in a tone language, whereas CTP can also arise in non-tonal languages undergoing tonogenesis (Svantesson 1989). Second, CTP 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 & Amano 1999, Newport & Aslin 2004, Creel, Newport & Aslin 2004). The next section investigates whether the preponderance of TTP over CTP can be explained by differing precursor robustness.

3. Phonetic typology: Tone-tone is no more robust than tone-voice

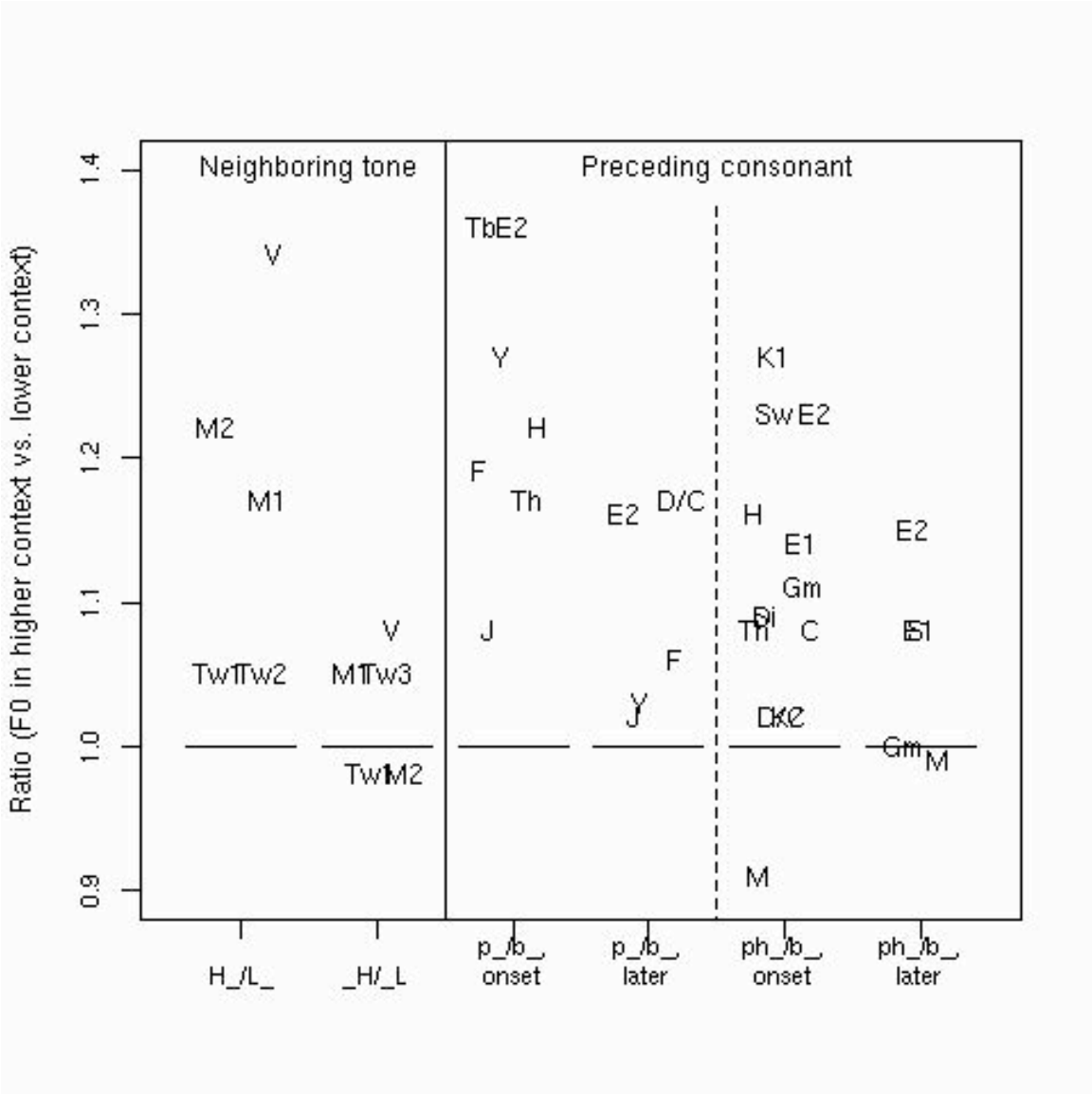
The less phonetic interaction there is between X and Y, the fewer opportunities there are for the listener to misinterpret phonetic covariation as phonological, and hence the less often the X-Y pattern should become phonologized (Ohala 1994, Kavitskaya 2002:123–133, Barnes 2002:151–159, Myers 2002, Blevins 2004:108–109). If this explanation applies to the typology of tone processes, it must be true that, across a wide range of languages, the phonetic influence of tone on neighboring F_0 is substantially greater than that of voicing.

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 I 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, I chose the point closest to the context. If the study used multiple target tones, I computed the effects for each target, then averaged them together within each speaker, and then averaged the speaker effects together.

A similar procedure was followed for interaction between F_0 and voicing or aspiration 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), so the survey included both tonal and non-tonal languages. The voiceless, aspirated, or fortis consonant was the Raising context (Hombert et al. 1979). Two measurement points in the target tone were plotted, the onset of voicing, and a second point at least 40ms later.

Figure (3) 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.

(3) 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.



(4) Key to (3)

(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): 0.98 (Xu 1997, Figures 4 and 7).

(Tw1) 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_: 0.98 (Chang 1988, Tables 6a–e and 7a–e).

(Tw2) 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).

(Tw3) 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: 1.054 (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 & Kim 1974, Figure 2).

(b) Effects of preceding consonant voicing:

(D/C) 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 F0 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).

(F) 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 & Hirose 1975, Table II).

(J) Japanese 3 speakers. Disyllables /kV.CV/, 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).

(D) 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).

(D/C) 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 & Hirose 1975, Table II).

(K1) Korean 2 speakers; 1400 tokens. /ph th kh/ vs. /p t k/. Measured at voice onset: 1.27 (Han & 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: 0.91; 40 ms later, 0.99 (Xu & Xu 2003, Figure 7).

(Si) SiSwati 4 speakers. 8 real-word CVX stimuli, /ph/ vs /ɓ/. Measured at vowel onset: 1.09; 40 ms later, 1.08. (Wright & 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)

There is no evidence that phonetic interaction is greater between two tones than between a tone and obstruent laryngeal features.[2] If anything, the opposite is true. Thus, the different frequency of TTP and CTP is not due to differences in precursor robustness. When vowel height is substituted for tone height, the same pattern emerges: Height-height processes (VVP) are more common than height-voice processes (VCP), but vowel F1 is less affected by other vowels than by consonant voicing (Moreton, in preparation).

[2] The effect of tonal anticipatory coarticulation is obviously no larger than that of voicing or aspiration. That of tonal carryover articulation at target-vowel onset is not significantly different from that of voicing (p = 0.600) or aspiration (p = 0.299, linear-mixed effects model with Language as a random effect).

4. Modularity bias via Bayesian Constraint Addition

With precursor robustness eliminated as an explanation, it is the turn of pattern selectivity. Why are CT and VC patterns underphonologized relative to TT and VV patterns? It will not do to say that TT and VV 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" pattern selectivity, 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 by Universal Grammar 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:

(5)

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 (5) 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 & 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/. I assume that both tone and voice are privative (H and [voice]). For tone spreading, I adapt the analysis of vowel harmony proposed by Pulleyblank (2004):

(6) Constraints for tone spreading (after Pulleyblank 2004)

(a) MAX-H: "Don't delete an H tone." *= underlyingly H-toned vowel with no H-toned surface correspondent.

(b) *LH: "No LH tone sequences." * = a H tone in the next syllable following an L tone.

These constraints allow only two grammars, distinguished by their effect on /LH/:

(7)

(a) Faithful realization

/LH/	MAX-H	*LH
→ LH		*
LL	*!	

(b) Righward L-tone spreading.

b. /LH/	*LH	MAX-H
LH	*!	
→ LL		*

I assume analogous constraints are 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	*!		
→ bL			*
pH		*!	

(c) Pre-/H/ devoicing

/bH/	*bH	MAX-H	MAX-VOICE
bH	*!		
bL		*	
→ 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 (Nagy & Reynolds 1997, Boersma 1998:269–273, Boersma & 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)

$$\frac{P(H_1 \mid D)}{P(H_0 \mid D)} = \frac{P(D \mid H_1)P(H_1) / P(D)}{P(D \mid H_0)P(H_0) / P(D)}$$
$$= \frac{P(D \mid H_1)P(H_1)}{P(D \mid H_0)P(H_0)}$$

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, I assume that the Learner is initially unbiased, so that the ratio is 1.

When the data arrives, that estimate is multiplied by P($D \mid H_1$)/P($D \mid H_0$), 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 H_0 the true grammar, and the "pattern" merely a statistical fluke? Or is H_1 the true grammar, 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* grammars H_1 allows. If the pattern is a one-in-a-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 grammars H_1 allows, the smaller P($D \mid H_1$), and hence, by (10), the smaller P($H_1 \mid D$). The 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) grammar contains two types of constraint, *m* involving only tone, and *n* involving only segments. The grammar 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 constraint 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)!*.[3] Adding a non-modular constraint causes a steeper increase in the number of different grammars, reduces P($D \mid H_1$) more, and so incurs a stronger penalty from the Razor.

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

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 (11):

(11) Corpus frequencies, actual and predicted. See text below for explanation of p, q, s.

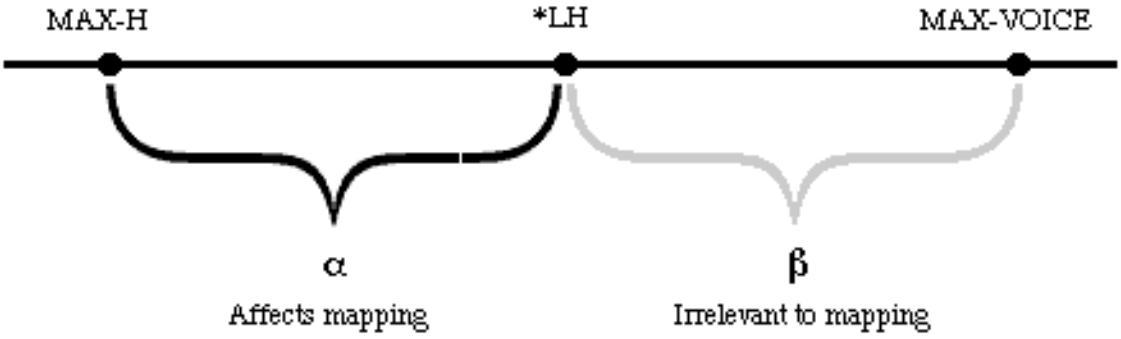
		D	Predicted by Learner's hypotheses		
			H_0: *No change*	H_1: *New constraint needed*	
Word type	Intended by Speaker	Perceived by Learner		LH condition	bH condition
HH, pH	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}(1+s_{a,\beta})$
HL, pL	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
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_1 differs depending on condition.

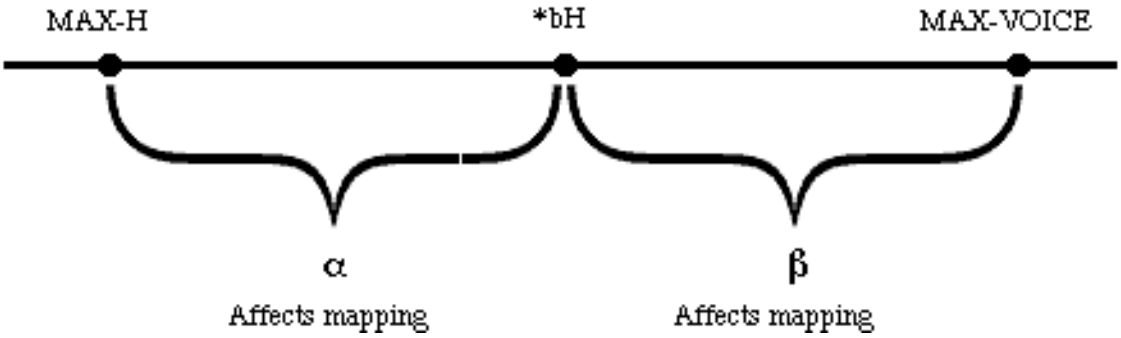
In the LH condition, H_1 says that the constraint set is {MAX-VOICE, MAX-H, *LH}. Under H_1(LH) the grammar has one genuinely adjustable parameter, α, the ranking distance between the *LH and MAX-H. Changing α changes p_α = P(*LH >> MAX-H | H_1(LH), α). The distance β between *LH and MAX-VOICE can also be adjusted, but that has no effect on the output of the grammar, because the two constraints do not interact. By setting α so that $p_\alpha = r$, the H_1(LH) grammar can be made to match D, while the H_0(LH) grammar cannot. This is shown in (12a).

(12) H_1: Nominally- versus actually-adjustable parameters.

(a) L-H condition: Two nominal parameters, but only one actual parameter.



(b) bH condition: Two actual parameters.



In the bH condition, H_1 says that the constraint set is {MAX-VOICE, MAX-H, *bH}. Unlike H_1(LH), H_1(bH) gives the grammar two genuinely adjustable parameters (12b). Both α and β 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 (11). Matching the Speaker's data requires setting α and β so that $q_{\alpha,\beta} = r$ and $s_{\alpha,\beta} = 0$. Again, H_0(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_1(LH) and H_1(bH) can accommodate the data, but H_1(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_1(bH) than by H_1(LH). As a result, H_1(LH) is a better alternative to H_0(LH) than H_1(bH) is to H_0(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 α and β. I follow Boersma (1998, Boersma & 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_1, then, we can

thus 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 α and β and standard deviation of 1. (For H_0, the rankings are irrelevant.) The Learner initially assumes that all rankings are equally probable (i.e., a uniform prior distribution for α and β, reflecting the Learner's 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 (11) shows that, for H_0(LH) or H_0(bH), the probability is always 1/4. Hence, if there are *N* tokens in *D*,

(13)

$$P(D \mid H_0(LH)) = P(D \mid H_0(bH)) = \left(\frac{1}{4}\right)^N$$

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 α,

(14)

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

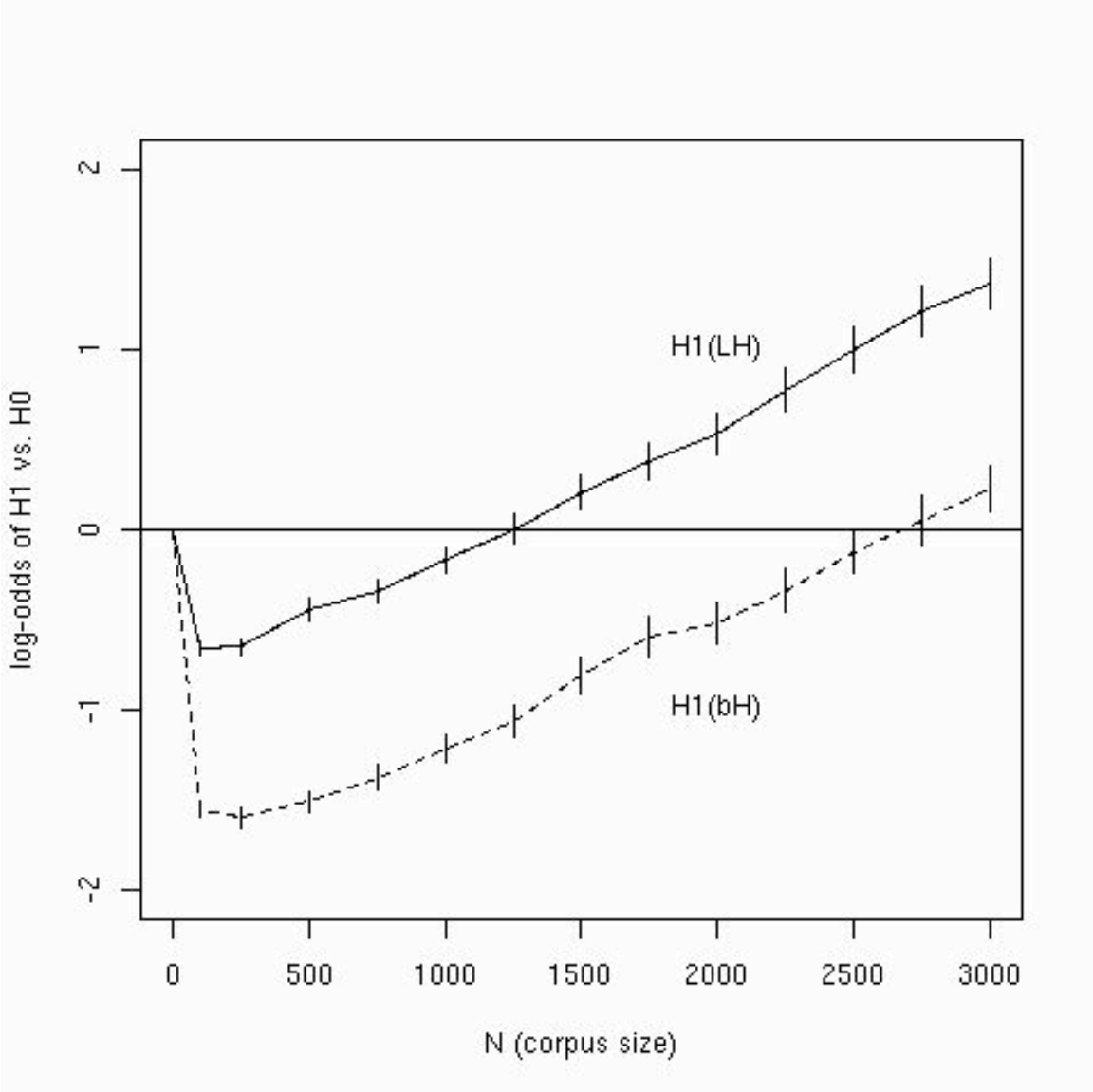
The probability of *D* under H_1(LH) is obtained by integrating (14), times the probability density of each α, over all α. Likewise, in the bH condition, for given α and β,

(15)

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

with P(D| H_0(bH)) obtained by integrating (15), times the probability density of each α and β, over all α and β. For practical reasons, the integrals were approximated by discretizing α and β as a grid of points spaced 0.1 units apart in the region [–4, 4]×[–4, 4]. These limits allow p_α, $q_{\alpha,\beta}$, and $s_{\alpha,\beta}$ to vary from < 0.0001 to > 0.9999. Random corpora *D* of size *N* = 0, 250, 500, ... 3000 were simulated using R (R Development Core Team, 2004). The resulting likelihood ratios were logit-transformed and plotted in (16).

(16) Simulation results. Solid: LH. Dashed: 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.



Consider first the LH condition (solid curve). 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.) The more data, the clearer it becomes 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 declines. 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 equipoised 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 α and β 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 process. 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 process.

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_1(LH) and H_1(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. Now suppose that this generation also retains the phonetic precursor. Then their speech contains even fewer 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.[4]

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 & 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 & 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 & Smolensky 1993:101). 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."

[4] 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 & Thomas in press), but could also be due to growth in the precursor itself.

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 (Kiparsky 2004) 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 & 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* pattern selectivity. The issue will only be settled in the lab. Some evidence argues for pattern selectivity: English-speaking adults trained on non-coarticulated $C_1V_1C_2V_2$ stimuli in one artificial-language paradigm learned to recognize height dependencies between V_1 and V_2 better than height-voice dependencies between V_1 and C_2 (Moreton, in preparation).

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.

References

Ansre, Gilbert. 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, Arto. 1995. Deriving variation from grammar: a study of Finnish genitives. MS, Rutgers Optimality Archive.

Anttila, Arto, and Young-Mee Cho. 2004. Variation and change in Optimality Theory. In.: John J. McCarthy (ed.), Optimality Theory in Phonology, 569–580. Malden: Blackwell.

Archangeli, Diana, and Douglas Pulleyblank. 1994. Grounded Phonology. MIT Press.

Barnes, Jonathan. 2002. Positional neutralization: a phonologization approach to typological patterns. Ph.D. dissertation, University of California, Berkeley.

Baumbach, E. J. M. 1987. Analytical Tsonga Grammar. Pretoria: University of South Africa.

Boersma, Paul. 1998. Functional Phonology: Formalizing the interactions between articulatory and perceptual drives. Ph. D. dissertation, University of Amsterdam..

Boersma, Paul, and Bruce Hayes. 2001. Empirical test of the Gradual Learning Algorithm. Linguistic Inquiry 32(1):45–86.

Bradshaw, Mary M. 1999. A crosslinguistic study of consonant-tone interaction. Ph. D. dissertation, Ohio State University.

Capell, A. 1949. Two tonal languages of New Guinea. Bulletin of the School for Oriental and African Studies 8: 184-199.

Chang, Yueh-chin. 1988. Sandhi tonal des syntagmes dissyllabiques du Min-Nan parlé à Taiwan. Cahier de Linguistique Asie Orientale 17(2):193-234.

Chen, Matthew. 2000. Tone sandhi. Cambridge: Cambridge University Press.

Chomsky, Noam, and Morris A. Halle. 1968. The Sound Pattern of English. Cambridge: MIT Press.

Coetzee, Andries. 2002. Between-language frequency effects in phonological theory. MS, University of Massachusetts, Amherst.

Connell, Bruce. 2002. Tone languages and the universality of intrinsic F0: evidence from Africa. Journal of Phonetics 30:101-129.

Creel, S.C., Newport, E.L., & Aslin, R.N. 2004. Distant melodies: Statistical learning of non-adjacent dependencies in tone sequences. Journal of Experimental Psychology: Learning, Memory, and Cognition, 30, 1119-1130.

Donohue, Mark. 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. B. 1961. Phonology of the word in Highland Yao. Bulletin of the School of Oriental and African Studies 24(3):531-541.

Downer, Gordon. 1967. Tone-change and tone-shift in White Miao. Bulletin of the School of Oriental and African Studies 30(3):589-599.

Duanmu, San. 1992. An autosegmental analysis of tone in four Tibetan languages. Linguistics of the Tibeto-Burman Area 15(1):65–91.

Egerod, Søren. 1956. The Lungtu dialect: a descriptive and historical study of a South Chinese idiom. Copenhagen: Ejnar Munksgaard.

Gandour, Jack. 1974. Consonant types and tone in Siamese. Journal of Phonetics 2:337-350.

Gessner, Suzanne. 2003. The prosodic system of the Dakelh (Carrier) language. Ph. D. dissertation, University of British Columbia.

Gnanadesika, Amalia. 1995, Markedness and faithfulness constraints in child phonology. MS, Department of Linguistics, University of Massachusetts, Amherst.

Gomez-Imbert, Elsa, and Michael Kenstowicz. 2000. Barasana tone and accent. International

Journal of American Linguistics 66(4):4-19.
Gordon, Raymond G., Jr. (ed.) 2005. Ethnologue: Languages of the World, 15th ed. Dallas: SIL International. http://www.ethnologue.com.
Hale, Mark, and Charles Reiss. 2000. "Substance abuse" and "dysfunctionalism": current trends in phonology. Linguistic Inquiry 31(1):157–169.
Ham, William 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 R. S. Weitzman. 1970. Acoustic features of Korean /P T K/, /p t k/, and /ph th kh/. Phonetica 22:112-128.
Han, Mieko S., and Kong-On Kim. 1974. Phonetic variation of Vietnamese tones in disyllabic utterances. Journal of Phonetics 2:223-232.
Hayes, Bruce. 1999. Phonetically driven phonology: the role of Optimality Theory and inductive grounding. In: Michael Darnell, Edith A. Moravcsik, Frederick NewMeyer, Michael Noonan, and Kathleen M. Wheatley (eds.), Formalism and Functionalism in Linguistics, Volume I, 243–285. Amsterdam: Benjamins.
Hayes, Bruce, Robert Kirchner, and Donca Steriade (eds.). 2004. Phonetically-based phonology. Cambridge: Cambridge University Press.
Hombert, Jean-Marie, John J. Ohala, and William G. Ewan. 1979. Phonetic explanations for the development of tones. Language. 55(1). 37-58.
Huziwara, Keisuke. 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, Larry M. 1976. Phonologization. In: Alphonse Juilland (ed.), Linguistic Studies Offered to Joseph Greenberg: Second Volume: Phonology. Studia linguistica et philologica 4. Saratoga, California: Anma Libri.
Jeel, Vivi. 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, Ryohei, and Hajime Hirose. 1975. Fiberoptic electromyographic and acoustic analyses of Hindi stop consonants. Annual Bulletin of the Research Institute for Logopedics and Phoniatrics. 9. 27-46.
Kagaya, Ryohei. 1974. A fiberscopic and acoustic study of the Korean stops, affricates, and fricatives. Journal of Phonetics 2:161-180.
Kavitskaya, Darya. 2002. Compensatory lengthening: phonetics, phonology, diachrony. New York: Routledge.
Kawahara, Shigeto. 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.
Kiparsky, Paul. 2004. Universals constrain change; change results in typological generalizations. MS, Stanford University. Accessed on August 27, 2005, from http://www.stanford.edu/~kiparsky/Papers/cornell.pdf.
Kjellin, Olle. 1977. Observations on consonant types and "tone" in Tibetan. Journal of Phonetics 5:317-338.
Kohler, K. J. 1982. F_0 in the production of lenis and fortis plosives. Phonetica 39(4-5):199-218.
Li Jinfang. 1996. Bugan, a new Mon-Khmer language of Yunnan Province, China. Mon-Khmer Studies 26: 135-60.
Lin, Hwei-Bing. 1988. Contextual stability of Taiwanese tones. Ph. D. dissertation, University of Connecticut.
Lofqvist, A. 1975. Intrinsic and extrinsic F_0 variations in Swedish tonal accents. Phonetica 31:228-247.
Malone, Terry. 2006. Tone and syllable structure in Chimila. International Journal of American

Linguistics 72(1):1-58.
MacKay, David J. C. 2003. Information theory, inference, and learning algorithms. Cambridge: Cambridge University Press.
Macken, Marlys A. 1995. Phonological acquisition. In: John A. Goldsmith (ed.), The handbook of phonological theory, 671–696. Cambridge: Blackwell.
Manuel, Sharon 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, John J. 1988. Feature Geometry and dependency: a review. Phonetica 45:84–188.
Moreton, Elliott, and Shigeaki 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, Elliott, and Erik R. Thomas. In press. The origins of Canadian Raising in voiceless-coda effects: a case study in phonologization. To appear in: Jennifer S. Cole and José Ignacio Hualde (eds.), Papers in Laboratory Phonology 9.
Moreton, Elliott (in preparation). Phonological typology and phonotactic learning. Manuscript, Department of Linguistics, University of North Carolina, Chapel Hill.
Myers, Scott. 2002. Gaps in factorial typology: The case of voicing in consonant clusters. MS, University of Texas at Austin.
Nagy, Naomi, and William T. Reynolds. 1995. Accounting for variable word-final deletion within Optimality Theory. In: J. Arnold, R. Blake, B. Davidson, S. Schwenter, and J. Solomon (eds.), Sociolinguistic variation: theory, data, and analysis, 151–160. Stanford: CSLI.
Newport, E.L., & Aslin, R.N. 2004. Learning at a distance: I. Statistical learning of non-adjacent dependencies. Cognitive Psychology, 48, 127-162
Noyer, Rolf. 1991. Tone and stress in the San Mateo dialect of Huave. ESCOL 1991 proceedings, 277–288.
Odden, David. 1995. Tone: African languages. In: John A. Goldsmith (ed.), Handbook of Phonological Theory. Cambridge: Blackwell.
Ohala, John J. 1993. The phonetics of sound change. In: Charles Jones (ed.), Historical Linguistics: Problems and Perspectives, 237-278. Harlow: Longman.
Ohala, John J. 1994. Hierarchies of environments for sound variation; plus implications for 'neutral' vowels in vowel harmony. Acta Linguistica Hafniensia 27:371–382.
Ohde, R. N. 1984. Fundamental frequency as an acoustic correlate of stop-consonant voicing. Journal of the Acoustical Society of America. 75. 224-230.
Pearce, Mary. 2003. Vowel harmony in Kera (Chadic). M.A. thesis, Department of Phonetics and Linguistics, University College London.
Peng, Shui-hui. 1997. Production and perception of Taiwanese tones in different tonal and prosodic contexts. Journal of Phonetics 25:371-400.
Pike, Eunice Victoria. 1948. Problems in Zapotec tone analysis. International Journal of American Linguistics 14(3):161-170.
Prince, Alan S., and Paul Smolensky. 1993. Optimality Theory: Constraint Interaction in Generative Grammar. MS, Department of Linguistics, Rutgers University.
Prince, Alan S., and Bruce Tesar. 1999. Learning phonotactic distributions. MS, Rutgers Optimality Archive, ROA-353.
Pulleyblank, Douglas. 2004. Harmony drivers: no disagreement allowed. In: Proceedings of the 28th Meeting of the Berkeley Linguistics Society, 249–267.
Schuh, Russell G. 2002. Class handout for Linguistics 252, Bade/Ngizim Phonology and Morphology. MS, Department of Linguistics, UCLA, accessed from http://www.humnet.ucla.edu/humnet/linguistics/people/schuh/Bade_Ngizim/Handout_04.pdf on August 25, 2005.

Serniclaes, Willy. 1992. Etude expérimentale de la perception du trait de voisement des occlusives du Français. Doctoral dissertation, Universite Libre de Bruxelles, Institut de Phonétique.

Shen, Xiaonan Susan. 1990. Tonal coarticulation in Mandarin. Journal of Phonetics 18:281-295.

Smith, Jennifer L. 2004. Making constraints positional: towards a compositional model of CON. Lingua 114(12):1433–1464.

Smolensky, Paul. 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, Wil C., and Lu Tianqiao. 1997. Wuming Zhuang tone sadhi: a phonological, syntactic, and lexical investigation. In: Jerold A. Edmondson and DavidB. Solnit (eds.), Comparative Kadai: the Tai branch, 107-139. SIL International.

Stairs Kreger, Glenn Albert, and Emily Florence Scharfe de Stairs. 1981. Diccionario Huave de San Mateo del Mar. Serie de Vocabularios y Diccionarios Indegenas "Mariano Silva y Aceres", Num. 24. Instituto Lingüístico de Verano (SIL).

Svantesson, Jan-Olof. 1989. Tonogenetic mechanisms in northern Mon -Khmer. Phonetica. 46:60-79.

Watkins, Laurel J. 1984. A grammar of Kiowa. University of Nebraska Press.

Weber, David, and Wesley Thiesen. 2001. A synopsis of Bora tone. Work Papers of the Summer Institute of Linguistics, University of North Dakota Session, 45.

Wolf, Ekkehard. 1983. A grammar of the Lamang language (Gwàd^ LàmàN). Glückstadt: Verlag J. J. Augustin GmbH.

Wright, Richard, and Aaron Shyrock. 1993. The effect of implosives on pitch in SiSwati. Journal of the International Phonetic Association 23(1):16-23.

Wright, Richard. 1996. Tone and accent in Oklahoma Cherokee. In: Pamela Munro (ed.), Cherokee Papers from UCLA, 11-22. Los Angeles: UCLA Department of Linguistics.

Xu, Ching X., and Yi Xu. 2003. Effects of consonant aspiration on Mandarin tones. Journal of the International Phonetic Association 33(2):165-181.

Xu, Yi. 1997. Contextual tonal variations in Mandarin. Journal of Phonetics 25:61-83.

Yip, Moira. 1995. Tone in East Asian languages. In: John A. Goldsmith (ed.), Handbook of Phonological Theory. Cambridge: Blackwell.

Zee, Eric. 1980. The effect of aspiration on the F_0 of the following vowel in Cantonese. UCLA Working Papers in Phonetics 49:90-97.