Contrasts and contexts in palatalization

Lucien Carroll

January 4, 2012

1 Introduction

According to Pandit (1954), three coronal fricatives (dental/alveolar /s/, palato-alveolar /ʃ/ and retroflex /ʂ/) in Old Indo-Aryan (classical Sanskrit and contemporary vernaculars) neutralized to a single phoneme in most Middle Indo-Aryan languages, and the orthography of Old Gujarati literature (before ca. 1500) suggests this was the case in Gujarati, producing /s/ only.[1] However, later influence from Sanskrit reintroduced a contrast between /s/ and /ʃ/. Followed by /i, e/, Sanskrit /ʃ, ʂ/ correspond to /ʃ/ in the Gujarati data (1a), while before the other vowels they were still adapted as /s/ (1b).

(1) a) /**ʃ**ikh/ 'advice' (cf Sanskrit /ʃiksaː/)
/**ʃ**eẓaː/ 'nose mucus' (cf Sanskrit /ʃleɳmaː/)
/dʒo**ʃ**iː/ 'astrologer' (cf Sanskrit /dʒjotiʂikah/)

b) /**s**uːḍlo/ 'basket' (cf Sanskrit /ʃuɳḍaː/)
/aːr**s**o/ 'mirror' (cf Sanskrit /aːdarʃah/)
/**s**amniː/ 'small wedge' (cf Sanskrit /ʃamjaː/)
/maː**s**o/ 'unit of weight' (cf Sanskrit /maːʂa/

This new pattern was not simply turning all the coronal fricatives into /ʃ/ before /i, e/, since words that originated with a /s/ in Sanskrit remain as such in Gujarati, regardless of the following vowel.

(2) /**s**indur/ 'vermillion' (cf Sanskrit /sinduram/)
/**s**evvuː/ 'to serve' (cf Sanskrit /sevate/)
/**s**amaːrvuː/ 'to repair' (cf Sanskrit /samaːrakajati/

[1] Here and throughout, I have adapted original transcriptions to standard IPA, though the sources are not always clear about the intended phonetic realization.

As a result of these changes, this dialect of Gujarati has many words with /ʃi/ and /ʃe/, and /s/ can be followed by any of the vowels, but there are few or no words with /ʃ/ followed by /a, o, u/.

The adaptation of each sound in each context is unremarkable on its own. When adapting loanwords, speakers do imitate the sounds of the source language, which would be easy when the Sanskrit word had an /s/. The retroflex /ʂ/ sounds similar to /ʃ/ and is articulated similarly, so they might easily be adapted as the same sound. And the vowels /i, e/ have a certain similarity to /ʃ/ that could make /ʃ/ easier in those contexts. On the other hand, a speaker unaccustomed to producing a particular sound (e.g. /ʂ/ for speakers of English) can find it quite difficult, so simplifying /ʃ, ʂ/ to a familiar /s/ is also understandable. But patterns of this type are actually typologically unusual. As I will show, it is quite unusual for a language to distinguish categories among coronal fricatives before front vowels like /i, e/ and yet not maintain that contrast before back vowels. Among languages that have more than one coronal fricative but don't contrast them in specific vocalic contexts, the overwhelming majority have the contrast everywhere *except* next to front vowels, especially /i/. In other words, many languages, including English, *palatalize* /s/ to /ʃ/ (or similar sounds) next to /i/ (e.g. 'bless you'), but very few languages could be described as having both /ʃi/ and /si/ sounds but depalatalizing /ʃ/ to /s/ next to /o/ or /a/.

How is it that sound patterns like the Gujarati fricatives are possible and yet uncommon? In order to show how it is possible, Hall (1995) develops a formal analysis that attempts to explain how the creation of this sound pattern in Gujarati is a natural consequence of the articulatory similarities and differences among the sounds. Though the analysis explains why /ʃ/ did not convert to /s/ next to /i/, the non-palatalization of /s/ in that context appears arbitrary. On the other hand, if the different aspects of the pattern are indeed natural, we would naturally expect the resulting patterns to occur relatively often. Another possibility is that the Gujarati pattern is an example of a 'crazy rule' (Bach & Harms, 1972), the result of an unlikely coincidence of a sequence of natural phonetic changes, helped along by human abilities to generalize (even hallucinate) patterns in variable complex data (Morley, 2008; Hayes, Zuraw, Siptár, & Londe, 2009; Hudson Kam & Newport, 2009). And because the Gujarati pattern consists of a particular balance among different factors (imitate the source, produce what is easy, adapt to the context), a third possibility exists, that the Gujarati pattern is the result of the simultaneous action of a few natural phonetic pressures. Both of the simultaneous explanations permit the existence of patterns like the Gujarati fricatives, but if either of these explanations creates the pattern too easily, than we would expect the pattern often. There is plenty of evidence that these 'unnatural' patterns are possible, and

collectively they are not uncommon, but individually they are each unusual (Mielke, 2008; Blevins, 2008), so the 'crazy rule' explanation could allow the Gujarati fricatives without making them common. In contrast, the simultaneous action of a few natural pressures is too good of an explanation. The phonetic details of language are deeply variable, influenced by everything from low-level aerodynamics and acoustics of the intended utterances to the social roles of the interlocutors and previous experience with the particular words and phrases in the utterance, and these can have simultaneous additive effects (Hume & Johnson, 2001a; Eckert & Rickford, 2001; Pierrehumbert, 2002; Tily et al., 2009). As I will show, if the weighted multi-factorial variation evidenced in phonetic detail were directly carried over to broad phonological patterns, patterns like the Gujarati fricatives would be typologically common. Wedel (2007) shows how the variation seen in phonetic detail is shaped by language change rather than directly carried over, but the explanation does not answer how patterns like the Gujarati fricatives could be rare and yet possible.

The status of patterns like the Gujarati fricatives have played an important role in development of constraint-based models of language typology. Optimality Theory (Prince & Smolensky, 1993/2004) attempts to describe the limits of cross-linguistic variation in terms of a formal computational system acting on linguistic representations, in the tradition of generative grammar. However, many of its innovations, including parallel evaluation of outputs and the interaction of multiple constraints, are carried over from its connectionist-based predecessor, Harmonic Grammar (Legendre, Miyata, & Smolensky, 1990). As originally conceived, Optimality Theory and Harmonic Grammar were levels of description (in the sense that the more abstract 'macrotheory' approximates the behavior of the more detailed 'microtheory') in a larger program of relating subsymbolic and symbolic cognition (Smolensky, 1988; Smolensky & Legendre, 2006). The primary innovation of Optimality Theory was the categorical ranking of constraints, rejecting the weighted system of Harmonic Grammar, in which optimization is based on the summed costs of constraint violations. Optimality Theory interprets constraint ranking typologically, so that exhaustive reranking of constraints is taken to reflect possible language patterns. The weight-based system of Harmonic Grammar allows certain language patterns (like the Gujarati fricatives) which were recognized as typologically strange, and constraint ranking eliminates the possibility of these cumulative-type patterns (Legendre, Sorace, & Smolensky, 2006). As a result of the typological motivations for constraint interaction and for the constraints themselves, Optimality Theory is ambiguously a model of cognition and a model of typology, a puzzle that has prompted explorations of how they could be intimately related (Hayes, Kirchner, & Steriade, 2004) as well as recommendations that they should be reexamined separately (Blevins, 2004; de Lacy, 2010).

However, issues with the status of cumulative patterns in typological predictions persisted. Early on, a formally similar set of language patterns were readmitted into Optimality Theory's predictions via the mechanism of local constraint conjunction (Smolensky, 1995). A variety of proposals on how to restrict local conjunction were suggested (e.g. Bakovic, 2000; Lubowicz, 2005) but no consensus emerged. Other proposals called for weighted constraints in a phonetics module only (Zsiga, 2000), in an integrated phonetic/phonological model (Flemming, 2001; Gafos & Benus, 2006), or in syntax (Sorace & Keller, 2005; Jäger & Rosenbach, 2006). And recently, Harmonic Grammar has received new interest in phonology, in both learning and typology (e.g. Hayes & Wilson, 2008; Potts, Pater, Jesney, Bhatt, & Becker, 2010). Considering that even within Optimality Theory oriented research, issues of learnability and language change have been acknowledged as also shaping typology (e.g. Myers, 2002; M. Gordon, 2007; Alderete, 2008), this study pursues the hypothesis that the observability of different cumulative patterns is better understood by more directly modeling language change.

This paper focuses on the typology of coronal palatalization patterns, a relatively simple case that shows that some cumulative patterns are highly underrepresented. After first motivating some preliminary assumptions, I compare the coronal palatalization contrasts in a sample of 198 languages to the type probabilities expected under naive models of language generation. Finally, I summarize a few proposals which attempt to explain the shortage of cumulativity, concluding that these proposals are inadequate but suggestive of the kind of explanation required.

2 Patterns of coronal palatalization

A wide variety of related but distinct phonological alternations are referred to as palatalization (Kochetov, 2010; Bateman, 2011). Palatalization prototypically involves either a change in primary place of articulation towards the palatal region ('full palatalization', e.g. /s/ → [ɕ] in (3a)) or addition of a secondary palatal articulation ('secondary palatalization', e.g. /s/ → [sʲ] in (3b)).

(3) a) Polish full palatalization (Rubach, 2003)
[gwɔ**ɕ**itɕ] INF verb of [gwɔs] 'voice' (noun)
[bra**tɕ**iʃɛk] DIMIN of [brat] 'brother'

b) Ukrainian secondary palatalization (Rubach, 2006)
[ly**sʲ**iv] GEN PL of [lys] 'fox' NOM SG
[vɔ**dʲ**i] DAT of [vɔda] 'water' NOM

Many examples of stop palatalization also include addition of fricative release ('assibilation' e.g. /t/ → [tʃ]) or conversion to a fricative ('spirantization' e.g. /k/ → [s]), and in some treatments, alternations that consist only of assibilation or spirantization, without a change in place of articulation, are also included under the rubric of palatalization. Palatalization can be triggered by assimilation to neighboring consonants or by morphological marking, but the prototypical palatalization alternations are triggered by vocoids, especially high front 'palatal' ones (e.g. [i, j]). Non-high front vowels (e.g. [e]) or non-front high vowels (e.g. [u]) trigger palatalization in some cases, but only if high front vowels also trigger it.

This section summarizes a typological survey of palatalization alternations and associated phonemic contrasts. I focus here on palatalizations of coronals that involve a change in place of articulation (primary or secondary) triggered by vocoids, and I distinguish the palatalization of fricatives from the palatalization of plosives (i.e. oral stops and affricates) in order to reduce some of the complexity. Coronal palatalization at least does not involve major changes in place of articulation (e.g. dorsal to coronal /k/ → [tʃ]), and the assibilation and spirantization are not issues in fricative palatalization. Given the fact that phonological alternations do not happen in isolation but occur as part of a system, it is important to remember that palatalization of fricatives frequently cooccurs with palatalization of plosives, and palatalization of coronals often cooccurs with palatalization of dorsals or labials (Bateman, 2007), but in the interest of perspicuity, these complications are not directly dealt with here.

2.1 Contrast and neutralization

Both articulatorily and acoustically, coronal fricatives lie in a continuous space. Fricatives are distinguished from other speech sounds by the noise produced through turbulent air flow at the point where the oral channel is constricted. Among coronal fricatives, the primary dimension of acoustic variation is the frequency of the spectral noise, which is associated with the distance between the point of frication and the front of the mouth, because the passageway forms a resonant cavity which filters the noise spectrum (Stevens, 1989; Zygis & Hamann, 2003; Hamann & Avelino, 2007). Additional factors include the peaks within the band of noise and formant transitions from and to neighboring vowels, which are associated with the orientation of the tongue tip, presence and size of a gap under the tongue tip, and the proximity of the tongue body to the palate (Stevens, 1989; M. Gordon, Barthmeier, & Sands, 2002). Within this continuous space, there is of course a great deal of variability, influenced by many factors in addition to phonetic context, but a systematic alternation dependent on phonetic context can result in the neutralization or near-neutralization of

otherwise contrastive sounds. In palatalization, tokens of the affected phoneme have a noticeably different phonetic distribution in the palatalizing context than in other contexts, and in some cases, this creates situations where phonemes are distinguishable in one context but indistinguishable in another.

For example, American English has palatalization alternations in two kinds of contexts, one at word boundaries and the other word-internally, and the word-internal palatalization is neutralizing, while the palatalization at word boundaries is non-neutralizing (Zsiga, 1995). According to the standard generative analysis (Halle & Mohanan, 1985), the word-internal palatalization is conditioned by an initial underlying /j/ in certain derivational suffixes, converting underlying /s, z/ to palato-alveolar /ʃ, ʒ/ at the morpheme boundary.[2] In the following examples, palato-alveolar fricatives appear word-internally, at the juncture of a stem that ends in an alveolar fricative and a suffix that (at least in other words) begins with /j/.

(4)	/kənfɛ**s**/	/kənfɛ**ʃ**ən/	cf	/ɹəbɛl/	/ɹəbɛl**j**ən/
	confess	confession		rebel	rebellion
	/ɹɪvaj**z**/	/ɹɪvɪ**ʒ**ən/			
	revise	revision			
	/pɹɛ**s**/	/pɹɛ**ʃ**ɚ/	cf	/fejl/	/fejl**j**ɚ/
	press	pressure		fail	failure
	/si**z**/	/si**ʒ**ɚ/			
	seize	seizure			
	/ɹej**s**/	/ɹej**ʃ**əl/	cf	/pɹɑvɚb/	/pɹɑvɚb**j**əl/
	race	racial		proverb	proverbial
	/spej**s**/	/spej**ʃ**əs/	cf	/vɪktɚ/	/vɪktɔɹ**j**əs/
	space	spacious		victor	victorious

Zsiga (1995) found that instances of /ʃ/ in *confession* and *impression* were indistinguishable from instances of /ʃ/ in *mesh* and *fresh*, which have no associated alternation. Zsiga mentions (but does not discuss) the possibility that the /ʃ/ tokens of word-internal palatalization are simply palato-alveolar underlyingly. Even if it turns out that these alternations have limited psychological reality, note that they satisfy a phonotactic regularity that is quite general and productive: in contrast to other places of articulation (e.g. *few*, *view*, *hue*), very few words in American English have [s] or [z] before [j] within a syllable, and speakers avoid creating

[2]Neither Zsiga (1995) nor Halle and Mohanan (1985) provide transcriptions, and I am suspicious that my own pronunciations differ on some of their example words. The transcriptions of these words were adapted from the CMU Pronouncing Dictionary (`http://www.speech.cs.cmu.edu/cgi-bin/cmudict`).

that structure in loan word adaptation, e.g. /əˌfɪ.**ʃə**ˈna.do/ 'aficionado' (< Spanish /sjo/), /ɪˈglej.**si**.əs/ 'Iglesias' (< Spanish /sjas/).[3] Palatalization also occurs at word boundaries when the following word starts with /j/, particularly in common phrases like *this year* and *bless you*. Though it is sometimes described as producing /ʃ/, Zsiga (1995, 2000) shows that the production in this context is better described as only partially palatalized [sʲ], with the beginning quite [s]-like but ending with lower frequency noise and closer palatal constriction. It is also quite variable, showing gradient differences among speakers, words and pragmatic contexts. So though the word-internal palatalization does produce neutralization, the word-boundary palatalization does not.

When looking at typological data about contrasts and neutralizations, there are subtleties about what constitutes 'distinguishable' contrasts and 'noticeable' alternations which cannot be ignored, since native speakers, field linguists and instrumental measurements frequently disagree on the distinct alternations and contrasts. In particular, native speakers and fieldworkers are each influenced by their native phonologies (Lee, Vakoch, & Wurm, 1996; Ladd, 2011), and there are often measurable statistical differences for ostensibly neutralized contrasts (Warner, Jongman, Sereno, & Kemps, 2004; Yu, 2007). In addition, because typological databases, even more than the original grammatical descriptions, reduce the patterns to simple abstract representations, phonological systems in a typological database are only broadly comparable (Simpson, 1999). Nevertheless, the primacy of phonemic contrasts in phonological fieldwork has made (approximate) neutralizations among the better documented cross-linguistic phenomena.

2.2 Contrast types

Knowing only that different phonetic contexts can condition alternations, and that alternations can neutralize contrasts, we might naively assume:

(5) Hypothesis: Palatalizing contexts and non-palatalizing contexts are independently capable of supporting a contrast or of neutralizing an underlying contrast.

In that case, there are four possible patterns of contrast, one of which may be usefully subdivided. I will refer to these types by the names in parentheses, though the reasons for calling them as such might not be clear until the formal analysis.

a) No neutralization (Full contrast): Phonemic contrasts based on place of articulation

[3]Among possible exceptions, the CMU Pronouncing dictionary lists a few foreign proper names, and *dossier*, *monsieur*, *tissue* and *Hoosiers*. The Merriam-Webster pronunciations of these words either resyllabify (/das.jej/, /da.si.jej/) or indicate /ʃ/.

within the coronal region are maintained regardless of the preceding or following vocoids.

Catalan contrasts alveolar /s, ts/ with alveopalatal /ɕ, tɕ/, and though front vocalic contexts condition primary constriction locations that are further forward for both series, the contrasts are not neutralized (Recasens & Espinosa, 2007).

[ˈmi**s**il]	'missile'	[pəˈ**s**at]	'past'
[təˈ**ɕ**i]	'weaved'	[bəˈ**ɕ**a]	'go down'

b) Contrast in non-palatalizing contexts (Special-case neutralization): A phonemic contrast is supported except in a particular palatalizing environment.

Polish has a dental series /s, z, t, d/ which contrasts with an alveolo-palatal series /ɕ, ʑ, tɕ, dʑ/ in most environments (Rubach, 2003).

[gɫɔ**s**]	'voice'	[ɫɔ**ɕ**]	'moose'
[**z**a**t**ɔka]	'bay'	[**tɕ**ɔ**tɕ**a]	'aunt'
[**z**ɛgar]	'clock'	[**ʑ**ɛlɛ]	'herb'

But they neutralize with the alveolo-palatals just at morpheme boundaries before /i, j, ɛ/.

[gɫɔ**ɕ**itɕ]	V.INF of [gɫɔs]	[pɔ**tɕ**itɕ]	V.INF of [pɔt] 'sweat'
[gɫɔ**ɕ**ɛ]	LOC.SG of [gɫɔs]	[pɔ**tɕ**ɛ]	LOG.SG of [pɔt]
[bra**tɕ**iʃɛk]	DIMIN of [brat] 'brother'		
[ʃkɔ**tɕ**i]	NOM.PL of [ʃkɔt] 'Scot'		

c) Contrast in palatalizing contexts (General-case neutralization): A phonemic contrast is supported only in a particular palatalizing environment.

As discussed previously, borrowing from Sanskrit into Gujarati reintroduced a contrast between /s/ and /ʃ/ before /i,e/, while Sanskrit /ʃ/ before the other vowels was adapted as /s/ (Pandit, 1954).[4] Thakur (2011) shows that some speakers do maintain this pattern of contrast.

[4]Several examples of Sanskrit /ʃja/ were adapted as /ʃa/, indicating there was a contrast before /a/ in the data Pandit examined. The focus of his paper is the diachronic change as reflected in orthography, rather than the synchronic phonology, and he specifically refrains from making a claim about the phonemic contrasts of any particular Gujarati dialect.

/**ʃ**ikh/	'advice'	(cf Sanskrit /ʃiksaː/)
/**s**indur/	'vermillion'	(cf Sanskrit /sinduram/)
/aːr**s**o/	'mirror'	(cf Sanskrit /aːdarʃah/)

d) No contrast (Total neutralization or Complementary distribution): no context supports a phonemic place contrast within the coronal region. In complementary distribution, the allophone found in palatalizing contexts is different than the allophone found in non-palatalizing contexts, while in total neutralization, the allophones are nominally the same.

Minais Gerais Portuguese (as well as several other Brazilian dialects) has affricates [tʃ, dʒ] before /i, j, ĩ/ in complementary distribution with alveolar stops [t, d] elsewhere (Cristófaro-Silva, 2003).

[a**dʒ**iʒɛˈ**tʃ**ivu]	'adjective'	[ˈ**t**ɛha]	'earth'
[ˈ**tʃ**ĩta]	'paint'	[ˈ**d**a**d**u]	'dice'
[ˈɔ**dʒ**ju]	'hate'	[kaˈ**d**ejra]	'chair'

Sao Paolo Portuguese has [t, d] even when followed by /i, j, ĩ/, with no affricates in the language (Cristófaro-Silva, 2003).

[a**d**iʒɛˈ**t**ivu]	'adjective'	[ˈ**t**ɛha]	'earth'
[ˈ**t**ĩta]	'paint'	[ˈ**d**a**d**u]	'dice'
[ˈɔ**d**ju]	'hate'	[kaˈ**d**ejra]	'chair'

In cases like Brazilian Portuguese, where the complementary distribution pattern involves both a large degree of phonetic change and a significant role in defining social identities (Pagotto, 2004), the alternation is consistently recorded in grammatical descriptions (even as it may gloss over variability across speakers and morphological contexts), while cases reported as total neutralization and the other three kinds of contrast patterns also involve subtle alternations (like word-boundary palatalization in English and fronting in Catalan), that might or might not be included in grammatical descriptions or represented in a typological database. Because complementary distribution like that of Minais Gerais does involve two distinct sound categories, I will distinguish it from total neutralization, but I acknowledge that the distinction is qualitatively different than the distinctions among contrast types.

These categories are only positional allophones rather than constrastive phonemes, and distinguishing between the categories requires comparison across phonetic contexts rather across units of meaning. As such, the representation in typological databases may be less consistent.

2.3 Type counts

The examples provided in §2.2 illustrate that each of the contrast types is a possible pattern, but one might justifiably still wonder about the validity of the preceding hypothesis (5) that palatalizing and non-palatalizing contexts are independently capable of supporting a phonemic contrast. There are very few absolute language universals, i.e. true of the set of all languages without exception (Hyman, 2008; Evans & Levinson, 2009), but there are still systematic trends across languages ('statistical universals') which can only be reliably revealed by broad surveys and quantitative analysis (Bickel, to appear). As a preliminary step, I am compiling a database of the palatalization contrast patterns in documented languages. Drawing primarily from the P-base database of sound patterns (Mielke, 2008), I have collected data on the phonemic contrasts and vocoid-triggered palatalizations among coronal fricatives and plosives. The data assembled so far represent 198 languages, a bit more than a third of the languages in P-base.

The palatalization patterns include both active allophonic alternations and static phonotactic regularities, though the overwhelming majority are active alternations. As with the above examples from English and Polish, some palatalization contexts are limited to particular prosodic or morphological domains, e.g. within syllables, at morpheme boundaries or across word boundaries, as long as they approximately apply throughout the language. The data sample does not contain patterns that are limited to a particular affix or a limited irregular portion of the lexicon. In the case of multiple documented dialects with different contrast patterns, I have only included the dialect that seemed to be best documented.

For each language, I recorded how many distinct place-of-articulation categories there were within the coronals, treating fricatives separate from stops and affricates, and including secondary palatalization but excluding labial, velar or pharyngeal articulatory features. For the plosives, a stop vs. affricate contrast could count as distinct places if the transcription of the affricate's release suggested a different place of articulation (e.g. /t/ vs. /t͡ɕ/) but not if it suggested the same place of articulation (e.g. /t/ vs. /ts/).

Among the languages with more than one coronal place, if the language neutralized the place contrast next to a high front vocoid, it was categorized as special-case neutralization, and if it maintained the contrast next to high front vocoids but neutralized in some other vocalic context, it was categorized as general-case neutralization. If there was no palatal

Table 1: Palatalization patterns by contrast type. The number of languages of each contrast type is on the right, with the number of language families in parentheses.

Type	Fricatives	Plosives
Full Contrast	(25) 68	(32) 86
Special-case Neutralization	(8) 15	(10) 15
General-case Neutralization	1	1
Complementary Distribution	(7) 19	(11) 23
Total Neutralization	(22) 81	(16) 72
No coronals	(4) 13	0

alternation recorded or if a recorded palatalization was non-neutralizing, the language was categorized as full contrast. Among the languages with just one coronal place, I distinguished between complementary distribution patterns which have a recorded palatalization alternation and total neutralization patterns which have no recorded palatalization. In addition, some languages have no coronal fricatives. The resulting language counts are recorded in Table 1.

To the left of the language counts in Table 1, I have indicated the number of language families that the counted languages represent, as an indication of the diversity of each type. The families counted are the highest level genealogical groups as assigned in P-base, which are based on that of the Ethnologue (R. G. J. Gordon, 2005). The Australian languages are grouped as a single family, while more conservative groupings are used in New Guinea and the Americas. The language families with the most languages in the sample are Niger-Congo (47 languages), Indo-European (23 languages), and Afro-Asiatic (22 languages). Table 2 shows the number of language families that the languages represent for the fricative contrast types, arranged by the six large areas used by Dryer (1992). Each language family is assigned to only one language area: The Africa area includes all of Afro-Asiatic and the other languages of mainland Africa; the Australia area includes the Australian languages and New Guinean languages; EurAsia includes the languages of Europe and Asia, except for those of Southeast Asia; The border between the North America and South America areas is drawn through central America between Mayan and Chibchan; The Southeast Asia area extends from Sino-Tibetan and Tai-Kadai out to Polynesia.

In addition to the contrast types listed, I catalogued which side of the fricative the conditioning vocoid was on, for each of the contrast types that involve an allophonic alternation or phonotactic regularity. The examples provided previously have all been right-side contexts,

Table 2: Language families containing each fricative constrast type, arranged by area.

	Afrc	Aust	EurAs	NAmer	SAmer	SEAs	Total
Full contrast	3	0	8	7	5	2	25
Special-case neutralization	2	1	3	0	2	0	8
General-case neutralization	0	0	1	0	0	0	1
Complementary distribution	2	0	2	0	0	3	7
Total neutralization	3	3	5	5	4	2	22
No coronals	1	2	1	0	0	0	4

Table 3: Numbers of languages reported to have each kind of conditioning context type

Conditioning Context	Languages
No alternation	[151]
Right side	38
Left side	3
Dual side	4
Double side	0

i.e. /s/ → [s]//_i/. The full set of context types are:

(6) a) right-side (/_i/): English /_j/

b) left-side (/i_/): Lower Grand Valley Dani (Austronesian) /{i,u}_{e,o,a}/

c) dual-side (/i_/,/_i/): Baré (Arawakan) /{i,e}_/,/_{i,e}/

d) double-side (/i_i/): None attested

The dual-side context type conditions palatalization when there is a high front vocoid on either side of the fricative, while the double-side context type conditions palatalization only when there is a high front vocoid on both sides. The context type counts are shown in Table 3, collapsing fricatives and plosive together, and collapsing across contrast type. Languages that were non-palatalizing in both fricatives and plosives are included as 'no alternation'.

2.4 Discussion

The type counts in Table 1 indicate that contrasts in palatalizing contexts and non-palatalizing contexts are closely related, and that there is an asymmetry between the context types. Whether comparing language counts or family counts, and whether looking at fricative contrasts or plosive contrasts, the most common contrast types are full contrast and total neutralization—where approximately the same allophones are found in each context.

Special-case neutralization and complementary distribution are well-attested but less common, while general-case neutralization is barely attested, with just one case with fricatives and one with plosives. The asymmetry between special-case neutralization and general-case neutralization can be expressed as an 'implicational statistical universal': if a language has a phonemic contrast that is based on place of articulation among coronal obstruents and it is observed before high front vocoids, then it probably also has the contrast in other vocalic contexts; if a language does not have such a contrast before low back vowels, then it probably does not have a contrast by high front vocoids.

It is worth considering to what extent the typological counts in Table 1 might be influenced by areal or genealogical factors rather than more universal characteristics of articulation and perception. For example, 10 of the 13 languages with no coronal fricatives are Australian languages, and these are all the Australian languages in the sample, as in fact Australian languages characteristically lack fricatives entirely (Lindblom & Maddieson, 1988). The data in Table 2 is too sparse to obtain much power with the statistical methods that are conventional in areal typology (partly because I have only counted families rather than genera), so we cannot rule out the possibility that areal factors have influenced the overall distribution, but the distributions outside of the Australia (and New Guinea) area are consistent with the overall trends. In all the areas except Australia, special-case neutralization is observed in fewer families than full constrast is, and in all the areas except Southeast Asia, it is observed in fewer families than total neutralization is. And general-case neutralization is observed in just one area, while the attestations of special-case neutralization (4 areas) and complementary distribution (3 areas) are more widely distributed. A statistical analysis of a suitably expanded sample (cf. Bickel, to appear) may still reveal areal biases that require a more careful formulation of the generalizations, but the simple interpretation of these type counts appears reasonable.

In discussions of comparable contrast typologies of syllable augmentation (Smith, 2000) and vowel nasalization (Kager, 2006), general-case neutralization (there called 'reverse positional neutralization') is reported as an unattested pattern. Since the general-case neutralization type is barely attested compared to the other contrast types, the two cases of general-case neutralization deserve further inspection.

The perspective of Pandit (1954) on the Gujarati fricatives is primarily historical, and he specifically refrains from making claims about the phonemic status in any particular dialect. Thakur (2011) contrasts Pandit's description with some previous descriptions that called [ʃ] an allophone of /s/ (i.e. complementary distribution), and with later descriptions that describe /s/ and /ʃ/ as contrastive in all environments, primarily on the basis of additional loanwords from Hindi and English. Thakur performed a production study of 20 Gujarati

speakers (all multilingual) and found that 2 of them evidenced the pattern of contrast suggested by the data Pandit described, while 6 others had some other system of marginal contrast in other contexts (e.g. codas) and the other speakers maintained a contrast in all environments. Thakur speculates that the number of full contrast speakers in the study may be related to the multilingualism of the participants (exposed to Hindi and English which both have a contrast), and that monolingual Gujarati speakers are more likely to have an allophonic system. It is still somewhat unclear how the general-case neutralization pattern and other patterns of contextual contrast fit in to the variation among speakers.

In the plosives column, the language categorized as having general-case neutralization is Chama (Tacanan, Bolivia). The categorization is based on the observation that /tʃ/ is not found before /e/ and the sequence /tʃa/ is only observed in word-initial syllables (Key, 1968), suggesting that /t/ and /tʃ/ contrast before the other vowels (/i,o/) but not before /e/, and only marginally before /a/. However, the original description also mentions that the gap may be accidental rather than structural, because /tʃ/ is rare. On the other hand, similar depalatalization neutralizations are observed in Ukrainian fricatives and plosives (Rubach, 2006) and Japanese fricatives (Ito & Mester, 2003).

To illustrate, Japanese maintains a palatalization contrast before low and back vowels

/**s**akɯ/	'fence'	/**ʃ**akɯ/	'wine-serving'
/**s**ɯgata/	'figure'	/**ʃ**ɯgo/	'protect'

but the contrast is neutralized to the non-palatalized allophone before /e/

/**s**epaado/	< 'shephard'	/**s**ekai/	'world' (< earlier ʃ)
/hana**s**e/	'speak' IMP		

and to the palatalized allophone before /i/.

/**ʃ**inema/	< 'cinema'	/**ʃ**ita/	'low '
/hana**ʃ**i/	'speak' INF		

Japanese and Ukrainian do not count as general-case neutralization because they do neutralize before /i/, but they confirm the possibility of depalatalizing neutralization before /e/.

For Gujarati and Chama, the two cases of general-case neutralization, the contrasts and alternations have not been studied as extensively as those of some other languages like English and Catalan, but we can at least locate the status of general-case neutralization between two extremes. The general-case neutralization is a possible distributional language pattern, and it apparently does have some psychological reality for some speakers. On the other hand, it is not clear that a community of speakers can maintain the pattern, and it is clearly typologically unusual.

3 Phonetic grounding

A long-running question within phonological and phonetic typology, as well as in associated theory construction and experimental research, is the relationship between cross-linguistic phonological patterns and associated phonetic pressures. For example, the exceptional cases of Japanese and Ukrainian highlight an unsurprising but profound generalization: the neutralizations triggered by high front vocoids are characterized as palatalization, not depalatalization.[5] Surely it is not a coincidence that palatalizing vocoids are phonetically similar to palatalized obstruents, but how does that association pass from a phonetic association to an observable regularity in phonology? This section reviews some of the issues in attempting to ground phonological patterns in phonetics and decribes the phonetic factors that are likely responsible for most palatalization contrast patterns.

3.1 Naturalness and grounding in phonology

There is a long tradition of phonetic explanations for trends in phonology, as well as counter-arguments that phonology is not as closely tied to phonetics as some of these explanations have suggested. Much of the literature has focused on speaker-oriented or 'ease of articulation' explanations and listener-oriented or 'ease of perception' explanations. Speaker-oriented explanations have focused on the articulatory difficulty of coordinated phonetic targets (Jakobson, 1962/1971; Marsico, Maddieson, Coupe, & Pellegrino, 2004), while listener-oriented explanations have focused on the difficulty of recognizing and reproducing a particular acoustic signature (Ohala, 1981; Hume & Johnson, 2001b; Blevins, 2004). On top of that, integrated models of adaptive speaker-hearers take the perspective that through experience of hearing their own voices and coping with communication difficulties, speakers gain unconscious knowledge of aspects of the relationship between production and perception, and actively manage them (Lindblom, 1990; Goldstein & Fowler, 2003; Hayes et al., 2004). The speaker-oriented and integrated perspectives tend to favor a tight integration of phonetics into synchronic phonology, while listener-oriented and formalist perspectives tend to caution against 'phonetic determinism', warning that conceptions of naturalness need to leave room for unnatural patterns, and taking the perspective that these trends are only indirectly grounded in phonetics, via diachronic sound changes that speakers and hearers have little active control over. (Anderson, 1981; Hyman, 2001; Blevins, 2008). These perspectives

[5]Standard Bulgarian is apparently an exception (Scatton, 1984). Some non-neutralizing alternations also involve motion away from the palate in 'palatalizing' contexts, e.g. fronting in Catalan (§2.2). In both Koya Gondi and Adilabad Gondi (Dravidian), alveolopalatals alternate with palatoalveolars, but in Koya it is the alveopalatals that are found in before front vowels, while in Adilabad it is the palatoalveolars (Subrahmanyam, 1968, cited in P-base).

disagree on the extent that phonetics is actively integrated into the learner's phonology, but they agree that many of the trends in phonological typology are grounded in phonetics.

An important distinction for the current discussion is the division between pressures in the physical system, in the perceptuo-motor system, and in higher level cognitive and sociocultural systems. In a taxonomy of 'filters' on phonology, Hume and Johnson (2001a) list perception and production as low level phonetic effects alongside the higher level effects of (cognitive) generalization and (social) conformity. Focusing on the phonetic biases in sound change, Garrett and Johnson (in press) make a four-way distinction among motor planning, aerodynamics, gestural mechanics and perceptual parsing. Aerodynamics and gestural mechanics biases deal with the physical system, in which sounds are physically or physiologically more or less difficult and more or less similar. The motor planning and perceptual parsing biases deal with the perceptuo-motor system, in which sequences of sounds may be prone to change due to speaker or listener error in adapting to the physical system. None of the treatments of naturalness claim that any of these kinds of biases is uniquely responsible for producing trends in typological patterns, but some treatments focus on one aspect or another at the expense of others.

The debate about to what extent phonology is phonetically natural apparently hinges on how much the perceptuo-motor system adapts to the physical system and to the particular features of the language being learned. If the perceptuo-motor system were non-adaptive, or completely adaptive to the particularities of the learned language but not to the physical system, then the trends in typology would only be dependent on the physical system and the higher level systems, and the phonetic grounding would be indirect. On the other hand, if the perceptuo-motor system were adaptive to the physical system but not to the language being learned, then the phonetic grounding of typological trends would be direct. Experimental evidence indicates that the perceptuo-motor system is adaptive both to the characteristics of the physical system (e.g. Lindblom, 1990; Berent, Steriade, Lennertz, & Vaknin, 2007) and to the pecularities of the language being learned (e.g. Beddor & Krakow, 1999; Dell, Reed, Adams, & Meyer, 2000), suggesting that some kind of mixed model is required.

In an induced constraint grammar (Wilson, 2006; Hayes & Wilson, 2008; Boersma, 2008; Moreton & Pater, 2011), all of the constraints and their associated weights are induced from experience. Most of the learning is based on the distributional regularities of the language itself, but certain aspects of the learners' experiences are independent of the language being learned, or they have similarities across language exposure. The learner is able to evaluate a wide variety of generalizations, and reify any that coincidentally appear, adapting to the pecularities of the language. But low-level phonetic experience creates a weak a bias that can influence generalizations at a slightly higher level, though it won't necessarily do

so, depending on the rest of the system (Boersma, 2008; Hayes et al., 2009). Similarly, the learning process creates a bias for certain kinds of simplicity in generalizations, but it won't necessarily reach that effect (Moreton & Pater, 2011). Within such a system, the constraints that represent high-level generalizations are qualitatively different than the biases that influence the development of the system, but the high-level constraints are grounded in the low-level biases inasmuch as the high-level constraints reflect generalizations about the historical pathways made possible by the biases.

3.2 Functional biases and formal constraints

There are three groups of functional biases involved in a palatalization alternation, which together suggest how each of the contrast types is phonetically natural.

Articulatorily and perceptually, [ʃ] is more like [i] than [s] is, making the sequence [si] articulatorily more complicated than [ʃi] as well as more easily misperceived or mispronounced as /ʃi/, while there is no particular similarity between [s] and other vowels. In the terminology of Articulatory Phonology (Gafos, 1999; Goldstein & Fowler, 2003; Bateman, 2007), the gestural goal for production of sounds usually transcribed as [s] is critical constriction (i.e. a distance producing turbulent airflow) in the alveolar region by the tongue tip, with constriction orientation varying from 'up' (apical) to 'down' (laminal) across speakers and languages. The gestural goal for [ʃ] is critical constriction in the post-alveolar region by the tongue tip, with constriction orientation usually 'down' (laminal). In comparison, the gestural goal for [i] is narrow constriction by the tongue body in the palatal region. Because the tongue tip and the tongue body are physically coupled, the coordinated production of post-alveolar laminal tongue tip constriction and palatal tongue-body constriction requires minimal motion or minimal adjustment in the case of gesture merger, while the transition from alveolar constriction requires more careful timing and a bit more motion, or a bit more adjustment in the case of gesture merger. When gesture merger does occur, the result is slight retraction of the point of tongue tip constriction and closer constriction by the tongue body, i.e. sounds transcribed as [sʲ, ɕ] or [ʃ], palatalization. Acoustically, sibilants are primarily distinguishable by the frequency of frication noise peak intensity, which varies from about 2 kHz for retroflex fricatives to about 8 kHz for dental fricatives, with palatal fricatives lying near the lower end of the scale (Stevens, 1989; Boersma & Hamann, 2008). In comparison, the formant structure of [i] (as well as [j]) is characterized by a relatively high F2, generally exceeding 2 kHz (Peterson & Barney, 1952). The similar formant structure (which has a common origin in the resonance of the front cavity) makes the fricative-vowel transition perceptually similar to palatalization. I will call this cluster of biases No[si]: Avoid

non-palatalized articulations next to high front vocoids.

In contrast to these contextual biases, there is some reason to believe that outside of the particular environment of high front vocoids (or palatal consonants) constrictions in front of the alveolar ridge are somewhat phonetically preferable to post-alveolar constrictions. Because the alveolar ridge sticks out from the curve of the roof of the mouth, and in 'rest position' the tip of the tongue is in front of the ridge, constriction in front of the ridge tends to require smaller adjustments to the tongue's trajectory than post-alveolar constriction (Lindblom, Diehl, Park, & Salvi, 2011). In addition, the noise of frication and stop bursts has the greatest intensity (for a given air pressure) when the location of the constriction is approximately two centimeters away from the lips, at about the alveolar ridge when the lips are relaxed (Stevens, 1989). I will call this bias NO[ʃ]: Avoid palatalized articulations

In opposition to these structural pressures, which tend to reduce contrast, languages must have a sufficiently informative phonetic encoding to permit recovery of the intended meanings. In conventional Optimality Theory, these pressures are encoded as categorical faithfulness constraints, but approaches interested in the interface between phonetics and phonology have explored other representations such as attractive forces in a dynamical system (Gafos & Benus, 2006), a combination of penalties for contrast loss and divergence from articulatory targets (Flemming, 2001), or an emergent property of managing the articulatory-acoustic map in a production-perception loop (Boersma & Hamann, 2008). I will call this bias IDENT: Maintain the constriction location targets specified by the phonemic input.

In the following discussion, I will refer to these biases as constraints that make categorical distinctions among phonological-phonetic mappings that are more or less preferred by that bias. Though this convention is a simplification of the synchronic system, this allows a clearer view of their interactions in typology.

4 Typological predictions of naive models

In this section, I compare the typological predictions under Optimality Theory with the predictions under Harmonic Grammar, assuming the phonetically-grounded constraints introduced in the previous section.

4.1 Cumulativity

Both Optimality Theory and Harmonic Grammar describe mappings in symbolic systems in terms of the interaction of multiple violable constraints. In Harmonic Grammar, the constraints are weighted so that each violation has a corresponding numeric cost, and the

optimal input-output mapping is the one with the least cost. The tableau in (7) shows a trivial faithful mapping of /ʃa/ to [ʃa], when the IDENT constraint (penalizing changes) has twice the weight of NO[ʃ] (penalizing instances of [ʃ] in the output).

(7) Harmonic Grammar Tableau

weights	20	10	
/ʃa/	F: IDENT	G: NO[ʃ]	H
☞ ʃa		*	−10
sa	*		−20

With these weights, the cost H of the faithful output is −10, better than the −20 of the alternative. If the weight of NO[ʃ] were increased to 30, or if an additional constraint with a weight greater than 10 preferred [sa] over [ʃa], then [sa] would be optimal. The tableau as presented here assumes that there is no such higher weight constraint. More generally, tableaux like this which represent toy grammars assume that all other constraints and candidates are negligible: any constraints with non-negligible weights are satisfied equally by the given candidates, and any candidates favored over these by a high weight constraint are themselves eliminated by an even higher weight constraint.

In Optimality Theory, contraints are ranked, so that when constraint X in (8) outranks constraints Y and Z, if constraint X favors candidate A over candidate B (and no higher ranked constraint favors candidate B or any another candidate over candidate A), A will be the optimal output regardless of how much it violates constraints Y and Z.

(8) Optimality Theory Tableaux

a.)

/input/	X	Y	Z
☞ A		*	*
B	*!		

b.)

/input/	X	Y	Z
☞ A		***	***
B	*!		

In contrast, the weighting of Harmonic Grammar constraints means that even though constraint X in (9) is weighted above constraints Y and Z, and constraint X favors candidate A over candidate B, candidate B can still be optimal if the cumulative costs of candidate A's violations of constraints Y and Z have a greater cost than candidate B's violation of constraint X.

(9) Harmonic Grammar Tableaux

a.)

weights	40	30	20	
/input/	X	Y	Z	H
A		*	*	−50
☞ B	*			−40

b.)

weights	40	30	20	
/input/	X	Y	Z	H
A		**		−60
☞ B	*			−40

In (9a), the cumulative cost of candidate A exceeds the cost of candidate B because of violations of multiple constraints, while in (9b), the cost of candidate A exceeds the cost of candidate B because of multiple violations of a single constraint. The violation cumulativity in (9b) has played a major role in theorizing about the nature of constraint interaction, but it is influenced by a wider variety of assumptions than the constraint cumulativity in (9a), making it harder to distinguish what is responsible for any rarity of patterns generated by violation cumulativity.

Any phenomenon analyzed with ranked constraints can be similarly analyzed by weighting those same constraints, but because of cumulative constraint interaction, some patterns analyzed with weighted constraints cannot be produced by ranking the same set of constraints. For example, the tableaux in (10) show a possible Harmonic Grammar language pattern.

(10) Constraint cumulativity in contrast typology

	weights	30	20	20	
a.)	/sa/	G: NO[ʃ]	S: NO[si]	F: IDENT	
	☞ sa				0
	ʃa	*		*	−50
b.)	/ʃa/	G: NO[ʃ]	S: NO[si]	F: IDENT	
	☞ sa			*	−20
	ʃa	*			−30
c.)	/si/	G: NO[ʃ]	S: NO[si]	F: IDENT	
	☞ si		*		−20
	ʃi	*		*	−50
d.)	/ʃi/	G: NO[ʃ]	S: NO[si]	F: IDENT	
	si		*	*	−40
	☞ ʃi	*			−30

In (10a), /sa/ maps to [sa] because it violates none of these constraints, while the alternative violates both the general markedness constraint G, which penalizes /ʃ/ in any context, and the faithfulness constraint F, which penalizes differences between input and output forms. In (10b), /ʃa/ maps to [sa] because the penalty for the violation of constraint G is greater than that of constraint F. In (10c), /si/ remains the same because the unfaithful candidate violates both constraint G and constraint F. And in (10d), /ʃi/ remains the same because the unfaithful mapping violates both constraint F and constraint S, which penalizes /s/ next to a palatal vowel, and the cumulative cost is greater than the cost of the violation of G. While

the constraint cumulativity in (9a) could have been eliminated by changing the weights of the constraints, this constraint cumulativity in (10d) cannot be eliminated without changing the optimal output in (10b) or (10c), with the result that this pattern of mappings cannot be expressed by ranking these constraints.

4.2 Categorical predictions

Typological claims about Optimality Theory and Harmonic Grammar have primarily been directed at distinguishing possible languages from impossible languages, without attempting to distinguish common or usual language types from the rare or unusual. Prince and Smolensky (1993/2004) proposed that the exhaustive reranking of constraints revealed the set of possible languages (a 'factorial typology'), and following the traditional paradigm of generative grammar, attempts to summarize typological variation in terms of constraint interaction have used undergeneration and overgeneration of language types as a primary criterion for judging the validity of the analysis, where undergeneration is the failure to account for an attested language type, and overgeneration is the prediction of unattested language types.

In Harmonic Grammar, each intput-output pattern in a set of tableaux determines a system of inequalities (Potts et al., 2010), which means that the language type is a function of the ratios among the constraint weights.[6] For example, the tableau in (7) behaves one way when the weight of constraint G (W_G) is greater than the weight of F (W_F), and another way when W_F is less than W_G. Or equivalently, if $W_G : W_F < 1$, we get the faithful mapping, and if $W_G : W_F > 1$ we get the unfaithful mapping. A set of tableaux like those in (10) create a more complicated system of inequalities, but at least in this case, the inequalities reduce to fairly simple expressions. The pattern in (10) is in fact the typologically unusual general-case neutralization type exemplified by Chama and Gujarati, maintaining the contrast between /s/ and /ʃ/ before /i/ but neutralizing to [s] in other vocalic contexts. The weighting conditions for producing this pattern type are:

(11) General-case Neutralization (GN):
Occurs in HG when $W_G > W_F > |W_S - W_G|$.

The ratios of three values can be displayed in two dimensions as a ternary mixing plot, where proximity to each corner is associated with increased proportion of that corner's variable (Fig. 1 left). The corner labeled (1,0,0) is a weight configuration where the weight of constraint X is so large that the weights of Y and Z are negligible. The opposite point labeled (0,0.5,0.5) is a weight configuration where the weight of constraint X is negligible

[6] Potts et al. (2010) require the weights to be positive integers and the separations in the inequalities to be at least one, which complicates the system of inequalities but allows more efficient computational methods.

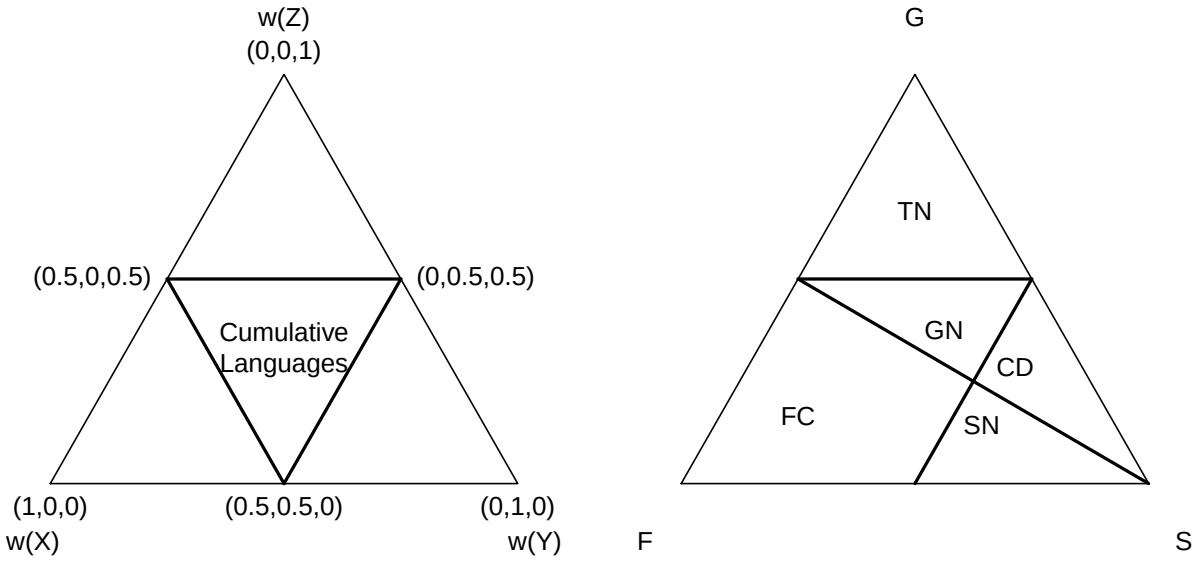


Figure 1: Language types in the space of ratios among three weighted constraint weights. The left plot shows the region of non-OT interactions in the space of a generic set of three constraints. The plot on the right shows the contrast typology.

and the weights of Y and Z are equal. The center of the plot is a weight configuration where all three weights are equal. Cumulative interactions are those with weight ratios lying inside the central triangle, where the sum of any two constraints is greater than the third. In the space of the ratios among F, S, and G (Fig. 1 right), general-case neutralization lies in the middle within the region of cumulative weight configurations, surrounded by four other pattern types which can be represented in Harmonic Grammar or Optimality Theory equally well.

(12) a) Full contrast (FC):
Occurs in OT when F » {S, G}.
Occurs in HG when $W_F > W_G > W_S - W_F$.

b) Special-case Neutralization (SN):
Occurs in OT when S » F » G.
Occurs in HG when $W_S - W_G > W_F > W_G$.

c) Complementary Distribution (CD):
Occurs in OT when S » G » F.
Occurs in HG when $W_S - W_F > W_G > W_F$.

d) Total Neutralization (TN):
Occurs in OT when G » {F, S}.
Occurs in HG when $W_G > W_F + W_S$.

Because Optimality Theory does not allow constraint cumulativity, and general-case neutralization can only be expressed with constraint cumulativity, Optimality Theory predicts that general-case neutralization is not possible. In Harmonic Grammar, general-case neutralization requires the constraint weights to have approximately equal ratios, but there is no reason to think that would be a problem, making general-case neutralization look like it should be an ordinary language pattern.

Recalling that the direct grounding assumption may need to be taken lightly, in order to leave room for unnatural patterns, the Optimality Theory typology could be forgiven for failing to account for a couple unusual languages. If the scope of phonetically-driven Optimality Theory is limited to natural patterns, and this general-case neutralization is in some sense unnatural, a 'crazy rule' resulting from coincidental recent changes, then this is merely a reminder of the limitations of the framework's assumptions. On the other hand, the Harmonic Grammar typology correctly predicts that general-case neutralization does exist, but fails to express the asymmetry between contrast in the special-case context and contrast in the general-case context.

4.3 Quantitative predictions

Considering that much of the systematicity in cross-linguistic variation consists of trends with exceptions rather than absolute universals, the traditional goal of generative linguistics to distinguish possible languages from the impossible is at once too ambitious and too simplistic. A robust model of cross-linguistic variation should be able to express the relationships between statistical universals and their functional explanations. Variationist constraint grammars potentially fulfill this role, and both the Optimality Theory typology and the Harmonic Grammar typology do predict that total neutralization and full contrast should be more common than the other contrast types, but the quantitative predictions do not help Harmonic Grammar express the asymmetry between the special-case and general-case contexts.

In a factorial typology, some rerankings crucially affect the pattern of mappings from underlying forms to surface forms, while other rerankings do not change the patterns. Partially Ordered Grammars (Anttila, 1997) model synchronic variation as the reranking of constraints within the limits of the ranking conditions established by observed non-variable related data. Coetzee (2002) extends this idea to cross-linguistic variation, modeling the

Table 4: Language type probabilities according to factorial typology r-volume, compared to observed distribution of coronal fricative palatalization type

Type	Partial Order	Total Order	r-vol Prob	Observed
Full Contrast	F » {S, G}	F » G » S F » S » G	2/6 = 33%	37.0%
Special-case Neutralization	S » F » G	S » F » G	1/6 = 17%	8.2%
General-case Neutralization	–	–	0%	0.5%
Complementary Distribution	S » G » F	S » G » F	1/6 = 17%	10.3%
Total Neutralization	G » {F, S}	G » F » S G » S » F	2/6 = 33%	44.0%

statistical universals that correlate backness with rounding in vowel inventories and aspiration with voicelessness in stop inventories. He argues that in addition to the constraints on input-output mappings, modeling of the trends in cross-linguistic variation requires 'preference constraints' which specify preferred rankings between pairs of constraints, comparable to universal constraint heirarchies (and subsuming them) but violable. Bane and Riggle (2008) dub the number of total constraint rankings that generate a given language pattern *r-volume*. Looking at Gordon's typology of quantity insensitive stress (M. Gordon, 2002), they show that r-volumes are indeed correlated with typological frequencies, so that the r-volumes associated with each stress pattern do at least generally distinguish the common stress patterns from the unattested stress patterns. Riggle (2010) develops a comprehensive argument that this correlation results from a learning strategy which assumes equal prior probability over permutations of the constraints and favors input-output mappings that have the highest estimated probability after previous observations. Riggle also develops a different analysis of the typologies Coetzee examined, rejecting 'preference constraints' but using stipulated redundancies in the constraint set to fit the quantitative trends.

Assuming neither preference constraints nor constraint redundancies, the sampled probabilities for the constrast typology are easy to compute manually. In the factorial typology of three constraints, there are just 3! = 6 total order permutations. By lining up those 6 permutations with the ordering relations shown in (12), we can calculate the r-volumes of each language type (Table 4). This model predicts that full contrast and total neutralization are similarly common, while special-case neutralization and complementary distribution are less common. As with the categorical interpretation of the constraint reranking, general-case neutralization is predicted to be impossible.

Similar quantitative predictions result from a Harmonic Grammar model that generates

Table 5: Language type probabilities according to area in ratio space, compared to observed distribution of coronal fricative palatalization type

Type	Area	Prob	Observed
Full Contrast	3/8	37.5%	37.0%
Special-case Neutralization	1/8	12.5%	8.2%
General-case Neutralization	1/8	12.5%	0.5%
Complementary Distribution	1/8	12.5%	10.3%
Total Neutralization	2/8	25.0%	44.0%

languages by sampling weights. Since the space of possible Harmonic Grammar weight configurations is most simply characterized by the weight ratios, the simplest assumption for sampling languages is a uniform distribution over the area in the ternary mixing plot (Figure 1). Under a uniform distribution over an area, the probability of drawing a sample in a particular subregion is proportional to the area of that subregion. Special-case neutralization, complementary distribution, and general-case neutralization each occupy 1/8 of the area of the whole triangle, so they are each predicted to occur in one out of eight languages, and total neutralization and full contrast are respectively expected to occur twice and three times as often (Table 5). The predicted probabilities correctly select full contrast and total neutralization as more common than the other contrast types, though underestimating the observed probability of total neutralization. But the expected probability of general-case neutralization is equal to the other alternation patterns, quite contrary to the observed probability.

Comparison of the predicted probabilities in Tables 4 and 5 with the observed language type probabilities shows that both of these models are not just quantitatively off but qualitatively wrong. The Optimality Theoretic r-volume leaves no possibility of general-case neutralization, while the Harmonic Grammar weight volume assigns too much probability to it, making it at least as ordinary as special-case neutralization and complementary distribution. If general-case neutralization is in some sense a phonetically unnatural pattern, the Optimality Theoretic model may have undergenerated it for that reason, reflecting a failure of the assumption of direct phonetic grounding, but the quantitative overgeneration of general-case neutralization by the Harmonic Grammar model has no obvious cause.

Table 6: Numbers of languages reported to have each kind of conditioning type

Conditioning Context	Palatalization	Nasalization
No alternation	151	67
Right side	38	9
Left side	3	12
Dual side	4	4
Double side	0	3

4.4 Evaluation

If Optimality Theory is still understood as a higher level description of a system that is described by Harmonic Grammar at a slightly lower level, the constraint ranking mechanism of Optimality Theory is an epiphenomenon without a clear corresponding phenomenon. The cumulative pattern of general-case neutralization is empirically unusual, but examination of the phonetic grounding gives no reason to expect a shortage of cumulativity, since the variation at the phonetic end of phonology (in post-lexical phonology and language-specific and word-specific phonetics) is deeply variable, with many factors, many of which have an additive effect (Cohn, 1993; Zsiga, 2000; Pierrehumbert, 2002; Tily et al., 2009). If phonetic variability is the source of cross-linguistic phonological variation, it is surprising that the result does not more closely resemble the source.

Another complication is that cumulative patterns seem to be more common in some phenomena than others. Most triggers of palatalization come after the target, but some precede the target, and some languages require triggers on both sides of the target at the same time. The data in Table 6 repeats the count data from Table 3 alongside a smaller sample of vowel nasalization contexts (second column). The 'no alternation' patterns are languages that don't license palatalization in any context (full contrast or total neutralization), 'right side' patterns have palatalization dependent only on the following vocoid, and 'left side' patterns have palatalization dependent only on the preceding vocoid. 'Dual side' patterns license palatalization if there are high front vocoids on either side of the obstruent, and a 'double side' pattern would license palatalization only when there are high front vocoids on both sides of the obstruent. In this typology (see Figure 2), the double side type is a cumulative pattern, and formally parallel cases have been discussed in the literature as two-trigger lenition (Kirchner, 1998), doubly privileged licensing (Jesney, 2009) and worst-of-the-worst neutralization (Smolensky, 2006), as evidence for quantitatively evaluated constraints,

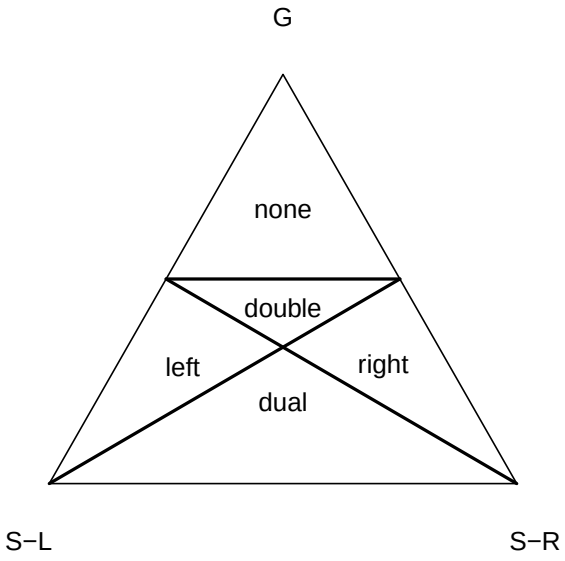


Figure 2: The typology of contextual conditioning in two contexts, in the space of ratios among the three constraint weights. The constraints S-L and S-R correspond to triggers in the left and right side contexts, and G is the general markedness or faithfulness constraints that resist contextual alternation.

weighted constraints or constraint conjunction. No cases of double side licensing are found for palatalization in this sample, in contrast to three cases of double side context in the smaller nasalization sample. The large difference between the counts of left side and right side context types in palatalization suggests that the left side context does not provide as strong a phonetic motivation for palatalization. If left side palatalization is itself relatively unnatural, that might reduce the probability of double side palatalization. Previous reports of double context phenomena, along with the attestations of double side nasalization, suggest that those cumulative patterns might be more common than the general-case neutralization and double side conditioning in the palatalization typologies. The samples are too small to know if the differences are reliable, but the differences are suggestive.

The model of language generation assumed in the discussion thus far has been synchronic and parallel, assuming that production, learning and language change can be approximated by one step of selecting weights or one step of selecting input-output mappings. But these are processes extended in time, and in particular, the sequential relationships of variation, phonologization and reanalysis are known to create asymmetries in the frequencies of language types (Myers, 2002; Blevins, 2004). A more careful treatment of language change

seems to be needed.

5 Predictions of evolutionary models

In this section I discuss three hypotheses about how language change restricts the abundance of cumulative patterns in typology, each of which has been implemented as a computational model. The first model is described by Wedel (2007), and I have also reimplemented it in order to be better understand its behavior. The following two models instantiate two other hypotheses I have considered, which at an abstract level are similar to Wedel's model, but they rely less on simulation and more on analysis. None of these models offers a convincing account of the shortage of typological cumulativity, but they suggest aspects which may be involved.

5.1 Wedel's exemplar model

Wedel (2007) discusses cumulativity as one example of how repeated cycles of perception and production in memory-rich processing can create categorical regularity. He shows that in a simple exemplar model of perception and production, a mild bias towards uniformity across the lexicon causes weak production biases to become entrenched. As a result of the entrenchment process, production biases interact non-cumulatively, so that the outcome depends on the typical sum of biases across the lexicon, rather than the sum of biases on individual exemplars. Though this model demonstrates biases that do not interact cumulatively, it does not demonstrate patterns of alternation or contextual regularities that are non-cumulative.

In these simulations, language production is idealized as replication of rich-memory detailed exemplars, and the roots of language patterns lie in error biases, gently pushing production or categorization in one direction or another. The repetition of recalled exemplars and an analogical bias across the set of exemplars creates positive feedback, which pushes the system toward categorical states. In the simplest model, without an analogical bias, each exemplar is replicated with random noise, modeled as a small probability that the exemplar recalled as one of two allophones will be flipped to the other allophone in production. Each exemplar produced is dependent only on the recalled exemplar plus noise, and so the set of exemplars wanders between the two allophone states, with about 50% of the exemplars in each state at any given time. Analogical error adds a bias toward conformity to generalizations over the set of exemplars, and this causes the exemplars in the basic model to all pattern together. At any one time, most of the exemplars are in the same allophone state, though the system as a whole does alternate between one allophone and the other. So

random variability in production becomes entrenched in the system as random alternation between categorical states, comparable to a language with a single coronal fricative undergoing sound changes that change the default allophone from alveolar to postalveolar and back again.

When production biases dependent on phonetic context are added into the system, these become entrenched as regular contextual patterns. In a system with two kinds of contexts, where one context favors one allophone (e.g. [ʃ] in [_i]) and another context favors the other allophone (e.g. [s] in [_a]),[7] the system has three states: complementary distribution, where each allophone is found in the context that prefers that allophone; total neutralization to one allophone ([s]); and total neutralization to the other allophone ([ʃ]). The phonetic biases in each production event are constant and dependent only on the phonetic context of that exemplar, but the positive feedback in the system causes their effects to accumulate differently depending on the states of the other exemplars.

The model that shows non-cumulativity of the phonetic biases includes three such biases: two independent biases towards one allophone (e.g. [ʃ] in [_i] and [ʃ] in [i_]), and one bias towards the other allophone in all contexts (e.g. [s] over [ʃ]). Under direct grounding approaches, biases like these are assumed to lie behind typologies like the contextual licensing data in Table 6, but again these biases are constant error rates, not constraints that have different weights in different languages. In this system, when the biases are set such that their ratios would interact cumulatively in a constraint system, the behavior is dependent on the distribution of context types in the lexicon. With a lexicon in which the phoneme is only found in single contexts, like {[_i],[i_]}, the context-free bias that prefers [s] dominates the system, and the only stable state is total neutralization to [s]. With a lexicon in which the phoneme is only found in combined contexts, like [i_i], the contextual biases that prefer [ʃ] dominate the system, and the only stable state is total neutralization to [ʃ]. When the lexicon is more diverse, the stable state is apparently dependent on the average sum of biases over the lexicon. When there are more single contexts than combined contexts, the system gravitates towards [s], and when there are more double contexts than single contexts, the system gravitates towards [s].

However, both of these stable states involve total neutralization across the lexicon, not contextual regularity of any kind. When the single contexts dominate the system, the whole lexicon shifts to the [s] state, and when the double contexts dominate the system, the whole lexicon shifts to the [ʃ] state. Not only is the double-side licensing (the cumulative pattern)

[7] Wedel introduces the allophony problem using the example of alveolar 'light' [l] versus velar 'dark' [ɫ], and discusses these systems primarily in abstract terms, as $+1, -1$ allophones of a category x, in contexts $\{a, b, c, \dots\}$. I have adapted the examples to coronal palatalization, for concreteness and integration with the previous discussion.

not possible, but neither are left-side and right-side licensing. In my own implementation of the model, lexicons that include exemplars where neither of the contextual biases apply ({[_a],[a_a]}) show that the state that might seem to be the dual-side licensing is actually total neutralization to [ʃ], since exemplars with these general contexts also pattern with the rest of the lexicon. In this model it is also more difficult than in the simpler models to create a system that can change from one state to another. It is still possible, but it requires more careful balancing of the different biases.

The emergence of categorical behavior from positive feedback over quantitative biases is striking, especially in a system that is schematically rather simple. This simulation provides a suggestion for why strict domination might be a generally useful assumption, but it does not actually show how a system could have the possibility of phonological alternations and yet reduce cumulative patterns. The simpler simulation demonstrates phonological alternations but doesn't have biases that can interact cumulatively, and the more complex simulation that demonstrates non-cumulative interaction can't produce phonological alternations, much less an occasional cumulative alternation pattern.

5.2 Bayesian learning bias

Further suggestive insights come from a Bayesian model of learning and production, where the effects of similarity biases comparable to those in Wedel's models can be more easily calculated. The learner's expectations during language learning, even weak ones, can strongly influence the resulting typological distribution, and similar results follow regarding biases in language use and interpretation.

Iterated Learning Models (Kirby, Smith, & Brighton, 2004) derive typological generalizations from the dynamics of language learning in the context of sociocultural and sociobiological evolution. For example, an early investigation (Kirby, 1999) developed models of how the implicational hierarchy of word order in noun phrases (e.g. in prepositional languages, nouns preceding their modifying adjectives implies that nouns precede their modifying genitives) could be seen as metastable states in the space of grammars, quasi-optimal because of the opposition between speaker economy and communicative success. Griffiths and Kalish (2007) provide a Bayesian analysis of iterated learning, showing that the learners' learning bias (their 'expections' prior to exposure to language) is essential to predicting the distribution of languages that is stable under iterated learning. Learners observe language data and produce a hypothesized language model by considering how likely it is that each possible hypothesis produced the given data, hedging their bets based on their prior expectations about possible language models. Assuming that learners maintain uncertainty in their productions

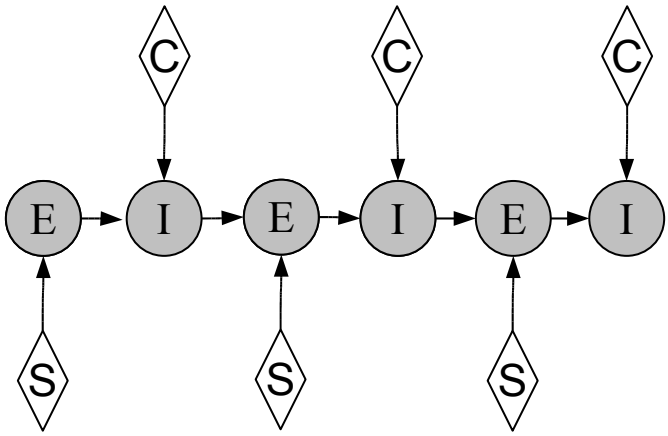


Figure 3: Influence network of language evolution under learning bias and utterance selection. The learning process induces a language model (I) from empirical language data (E), under a prior probability composed of cognitive biases (C). The production process generates E-language data from an I-language model, in concert with a sociocultural probability (S) over utterances, which is composed of the aggregated selection biases of shared experience and communicative and social goals.

and generalizations, iterated learning filters out any effect of the initial learning data, and assuming that any other biases in the induction-production loop are negligible, the final evolutionarily stable distribution approximately matches the distribution of the learners' prior expectations.

The model that Griffiths and Kalish use emphasized the role of cognitive bias, but Kirby, Dowman, and Griffiths (2007) extend the analysis to include a probability over meanings, one kind of functional requirement of language in use, and they show that the distribution of meaning frequencies is responsible for the morphological irregularity of frequent words, fairly independent of learner's expectations.

The Bayesian analysis of iterated learning is symmetric between language representation and language data, so biases in the generation of language data also shape the evolutionarily stable distribution. Just as the learning bias acts to subtly reshape the language model hypothesis space (I) during learning (Fig 3), utterance selection mechanisms shape the language data space (E). Even though this model still oversimplifies the system—such as not including the influence that our cognitive models of social context have on learning or the influence that physiological constraints have on production—it does provide a formal entry point for many of the functional pressures on language.

In a weighted constraint grammar, the extent to which the learner trusts idiosyncratic data versus generalizes beyond the data is controlled by implicit expectations about the distribution of constraint weights. And though the strength of this generalization bias does not influence the abundance of cumulativity, the shape of it does. The literature on learning Harmonic Grammars and training related industrial models (maximum entropy models; Berger, Della Pietra, & Della Pietra, 1996) suggests several different hypotheses about the shape of the generalization bias. Though some learning models do set different generalization biases for different constraints (Wilson, 2006; Jesney & Tessier, 2010) or make sets of constraint weights correlated (Riezler & Vasserman, 2004), I assume here that all constraint weights are independently drawn from the same distribution. The probability of a set of constraint weights is thus the product of the probabilities of the individual weights, and the weight space that the probabilities are most easily defined over is measured by the magnitudes of the weights, not by their ratios.

The most common weight distribution in maximum entropy language models is a normal (Gaussian) distribution centered around zero (Hoerl & Kennard, 1970; Chen & Rosenfeld, 1999), which represents the expectation that the weights have a finite variance and a mean of zero. The Maximum Entropy Phonotactic Learner (Hayes & Wilson, 2008) adapts the normal distribution to Maximum Entropy Grammar by removing the half of the distribution that falls on the negative side of zero, resulting in what is known as a half-normal distribution (Fig. 4 top). The probability of a set of weights is a function of the sum of the squares of the weights, or equivalently, a function of the distance from the origin. As a result, in the weight space for two constraints, the probability mass forms a circular hill around the origin, where the contours of equal probability are concentric arcs, and the surfaces of equal probability in the weight space for three constraints consist of sections of concentric spherical shells.

Another common weight distribution in maximum entropy language models is a Laplace distribution (Tibshirani, 1996), which consists of two exponential distributions back to back, and it represents the expectation that the magnitudes of the weights have a finite sum. Compared to the normal distribution, the Laplace distribution has a bit less probability mass near median magnitudes, and more probability mass near zero and very large magnitudes. Because of this, learning a model using a Laplace distribution tends to reduce redundancy in the active constraint set, by setting the weights of some of the redundant constraints to zero. If we remove the half of the distribution that falls below zero, the result is an exponential distribution (Fig. 4 mid). The probability of a set of weights is a function of the sum of the magnitudes of the weights. As a result, the weight space for two constraints has the probability mass in a sharply peaked and gently sloping hill, where the contours of equal

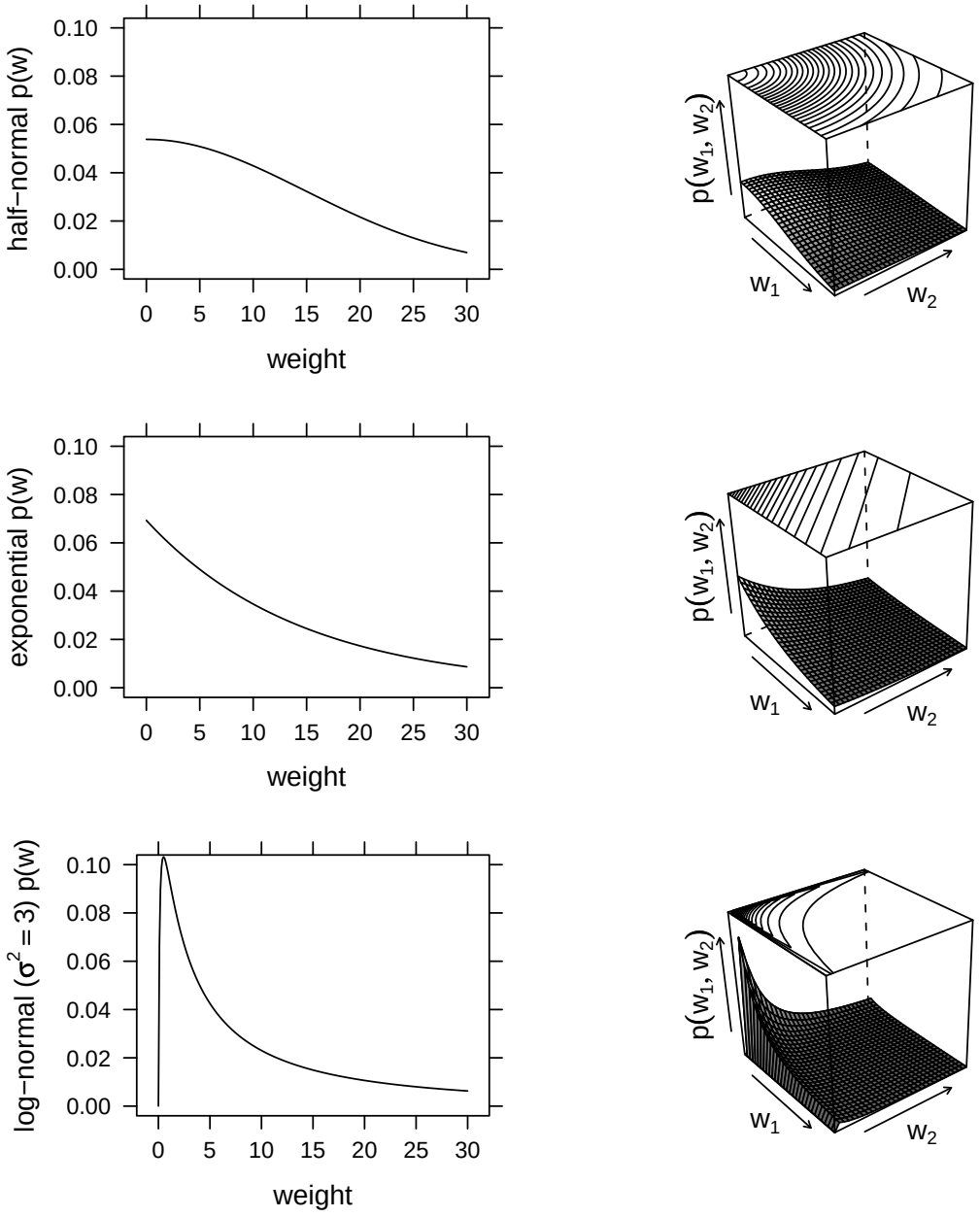


Figure 4: Probability density functions in one dimension (left) and two dimensions (right) for the half-normal distribution (top), exponential distribution (mid), and log-normal distribution (bottom). Lines of equal probability are projected onto the ceiling of the weight space plots. The median of each distribution is set to 10.

Table 7: Language type probabilities according to expected weight distribution, compared to observed distribution of coronal fricative palatalization type

Type Dist:	Half-normal	Exponential	Log-normal	Observed
Full Contrast	41.7%	37.5%	35.4%	37.0%
Special-case Neutralization	8.3%	12.5%	14.6%	8.2%
General-case Neutralization	25.0%	12.5%	6.2%	0.5%
Complementary Distribution	8.3%	12.5%	14.6%	10.3%
Total Neutralization	16.7%	25.0%	29.2%	44.0%

probability are parallel straight lines, and for three constraints, surfaces of equal probability are parallel planes.

Exponential Noisy Harmonic Grammar (Boersma & Pater, 2008), assumes that uncertainty in each constraint's weight is distributed log-normally. A log-normal distribution appears much like a normal distribution when the variance parameter[8] is small, but when the variance parameter is large, it is quite skewed, with a sharp peak at the low end and a long tail at the high end. Figure 4 bottom shows a log-normally distributed weight probability with a relatively high variance parameter, along with the corresponding weight space for two constraints. Under a log-normal probability, a weight equal to zero is impossible, but the highest probability region lies where both weights are close to zero.

In the limiting case where other biases in the system are negligible, the stable typological distribution would have weight configurations sampled randomly from the distribution that characterizes the generalization bias. Therefore the predicted distribution of constrast types can be estimated by sampling weight configurations from the distribution of the assumed generalization bias, and then categorizing the weight configuration according to their ratios. The probabilities esimated by this method are shown in Table 7.

The type probabilities under the exponential distribution are actually equal to the type probabilities under a uniform distribution over weight ratios (§4.3), because the ternary plot is drawn in the plane where the three weights sum to one, and this is one of the planes of equal probability under the exponential distribution. As shown in the plots of Figure 4, the lines of equal probability for the half-normal distribution and the log-normal distribution curve in opposite directions, with the half-normal distribution pushing probability mass towards the center where the weights are comparable, and the log-normal distribution pushing probability mass away from the center. In the plane of the ratios among the weights, this means that

[8]The 'variance' parameter is actually the variance of the underlying normal distribution, not the variance of the log-normal distribution

the half-normal distribution pushes languages towards the cumulative patterns, and the log-normal distribution pushes languages away. Under the half-normal distribution, general-case neutralization is predicted to be more common than the other alternation types and even more common than total neutralization. Under the log-normal distribution, general-case neutralization is predicted to be less common than any of the other types, but (with the given variance anyway) it is still not as uncommon as in the observed distribution.

I previously argued (Carroll, 2010) that learner tendencies to regularize unpredictable or meaningless variation (Hudson Kam & Newport, 2009) indicated that the generalization bias in learning must be shaped something like this log-normal distribution, but as I discuss in the next section, this kind of regularization apparently results from a bias in production and parsing under uncertainty, not learning per se. Not only is there no particular evidence that learners use a log-normal bias or something like it, but the evidence we do have about generalization behaviors suggests that learning is less constrained than even these distributions. The convention to restrict constraint weights to positive values is motivated by typology, since allowing grounded constraints to take negative values results in typologically strange effects (Potts et al., 2010). However, though they may be more difficult than phonetically natural patterns, people can learn arbitrary and unnatural patterns in the laboratory (Pycha, Nowak, Shin, & Shosted, 2003; Seidl & Buckley, 2005), and unnatural patterns do appear in natural languages (Blevins, 2008; Mielke, 2008; Morley, 2011). Though the generalization bias may still disprefer cumulative patterns, its effects must be quite weak.

5.3 Bayesian production bias

Parallel to the learning bias, which favors generalization and robust representations rather than idiosyncratic exceptionalism, a bias in production and comprehension favors predictability rather than meaningless variation (Hudson Kam & Newport, 2009). In the psycholinguistics literature, this has become known as regularization, and the effects of regularizing behavior probably do contribute to what linguists call regularization of grammatical patterns, but they are distinct. Regularizing behavior in language processing is primarily not an issue of generalized long-term representations, but rather an effect of working memory limitations (Hudson Kam & Chong, 2009).

Regularizing behavior is a compromise between two rational strategies: frequency maximizing and frequency matching (Gaissmaier & Schooler, 2008). For example, if 80% of the people in a certain region call a certain kind of snack a 'biscuit', while the rest of the people call it a 'cookie', effective communication can be achieved 80% of the time by maximizing (always assuming it is a 'biscuit'), whereas frequency matching (randomly guessing 'biscuit'

80% of the time) will lead to success only $80\% \times 80\% + 20\% \times 20\% = 68\%$ of the time. So the simple strategy of selecting the most frequent possibility is advantageous for the immediate task. However, if there is structure to the variation (such as if the 'biscuit' people tend to pronounce their vowels differently), then someone using the maximizing strategy has little opportunity to test hypotheses about the use of 'cookie', while someone using the probability matching strategy can do so, learning the structured variation more quickly.

The general trend on both linguistic and non-linguistic tasks is that when cognitive resources are available and well-adapted to the task, matching is favored, but when resources are more limited, maximizing is favored. Experiments with split brain patients (Wolford, Miller, & Gazzaniga, 2000; Miller & Valsangkar-Smyth, 2005) show asymmetries between left and right sides of the brain: for many sequence tasks, the left brain frequency matches and the right brain maximizes, while for predicting sequences of faces (a task that is right lateralized), the left brain maximizes and the right brain frequency matches. Young children tend to maximize but shift towards matching as they mature (Ramscar & Gitcho, 2007). Adults generally use a frequency matching strategy but will shift towards a maximizing strategy either when distracted by another task or after significant training (Shanks, Tunney, & McCarthy, 2002; Hudson Kam & Chong, 2009). And even when seeming to frequency match, adults are still subtly regularizing, an effect that only becomes noticeable after accumulating over several 'generations' of learners (Reali & Griffiths, 2009).

In variationist constraint grammars like Maximum Entropy Grammar (Hayes & Wilson, 2008) and Noisy Harmonic Grammar (Boersma & Pater, 2008), there is meaningless or unpredictable variation along the boundaries between language types. Regularizing production behavior tends to make the language data less variable, pushing languages away from these boundaries. Because cumulative language types lie in the center of the space of weight ratios, they are bordered by multiple type boundaries. As a result, regularizing behavior tends to reduce cumulative language types. However, the predictions of this model are not consistent with typology more generally.

A generic Harmonic Grammar tableau can be extended into a Maximum Entropy Grammar tableau (13) by exponentiating the costs to produce unnormalized probabilities ψ, which are then converted to proper probabilities p by dividing by their sum Z.

(13) Variation in a Maximum Entropy Grammar

weights	20	10			
/ʃa/	F: IDENT	G: NO[ʃ]	H	ψ	p
☞ ʃa		*	-10	$2^{-10} \approx 10^{-3}$	0.999
sa	*		-20	$2^{-20} \approx 10^{-6}$	0.001
				$Z \approx 1.001 \times 10^{-3}$	

If we generalize to arbitrary weights, but keep the simple violation profile, one candidate violates constraint C_1 which has weight w_1, and the other candidate violates constraint C_2 which has weight w_2. This example is also assuming that any other candidates have negligible probability due to violating some other high-weight constraint, and any other constraints violated by just one of these two candidates have negligible weight. The Maximum Entropy Grammar probabilities for this decision are determined by a single parameter $t = w_1 - w_2$, the logarithm of the odds ratio of the two outputs.

(14) Equal violations binary decision

weights	w_1	w_2			
/in/	C_1	C_2	H	ψ	p
out_1	*		$-w_1$	2^{-w_1}	$(1 + 2^{w_1 - w_2})^{-1}$
out_2		*	$-w_2$	2^{-w_2}	$(1 + 2^{w_2 - w_1})^{-1}$
				$Z = 2^{-w_1} + 2^{-w_2}$	

The probability $p(\text{out}_1) = 2^{-w_1}/(2^{-w_1} + 2^{-w_2})$ is simplified by dividing the numerator through, and after making the substitution $t = w_1 - w_2$, we have $p(\text{out}_1) = 1/(1 + 2^t)$ and $p(\text{out}_2) = 1/(1 + 2^{-t})$.

Figure 5 left shows lines of equal output probability (characterized by the parameter t) passing diagonally through the weight space for the two weights under an exponential distribution, close to the origin where variation is non-negligible. Along the line $t = 0$, $p(\text{out}_1) = p(\text{out}_2) = 0.5$, maximum uncertainty about the output. Along line $t = 2$, $p(\text{out}_1) = \frac{1}{1+2^2} = \frac{1}{5} = 0.2$, so the output is most likely out_2. Regularization tends to push languages away from the $t = 0$ line, decreasing the probability of weight configurations in that region.

The model of regularization used by Reali and Griffiths (2009) assumes that production uses a prior expectation over the relative frequencies of variants, and that this prior expectation causes people to hedge toward more predictability. The particular prior expectation they assume is a symmetric beta distribution, characterized by the parameter α, where a

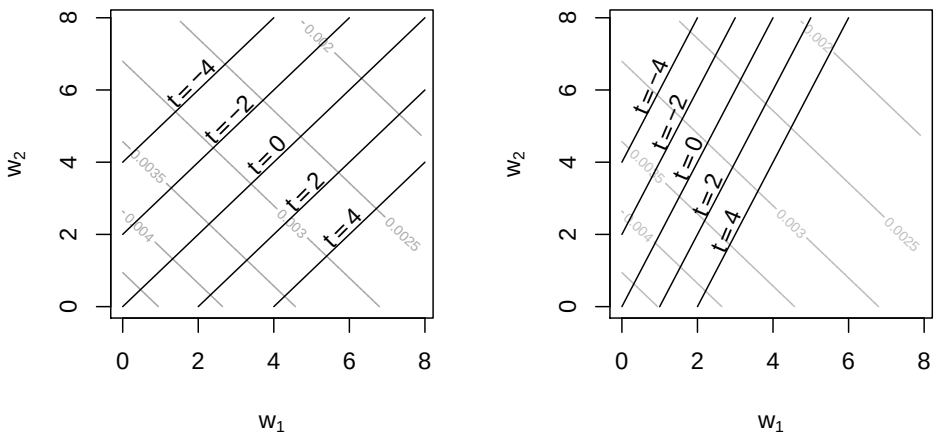


Figure 5: Variation in the simple binary decision problem, for the equal violation case (left) and unequal violation case (right). Regularizing behavior disfavors weight configurations near $t = 0$, where each outcome has $p = 0.5$

value of 1 corresponds to an unbiased (matching) prior, and lower values correspond to regularizing priors, down to 0 which creates a fully maximizing prior.

In order to understand how regularizing might influence the typological probability of cumulative languages, I simulated the type probabilities in a model where the typological distribution is a product of the learning bias and the regularizing bias. Of the four tableaux that characterize the patterns, three of them have variation—inputs /ʃa/, /si/ and /ʃi/ in the contrast typology, and inputs /asi/, /isa/ and /isi/ in the licensing typology. Each of these three tableaux produce a variant probability, and each variant probability is independently evaluated under the beta distribution. The (unnormalized) probability is the product of the three beta probabilities and the probability under the learning bias. Examples of the product of the beta probabilities are shown in Figure 6, where the contour lines show lines of equal unnormalized probability. The upper plot shows the distribution for the contrast typology, and the lower plot shows the distribution for the licensing typology. Each one is rotated compared to the ternary plots, such that the upper corner in the ternary plots (associated with constraint G) corresponds to the lower left corner in these plots. The other two weights correspond to the x and y dimensions, and the three weights sum to 20. The probabilities are calculated assuming that $\alpha = 0.7$ and that residual uncertainty (equivalent

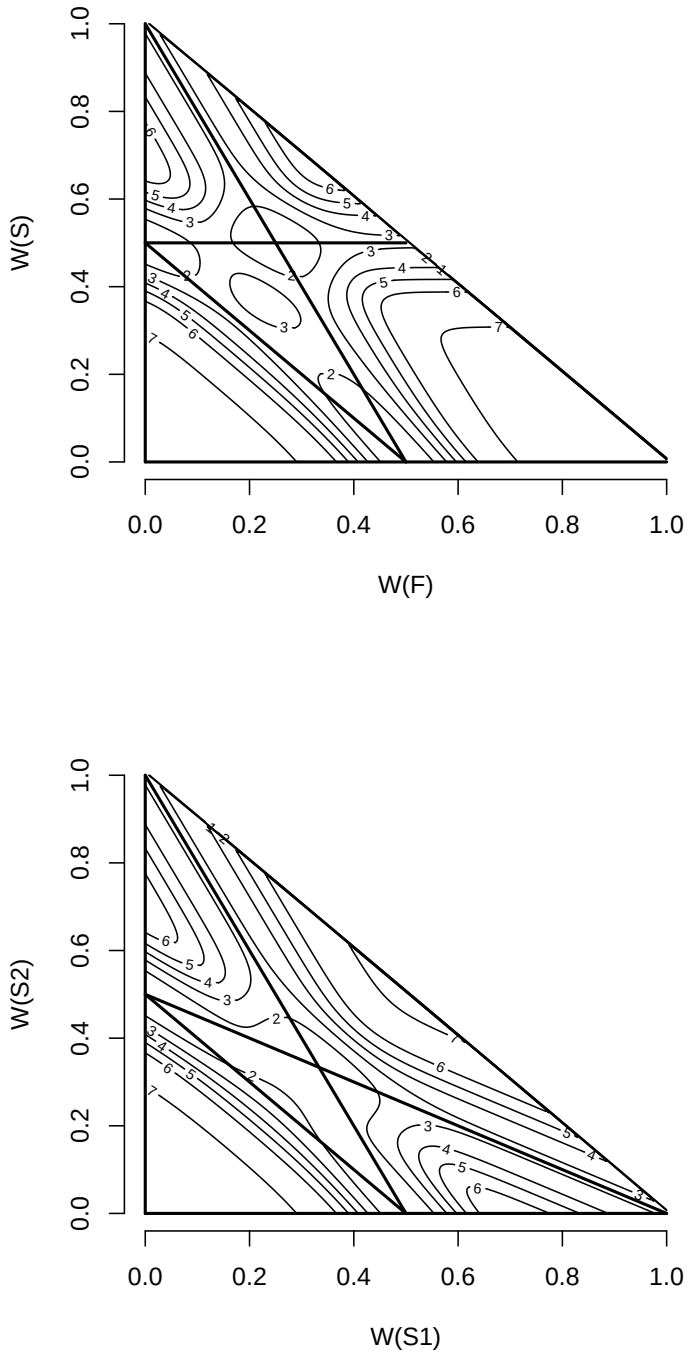


Figure 6: Probability density functions showing the influence of a variation reduction bias in the general-case neutralization typology (top) and the contextual conditioning typology (bottom). The regions near the boundaries between language types have relatively low probability.

Table 8: Estimated probability distribution under a beta regularization bias, using an exponential distribution for the learning bias

Contrast Type	$\alpha = 0.3$	$\alpha = 0.7$	$\alpha = 1.0$	Observed
Full Contrast	35.8%	35.9%	37.5%	37.0%
Special-case Neutralization	7.5%	10.5%	12.5%	8.2%
General-case Neutralization	7.5%	10.4%	12.5%	0.5%
Complementary Distribution	7.6%	10.8%	12.5%	10.3%
Total Neutralization	41.6%	32.3%	25.0%	44.0%

Licensing Type	$\alpha = 0.3$	$\alpha = 0.7$	$\alpha = 1.0$	Observed
No alternation	25.1%	25.9%	25.0%	76.3%
Right side	18.2%	18.1%	16.7%	19.2%
Left side	20.0%	17.1%	16.7%	1.5%
Dual side	31.6%	32.2%	33.3%	2.0%
Double side	5.1%	6.6%	8.3%	0.0%

to background variation of $\alpha/50$) keeps the probabilities finite.

The probabilities in the regions along the typology boundaries are reduced and these regions do extend into the cumulative regions. However, in the contextual neutralization typology (Fig. 6 top) the regularization reduces the general-case neutralization about the same amount as the special-case neutralization and complementary distribution, so that if this effect were strong enough to nearly eliminate general-case neutralization, there would be nearly no alternations at all, leaving only total neutralization and full contrast. Table 8 shows estimated probability distributions for the product of an exponential learning bias and the regularizing production bias, using an exponential distribution with a median of 10. The column with $\alpha = 1.0$ is calculated analytically, since it is equivalent to the bare exponential learning bias, while the columns for the other two values of α are calculated by numeric integration, sampling from the exponential distribution and multiplying by the calculated regularization bias. In the contextual licensing typology (Fig. 6 bottom), the regularizing bias can reduce the abundance of the cumulative pattern (double side) without reducing the other patterns (Tab. 8 bottom). However, the cumulative pattern already has a lower probability than the other types without the regularizing bias, and the relationship between the predicted probabilities and observed probabilities has more serious issues than the abundance of the cumulative pattern: for both the palatalization type counts in the table and the nasalization type counts mentioned previously, the no alternation count is overwhelmingly more likely. But even if it is appealing to invoke the regularizing bias to explain the unattested (so far) double side licensing in palatalization, that leaves unexplained

why the double side licensing is relatively well-attested in nasalization.

6 Future directions

Future research will continue the two branches of research that led up to this point: analyzing typological data, and quantitative modeling of particular hypotheses.

6.1 Typological data

The current data set is too small and too restricted to judge the reliability of many of the generalizations about typological probabilities that I have suggested here. As discussed previously, the counts of language families are too limited to obtain much statistical power. In addition, this paper's focus on the contrast types in palatalization leave open the question of whether contrast in other phonetic features or typologies with other kinds of constraints have more (or fewer) attestations of cumulative pattern types than observed in coronal palatalization.

So one priority is to expand the breadth of the data. I had previously started to compile data sets for vowel nasalization and NC effects, and I am also considering lenition effects.

a) Vowel nasalization: The vowel nasalization data (Table 6) is parallel in many ways to the palatalization data, but it has attested cumulative patterns in the licensing typology and (so far) does not have cumulative patterns in the contrast typology, the opposite of palatalization.

b) Lenition effects: The lenition effects (e.g. intervocalic spirantization) that Kirchner (1998) discusses are very comparable to the contextual conditioning typologies of palatalization and nasalization. The analytic mechanisms that Kirchner uses incorporate the additivity of the contexts into single constraints that participate in a ranked constraint hierarchy, but the distributional patterns seem to be equivalently expressible in terms of weighted constraints.

c) NC effects: A wide variety of alternations, conditioned by potential nasal+consonant clusters, avoid the structure of a nasal stop followed by a consonant that has a different oral closure or different voicing type. NC effects have played a central role in the debate over the grounding of naturalness, with some taking it as evidence for direct synchronic grounding (Hayes, 1999; Hayes & Stivers, 2000; Halpert, to appear), and others taking it as evidence against direct grounding (Hyman, 2001; Myers, 2002; Blust, 2004).

In addition, Farris-Trimble (2008) has argued that some NC effects are cumulative patterns.

The vowel nasalization data is relatively simple, in many ways simpler than the palatalization data, while the lenition effects data would be more complicated. The NC effects would be even more complicated, though the task would be made easier by a few typological survey studies (Hyman, 2001; Blust, 2004; Halpert, to appear).

A second priority is to extend the depth of the typological data. The 198 languages in the current palatalization sample are only a third of the languages in the P-base database of sound patterns, and there are other languages that have phonological documentation available but are not included in P-base yet.

A third priority is to make use of recently developed statistical methods that use all levels of the linguistic family trees for improved genealogical control (Bickel, 2008), and regression models that better distinguish biases within families and areas from cross-linguistic biases (Bickel, to appear).

6.2 Markov chain analysis of evolutionary models

In his closing chapter, Mielke (2008) outlines his vision of an emergentist approach to typology. Emphasizing that an understanding of emergent patterns in language necessarily requires a diachronic approach. Citing the suggestion of Greenberg (1978) that diachronic changes between linguistic types can be viewed as transitions in a state and process model, Mielke develops a comprehensive view of how the 'filters' on phonology (Hume & Johnson, 2001a) can be modeled using Markov chain analysis methods. Languages or languages types represent state nodes within a network of connected nodes, and languages pass from state to state probabilistically, subject to the functional biases of the 'filters'. Analytic and computational techniques exist for estimating the transitional probabilities and the long term probabilities of find a language in any particular state.

The analysis method is particularly applicable to exemplar models like Wedel's. In an exemplar model with just two system states, the behavior of the system can be characterized by a mixing matrix π, where the terms of the matrix specify the probabilities of staying in each state or transitioning to any other at each time step.

$$\pi = \begin{bmatrix} 1 - \pi_{AB} & \pi_{AB} \\ \pi_{BA} & 1 - \pi_{BA} \end{bmatrix}$$

The term π_{AB} is the probability that a system in state A will change to state B, and since the only alternative to changing to state B is staying in state A, the probability of staying

in state A is $1 - \pi_{AB}$. Similarly, the term π_{BA} is the probability of a system in state B changing to state A, and the probability of staying in state B is $1 - \pi_{BA}$. If there are 100 identical systems each described by this mixing matrix, and initially 40 of them are in state A and 60 of them are in state B, then after one time step, we would expect to find $40 \times (1 - \pi_{AB}) + 60 \times \pi_{BA}$ in state A and $40 \times \pi_{AB} + 60 \times (1 - \pi_{BA})$ in state B. In the stable distribution of system states, the number of systems in A and B remains the same after each time step, so the numbers (x_A and x_B) are described by the system of equations

$$\left\{\begin{array}{ccc} x_A(1-\pi_{AB}) & + & x_B\pi_{BA} = x_A \\ x_A\pi_{AB} & + & x_B(1-\pi_{BA}) = x_B \end{array}\right\}$$

Either of these equations can be rearranged as

$$\frac{x_A}{\pi_{BA}} = \frac{x_B}{\pi_{AB}}$$

In other words, the long term probability of finding a particular system in a particular state is proportional to the step-wise probability of transitioning into that state and inversely proportional to the step-wise probability of transitioning out of it. The same analysis can be performed for systems of three or four states, though the equations are more complicated. The corresponding equation for a system of three states is

$$\frac{x_A}{\pi_{BA}\pi_{CB} + \pi_{BA}\pi_{CA} + \pi_{CA}\pi_{BC}} = \frac{x_B}{\pi_{AB}\pi_{CA} + \pi_{AB}\pi_{CB} + \pi_{AC}\pi_{CB}} = \frac{x_C}{\pi_{AC}\pi_{BA} + \pi_{AC}\pi_{BC} + \pi_{AB}\pi_{BC}}$$

In this system, it is still true in a general sense that rates of transition into a state tend to increase the probability of that state, and rates of transition out of that state tend to decrease the probability of that state. For example, two out of the three times that π_{BA} appears, it is in the term that increases x_A, and all three times, it is in terms that decrease x_B. But the probability of any state is a function of all three terms, dependent on all of the transition rates, even for the transitions between the other two states.

Since the error biases in Wedel's models are dependent only on the current states of the examplars, and these are approximately constant for each state of the system, the transition probabilities can be approximated by a function of the error biases. This should lead to analytic results relating the error biases and composition of the lexicon to the typologically stable distribution. However, even expressed directly in terms of the transition probabilities, the solution of the system of equations is rather complicated for systems of three and four states, and there is no analytic solution for systems of more than four states. The approximations necessary to express the transition probabilities as a function of the system state

alone might make the interpretation of the solution somewhat opaque. In addition, it may be impossible to examine the full typologies of contextual contrast and conditioning, since they each have five types.

6.3 Conclusions

Attestations of contrast types in the palatalization typology demonstrate that the general-case neutralization pattern exemplified by Gujarati fricatives (Pandit, 1954) is underrepresented. However, issues raised by this case are not limited to general-case neutralization or palatalization. General-case neutralization is just one kind of cumulative pattern, and double-side contextual conditioning is another. Similarly the palatalization data is just one phenomenon that shows these two typologies, while others include vowel nasalization and consonant lenition. The shortage of cumulative patterns is more general than palatalization contrasts, though perhaps not extending to all phonological pattern types.

A better understanding of the distribution of cumulative patterns will come hand-in-hand with better models of the distinctions between learned synchronic grammars and typological variation, and better models of how phonetic and cognitive biases influence typology. Some of the path forward will involve more careful models of language change, but much must also come from better characterizing the phenomena and pattern types that are more or less disposed to producing cumulative patterns.

References

Alderete, J. (2008). Using learnability as a filter on factorial typology: A new approach to Anderson and Browne's generalization. *Lingua*, *118*, 1177–1220.

Anderson, S. R. (1981). Why phonology isn't "natural". *Linguistic Inquiry*, *12*(4), 493–539.

Anttila, A. (1997). Deriving variation from grammar. In F. Hinskens, R. van Hout, & L. Wetzels (Eds.), *Variation, change and phonology theory* (pp. 35–68). John Benjamins.

Bach, E., & Harms, R. T. (1972). How do languages get crazy rules? In R. P. Stockwell & R. K. S. Macaulay (Eds.), *Linguistic change and generative theory* (pp. 1–21). Bloomington: Indiana University Press.

Bakovic, E. (2000). *Harmony, dominance and control*. Unpublished doctoral dissertation, Rutgers University.

Bane, M., & Riggle, J. (2008). Three correlates of the typological frequency of quantity-insensitive stress systems. In *Proceedings of the 10th ACL SIGMORPHON.*

Bateman, N. (2007). *A crosslinguistic investigation of palatalization.* Unpublished doctoral dissertation, UC San Diego.

Bateman, N. (2011). On the typology of palatalization. *Language and Linguistics Compass*, *5*(8), 588-602.

Beddor, P. S., & Krakow, R. A. (1999). Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation. *Journal of Acoustic Society of America*, *106*(5), 2868–2887.

Berent, I., Steriade, D., Lennertz, T., & Vaknin, V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition*, *104*, 591–630.

Berger, A. L., Della Pietra, S. A., & Della Pietra, V. J. (1996). A maximum entropy approach to natural language processing. *Computational Linguistics*, *22*(1), 39–71.

Bickel, B. (2008). A refined sampling procedure for genealogical control. *Sprachtypologie und Universalienforschung*, *61*(3), 221–233.

Bickel, B. (to appear). Statistical modeling of language universals. *Linguistic Typology*.

Blevins, J. (2004). *Evolutionary Phonology: The Emergence of Sound Patterns.* Cambridge University Press.

Blevins, J. (2008). Natural and unnatural sound patterns: A pocket field guide. In *Naturalness and iconicity in language.* John Benjamins.

Blust, R. (2004). Austronesian nasal substitution: A survey. *Oceanic Linguistics*, *43*(1), 73–148.

Boersma, P. (2008). *Emergent ranking of faithfulness explains markedness and licensing by cue.* ROA-954.

Boersma, P., & Hamann, S. (2008). The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology*, *25*, 217-270.

Boersma, P., & Pater, J. (2008). *Convergence properties of a gradual learning algorithm for harmonic grammar.* ROA-970.

Carroll, L. (2010). *Cumulativity in constraint grammars: An iterated learning view of probabilistic typology.* (UC San Diego)

Chen, S. F., & Rosenfeld, R. (1999). *A Gaussian prior for smoothing maximum entropy models* (Tech. Rep. No. CMU-CS-99-108). Carnegie Mellon Unviersity.

Coetzee, A. W. (2002). *Between-language frequency effects in phonological theory.* ROA-973.

Cohn, A. C. (1993). Nasalisation in English: Phonology or phonetics. *Phonology*, *10*(1), 43–81.

Cristófaro-Silva, T. (2003). Palatalisation in Brazilian Portuguese. In S. Ploch (Ed.), *Living on the edge: 28 papers in honour of Jonathan Kaye* (Vol. 62). Mouton De Gruyter.

de Lacy, P. (2010). Phonological evidence. In S. Parker (Ed.), *Phonological argumentation:*

Essays on evidence and motivation. Equinox.

Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology*, *26*(6), 1355–1367.

Dryer, M. S. (1992). The Greenbergian word order correlations. *Language*, *68*, 81–138.

Eckert, P., & Rickford, J. (Eds.). (2001). *Style and sociolinguistic variation.* Cambridge University Press.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*, 429–492.

Farris-Trimble, A. W. (2008). *Cumulative faithfulness effects in phonology.* Unpublished doctoral dissertation, Indiana University.

Flemming, E. (2001). Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology*, *18*, 7–44.

Gafos, A. I. (1999). *The articulatory basis of locality in phonology.* Garland Publishing.

Gafos, A. I., & Benus, S. (2006). Dynamics of phonological cognition. *Cognitive Science*, *30*, 905–943.

Gaissmaier, W., & Schooler, L. J. (2008). The smart potential behind probability matching. *Cognition*, *109*, 416–422.

Garrett, A., & Johnson, K. (in press). Phonetic bias in sound change. In A. C. L. Yu (Ed.), *Origins of sound change: Approaches to phonologization.* Oxford University Press.

Goldstein, L., & Fowler, C. A. (2003). Phonetics and phonology in language comprehension and production. In N. O. Schiller & A. S. Meyer (Eds.), (pp. 159–207). Mouton de Gruyter.

Gordon, M. (2002). A factorial typology of quantity-insensitive stress. *Natural Language and Linguistic Theory*, *20*, 491–552.

Gordon, M. (2007). Typology in optimality theory. *Language and Linguistics Compass*, *1*(6), 750–769.

Gordon, M., Barthmeier, P., & Sands, K. (2002). A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association*, *32*(2), 141–174.

Gordon, R. G. J. (Ed.). (2005). *Ethnologue: Languages of the world.* Dallas, TX: SIL International.

Greenberg, J. H. (1978). Diachrony, synchrony, and language universals. In *Universals of human language: Word structure.* Stanford University Press.

Griffiths, T. L., & Kalish, M. L. (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive Science*, *33*(3), 441-480.

Hall, T. A. (1995). Inalterability and sibilant neutralization in Gujarati. *Lingua*, *98*, 11–20.

Halle, M., & Mohanan, K. P. (1985). Segmental phonology of modern english. *Linguistic Inquiry*, *16*(1), 57–116.

Halpert, C. (to appear). Overlap-driven consequences of nasal place assimilation. In *Consonant clusters and structural complexity.* Mouton De Gruyter.

Hamann, S., & Avelino, H. (2007). An acoustic study of plain and palatalized sibilants in Ocotepec Mixe. In *Proceedings of ICPhS-2007.*

Hayes, B. (1999). Phonetically-driven phonology: The role of optimality theory and inductive grounding. In M. Darnell, E. Moravscik, M. Noonan, F. Newmeyer, & K. Wheatly (Eds.), *Functionalism and formalism in linguistics, volume 1: General papers.* John Benjamins.

Hayes, B., Kirchner, R., & Steriade, D. (2004). *Phonetically Based Phonology.* Cambridge University Press.

Hayes, B., & Stivers, T. (2000). The phonetics of postnasal voicing. In *Working papers.* UCLA Linguistics.

Hayes, B., & Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*.

Hayes, B., Zuraw, K., Siptár, P., & Londe, Z. (2009). Natural and unnatural constraints in hungarian vowel harmony. *Language*, *85*(4), 822-863.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67.

Hudson Kam, C., & Chong, A. (2009). Investigating the cause of language regularization in adults: memory constraints or learning effects? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 815–821.

Hudson Kam, C., & Newport, A. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, *59*, 30-66.

Hume, E., & Johnson, K. (2001a). A model of the interplay of speech perception and phonology. In E. Hume & K. Johnson (Eds.), *The role of speech perception in phonology.* Academic Press.

Hume, E., & Johnson, K. (Eds.). (2001b). *The role of speech perception in phonology.* Academic Press.

Hyman, L. M. (2001). The limits of phonetic determinism in phonology. In E. Hume & K. Johnson (Eds.), *The role of speech perception in phonology.* Academic Press.

Hyman, L. M. (2008). Universals in phonology. *The Linguistic Review*, *25*, 83–137.

Ito, J., & Mester, A. (2003). Systemic markedness and faithfulness. In *Proceedings of CLS 39.*

Jäger, G., & Rosenbach, A. (2006). The winner takes it all—almost: cumulativity in

grammatical variation. *Linguistics*, *44*(5), 937–971.

Jakobson, R. (1962/1971). *Selected writings: Phonological studies* (Vol. 1; S. Rudy, Ed.). The Hague: Mouton.

Jesney, K. (2009). Licensing in multiple contexts: An argument for harmonic grammar. In *Proceedings of CLS 45.* Cascadilla Press.

Jesney, K., & Tessier, A.-M. (2010). Biases in harmonic grammar: the road to restrictive learning. *Natural Language and Linguistic Theory*, *28*(4).

Kager, R. (2006). Lexical irregularity and the typology of contrast. In *The nature of the word: Essays in honor of Paul Kiparsky.* MIT Press.

Key, M. R. (1968). Phonemic pattern and phoneme fluctuation in Bolivian Chama (Tacanan). *La Linguistique*, *4*, 35–48.

Kirby, S. (1999). *Function, selection, and innateness: The emergence of language universals*. Oxford University Press.

Kirby, S., Dowman, M., & Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *PNAS*, *104*(12), 5241–5245.

Kirby, S., Smith, K., & Brighton, H. (2004). From UG to universals. *Studies in Language*, *28*(3), 587-607.

Kirchner, R. (1998). *An effort-based approach to consonant lenition.* Unpublished doctoral dissertation, UC Los Angeles.

Kochetov, A. (2010). Palatalisation. In C. Ewen, B. Hume, & M. van Oostendorp (Eds.), *Companion to phonology.* Wiley Blackwell.

Ladd, R. (2011). Phonetics in phonology. In J. Goldsmith, J. Riggle, & A. Yu (Eds.), *Handbook of phonological theory* (2nd ed.). Blackwell.

Lee, W.-S., Vakoch, D. A., & Wurm, L. H. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, *25*(5), 527–542.

Legendre, G., Miyata, Y., & Smolensky, P. (1990). *Harmonic grammar – a formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations* (ICS Technical Report No. 90-5). University of Colorado.

Legendre, G., Sorace, A., & Smolensky, P. (2006). The harmonic mind: From neural computation to optimality-theoretic grammar. In P. Smolensky & G. Legendre (Eds.), (Vol. 2, pp. 339–399). MIT Press.

Lindblom, B. (1990). Speech production and speech modeling. In W. J. Hardcastle & A. Marchal (Eds.), (pp. 403–439). Kluwer.

Lindblom, B., Diehl, R., Park, S.-H., & Salvi, G. (2011). Sound systems are shaped by their users: The recombination of phonetic substance. In N. Clements & R. Ridouane (Eds.),

Where do phonological features come from? Cognitive physical and developmental bases of distinctive speech categories. John Benjamins.

Lindblom, B., & Maddieson, I. (1988). Phonetic universals in consonant systems. In *Language, speech and mind.* Routledge, London.

Lubowicz, A. (2005). Locality of conjunction. In J. Alderete (Ed.), *Proceedings of WCCFL 24.*

Marsico, E., Maddieson, I., Coupe, C., & Pellegrino, F. (2004). Investigating the "hidden" structure of phonological systems. In *Proceedings of BLS 30.*

Mielke, J. (2008). *The emergence of distinctive features.* Oxford University Press.

Miller, M. B., & Valsangkar-Smyth, M. (2005). Probability matching in the right hemisphere. *Brain and Cognition*, *57*, 165–167.

Moreton, E., & Pater, J. (2011, July 13). *Formally biased phonology: Complexity in learning and typology.* Handout from the Workshop on Testing Models of Phonetics and Phonology, Linguistics Institute, University of Colorado, Boulder.

Morley, R. (2008). *Generalization, lexical statistics, and typologically rare systems.* Unpublished doctoral dissertation, Johns Hopkins University.

Morley, R. (2011). *How likely are impossible languages? /t/ epenthesis reconsidered.* Ms. Ohio State University.

Myers, S. (2002). *Gaps in factorial typology: The case of voicing of consonant clusters.* ROA-509. (Ms. UT Austin)

Ohala, J. J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick, & M. F. Miller (Eds.), *Papers from the parasession on language and behavior* (pp. 178–203). Chicago Linguistic Society.

Pagotto, E. G. (2004). *Variação e (´) identidade.* Maceió: EDUFAL.

Pandit, P. (1954). Indo-Aryan sibilants in Gujarati. *Indian Linguistics*, *14*, 503–511.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, *24*(2), 175–184.

Pierrehumbert, J. B. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology* (Vol. 7, pp. 101–139). Mouton de Gruyter.

Potts, C., Pater, J., Jesney, K., Bhatt, R., & Becker, M. (2010). Harmonic grammar with linear programming: from linear systems to linguistic typology. *Phonology*, *27*, 77–117.

Prince, A., & Smolensky, P. (1993/2004). *Optimality theory: Constraint interaction in generative grammar.* ROA-537.

Pycha, A., Nowak, P., Shin, E., & Shosted, R. (2003). Phonological rule-learning and its implications for a theory of vowel harmony. In M. Tsujimura & G. Garding (Eds.),

WCCFL 22 proceedings (pp. 101–114).

Ramscar, M., & Gitcho, N. (2007). Developmental change and the nature of learning in childhood. *Trends in Cognitive Sciences*, *11*, 274–279.

Reali, F., & Griffiths, T. L. (2009). The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition*, *111*, 317-328.

Recasens, D., & Espinosa, A. (2007). An electropalatographic and acoustic study of affricates and fricatives in two Catalan dialects. *Journal of the International Phonetic Association*, *37*, 143–172.

Riezler, S., & Vasserman, A. (2004). Incremental feature selection and l_1 regularization for relaxed maximum-entropy modeling. In *Proceedings of EMNLP'04.*

Riggle, J. (2010). *Sampling rankings.* ROA-1075.

Rubach, J. (2003). Polish palatalization in derivational optimality theory. *Lingua*, *113*, 197–237.

Rubach, J. (2006). Mid vowel fronting in Ukrainian. *Phonology*, *22*, 1–36.

Scatton, E. A. (1984). *A reference grammar of modern Bulgarian.* Columbus, Ohio: Slavica Publishers.

Seidl, A., & Buckley, E. (2005). On the learning of arbitrary phonological rules. *Language learning and development*, *1*, 289–316.

Shanks, D. R., Tunney, R. J., & McCarthy, J. D. (2002). A re-examination of probability matching and rational choice. *Journal of Behavioral Decision Making*, *15*, 233–250.

Simpson, A. P. (1999). Fundamental problems in comparative phonetics and phonology: Does UPSID help solve them? In *Proceedings of ICPhS 99.*

Smith, J. L. (2000). Prominence, augmentation, and neutralization in phonology. In L. Conathan, J. Good, D. Kavitskaya, A. Wulf, & A. Yu (Eds.), *Proceedings of BLS 26* (pp. 247–257).

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, *11*, 1–23.

Smolensky, P. (1995). *On the internal structure of the constraint component* Con *of UG.* Handout of talk at UCLA, April 7. ROA-86.

Smolensky, P. (2006). The harmonic mind: From neural computation to optimality-theoretic grammar. In P. Smolensky & G. Legendre (Eds.), (Vol. 2, pp. 27–160). MIT Press.

Smolensky, P., & Legendre, G. (2006). *The harmonic mind: From neural computation to optimality-theoretic grammar.* MIT Press.

Sorace, A., & Keller, F. (2005). Gradience in linguistic data. *Lingua*, *115*, 1497–1524.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, *17*, 3–45.

Subrahmanyam, P. S. (1968). *A descriptive grammar of Gondi.* Annamalainagar: Annamalai

University.

Thakur, P. (2011, July). *Sibilants in Gujarati phonology.* (Poster from LSA Institute Workshop on Information Theoretic Approaches to Linguistics)

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistics Society B*, *58*(1), 267–288.

Tily, H., Gahl, S., Arnon, I., Snider, N., Kothari, A., & Bresnan, J. (2009). Syntactic probabilities affect pronunciation variation in spontaneous speech. *Language and Cognition*, *1*(2), 147–165.

Warner, N., Jongman, A., Sereno, J., & Kemps, R. (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: evidence from Dutch. *Journal of Phonetics*, *32*, 251–276.

Wedel, A. B. (2007). Feedback and regularity in the lexicon. *Phonology*, *24*, 147–185.

Wilson, C. (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science*, *30*, 945–982.

Wolford, G., Miller, M. B., & Gazzaniga, M. (2000). The left hemisphere's role in hypotheses formation. *The Journal of Neuroscience*, *20*(RC64), 1–4.

Yu, A. C. L. (2007). Understanding near mergers: the case of morphological tone in Cantonese. *Phonology*, *24*, 187–214.

Zsiga, E. C. (1995). An acoustic and electropalatographic study of lexical and post-lexical palatalization in American English. In B. Connell & A. Arvaniti (Eds.), *Phonology and phonetic evidence, papers in laboratory phonology IV.* Cambridge University Press.

Zsiga, E. C. (2000). Phonetic alignment constraints: consonant overlap and palatalization. *Journal of Phonetics*, *28*, 69–102.

Zygis, M., & Hamann, S. (2003). Perceptual and acoustic cues of Polish coronal fricatives. In *Proceedings of the 15th ICPhS.* Barcelona: Causal Productions.