

PHONOLOGICAL CONTRAST AND ARTICULATORY EFFORT

ROBERT KIRCHNER
University of Alberta

Phonological theory has long been guided by the assumption that the representational elements consist of a limited inventory of distinctive features. Ostensibly, this assumption is motivated by the limited range of contrasts observed in sound systems. For example, segments are standardly categorised, for purposes of phonological analysis, as either [+nasal] or [-nasal]; further phonetic details (e.g. the precise area of the velopharyngeal port) are excluded from the representational inventory. The theory thus excludes unattested sound systems in which, for example, unnasalised, slightly nasalised, moderately nasalised, and heavily nasalised vowels all behave as separate phonemes.

In this article, I will argue against this representational assumption. In section 1, I demonstrate that the assumption is superfluous to an adequate treatment of phonological contrast. Within the framework of Optimality Theory, the contrastive status of a featural distinction [F] within a sound system falls out from the ranking of the corresponding input-output faithfulness constraint. Under the further assumption that, for some [F], the universal constraint set lacks a corresponding I-O faithfulness constraint, it follows that [F] cannot have contrastive status under any constraint ranking. It is thus possible to include the full range of phonetic detail in phonological representations, without thereby predicting spurious contrasts.

I then argue, in section 2, that phonetic enrichment of phonological representations is crucial to an adequate characterisation of many phonological phenomena. As a case in point, we focus on the class of lenition processes, which, I contend, motivate direct reference to a universally non-contrastive phonetic property, namely the articulatory effort (qua biomechanical energy) expended in realising particular segments. An effort-based Optimality Theoretic approach to lenition is proposed; and this approach is illustrated with an analysis of voicing, spirantisation, flapping, and elision in Tümpisa Shoshone in section 3.

1. CONTRASTIVENESS AND FAITHFULNESS

To begin, I will assume a set of featural correspondence constraints, of the form PRESERVE(F). PRES(F) is violated just in case F is inserted or deleted, or the value of F changes, in mapping from input to output.

1.1. Language-specific predictable status

To illustrate the connection between PRES(F) and the predictability or contrastiveness of F in a sound system, let us consider predictable aspiration of stops in English. For descriptive purposes, we can encapsulate the conditions governing the distribution of aspiration in terms of the following constraint:

- (1) ASPIRATE: A stop is [+aspirated] iff it is [-voiced], occurring in initial position in a stressed or word-initial syllable.

The English pattern is obtained under the ranking in the following tableaux:

(2)		ASPIRATE	PRES(asp)
a.	pØl - ÆpØl	*!	
☞	pØl - Æp Øl		*
b.	p Øl - ÆpØl	*!	*
☞	p Øl - Æp Øl		
c.	spØl - ÆspØl		
☞	spØl - Æsp Øl	*!	*
d.	sp Øl - ÆspØl		*
☞	sp Øl - Æsp Øl	*!	

(I assume that PRES(voi), the stress assignment constraints, etc. all dominate PRES(asp), therefore candidates [ÆbØl][pØl](unstressed), etc. are ruled out.) Tableaux (2a) and (b) show that, regardless of underlying specification for [asp], a voiceless stop in initial position within a stressed syllable is aspirated on the surface. Tableaux (2c) and (d) show that, regardless of underlying specification for [asp], a voiceless stop in any other environment is realised as unaspirated. Under this ranking then, for any underlying stops which differ with respect to [asp], the PR neutralises to a particular value of [asp]: [+asp] in the aspiration environment, and [-asp] elsewhere. Therefore, stop aspiration is not contrastive under this grammar.

If, however, ASPIRATE is ranked below PRES(asp) (and there is no other higher-ranking constraint on the distribution of [asp] in voiceless stops in this context), then [asp] is contrastive in stops, as in Hindi:

(3)		PRES(asp)	ASPIRATE
☞	pi - Æpi		*
	pi - Æp i	*!	
	p i - Æpi	*!	*
☞	p i - Æp i		

That is, an underlying distinction in [asp] is maintained on the surface (/pi/ - [pi] and /p i/ - [p i]) under this ranking.

To summarise, with the OT framework, the predictable vs. contrastive status of stop aspiration in English and Hindi respectively in no way depends upon the absence of [aspirated] specifications from any level of representation. Rather, it depends upon the ranking of *faithfulness* to this feature relative to constraints on the *surface distribution* of this feature.

1.2. Universal predictable status

Phonetically, of course, aspiration is not a zero-sum thing, but a continuous dimension. The degree of stop aspiration actually varies gradiently, in English and other languages, depending on the stress level and phrasal position of the relevant syllable (see e.g. Pierrehumbert and Talkin 1992). I will now show that the surface gradiency of the aspiration pattern can be handled in terms of the same sort of formalism — with constraints on surface distribution interacting with faithfulness constraints — without thereby predicting spurious contrasts involving intermediate degrees of aspiration. We simply recognise the continuous aspiration dimension in the representation, and adopt a gradient version of the ASPIRATE constraint:

- (4) ASPIRATE (gradient): A voiceless stop in initial position in a degree n stressed syllable, or in a degree n prosodic constituent, is realised with degree n aspiration.¹

The faithfulness constraint, PRES(asp), however, still imposes a binary distinction, no aspiration (= [-asp]) vs. some positive degree of aspiration (= [+asp]). Other distinctions in aspiration, such as [\pm degree 33 aspiration], are assumed to lack corresponding faithfulness constraints; this stipulation is simply the analogue, in this approach, of the standard stipulation that only a binary aspiration distinction is represented in the phonology.

Now, just as in the previous tableaux, if the ASPIRATE constraint dominates PRES(asp), aspiration is predictable (albeit now gradiently assigned, depending on the prosodic prominence of the context); PRES(asp) is inactive. On the other hand, if PRES(asp) dominates ASPIRATE, an underlyingly unaspirated stop maintains its lack of aspiration on the surface, regardless of context; while a stop which is underlyingly specified for some positive degree of aspiration is realised with degree n aspiration, in accordance with the gradient ASPIRATE constraint, so long as $n > 0$.

(5)

	PRES(asp)	ASPIRATE (gradient)
☞ stop with degree 0 aspiration — degree 0 aspiration (in context for degree n aspiration), $n > 0$		*
☞ stop with degree 0 aspiration — degree n aspiration (in context for degree n aspiration), $n > 0$	*!	
☞ stop with degree m aspiration — degree 0 aspiration (in context for degree n aspiration), $m, n > 0$	*!	*
☞ stop with degree m aspiration — degree n aspiration (in context for degree n aspiration), $m, n > 0$		
☞ stop with degree m aspiration — degree m aspiration (in context for degree n aspiration), $m, n > 0$		*

Crucially, however, since there are, by hypothesis, no faithfulness constraints on intermediate degrees of aspiration (PRES(degree m aspiration)), an underlying distinction between degree m aspirated stops and other aspirated stops is not maintained on the surface: a stop with some particular positive degree of aspiration can only map to the positive surface aspiration value which best satisfies the ASPIRATE constraint. Thus, just as in the previous section, regardless of the ranking of the relevant constraints, distinctions among intermediate degrees of aspiration have no contrastive status — notwithstanding the introduction of a continuous aspiration dimension in the representation.

1.3. Discussion

Note that the foregoing result, does not translate elegantly into a rule-based framework. We would have to stipulate that *every* language has a rule or set of rules that neutralise distinctions in degree of aspiration in all contexts; but the specific rules which achieve this outcome would differ from language to language. For example, the distinction

¹ This assumes some function, the details of which do not concern us here, for relating degree n in the stress dimension or prosodic hierarchy to degree n in the aspiration dimension.

between released and unreleased stops is universally non-contrastive, though stop release is phonologically relevant in licensing contour segments (partially nasalised or affricated) (Steriade 1993). Nevertheless, this feature neutralises (pre-pausally) to [+released] in French, [-released] in Korean, and is in free variation in English. In a rule-based framework, we would have to posit three distinct neutralising rules for the three languages; thus there is no unified formal expression of the feature's non-contrastiveness. In the OT formalism, however, what unifies the three cases is the universal absence of a PRES(released) constraint; and what distinguishes them is the ranking of conflicting constraints on the surface value of [released], such as "Stops must be released" vs. "Coda stops must be unreleased."

Further note that this OT approach permits us to appeal to a *potential contrastiveness hierarchy*. Intuitively, the position of PRES(F) in the constraint hierarchy of a grammar for a given speaker corresponds to the degree to which the speaker attends to feature F in the mapping between input and output. Thus, for example, speakers of Hindi attend to stop aspiration distinctions in a way that English speakers do not. Although constraint ranking is generally a language-specific matter, it must be recognised that certain featural distinctions are inherently more salient than others, e.g. [consonantal] (characterised by abrupt, large-scale changes in amplitude) vs. [longitudinal vocal fold tension] (principally cued by subtle F₀ perturbations in the beginning of a following vowel), which is contrastive only in Musey (Shryock 1995). The notion of potential contrastiveness, or inherent salience, can thus be formalised in terms of a set of universal ranking conditions: if feature F is inherently more salient than feature G, then PRES(F) universally outranks PRES(G) (cf. Jun 1995). In sum, features which are inherently highly salient have corresponding faithfulness constraints which are universally highly ranked; while inherently subtler features have lower-ranked faithfulness constraints. And, as discussed above, universally non-contrastive features lack faithfulness constraints altogether.

Moreover, this approach captures the connection between the frequent non-contrastiveness of some feature and its usual restriction to narrow environments in languages where it is contrastive. For example, Kaun (1994) observes that most languages do not permit a contrast in [round] independent of [back]; but in those languages that do (e.g. Turkish), contrastive rounding is typically subject to vowel harmony, i.e. surface restrictions on the vowels that can occur with it (within some domain). The lower the ranking of PRES(F), the more constraints on the distribution of F which may dominate PRES(F), hence the narrower the contexts in which F is contrastive, and the greater the likelihood that F will not be contrastive in any context at all.

The traditional representational treatment of contrastiveness, on the other hand, is all-or-nothing. If a phonetic property is admitted to the pantheon of phonological features, it is formally equal to all other features in its potential for signalling contrasts, and no distinction can then be drawn between features which are frequently contrastive and those that are rarely contrastive, cross-linguistically. Similarly, if a feature is contrastive in a given language, it must be present underlyingly, and no distinction can then be drawn between features which are contrastive in a broad array of contexts and contrasts which surface only in narrow contexts. The representational treatment thus fails to capture the notion of a potential contrastiveness hierarchy, as well as the connection between potential contrastiveness and contextual restrictions.

For further discussion of this approach, including a formal proof of the relation between contrastiveness and ranking of faithfulness constraints, see Kirchner 1997. A more richly articulated system of OT constraints for handling gradient variation (albeit following the same general approach to contrastiveness) is proposed in Boersma 1998.

Alternatively, Flemming 1995 rejects input-output faithfulness constraints in favour of constraints which refer directly to the maintenance of contrast over sets of possible forms; but his proposal is in accord with Kirchner and Boersma, to the extent that he handles phonological contrast in terms of the constraint system, rather than in terms of representational abstraction.

2. ARTICULATORY EFFORT IN PHONOLOGY

We have seen, in the previous section, that a restrictive feature inventory is superfluous to an adequate treatment of contrastiveness; and that by dispensing with this assumption — thereby allowing rich phonetic detail in phonological representation — phonological theory can take on gradient variation. I will now argue that this move is motivated as well by phenomena from the phonological “heartland,” focussing on the class of processes generally referred to as “lenition,” or “weakening.”² These traditionally include:

- *degemination*, i.e. reduction of a long (geminate) to a short (singleton) consonant (e.g. t: – t);
- *voicing* (e.g. t – d);
- *flapping*, i.e. reduction of a stop to a flap (e.g. t – r);
- *spirantisation*, i.e. reduction from a stop (or affricate) to a fricative or approximant continuant (e.g. t – {θ, ʈ});
- reduction of other consonants to approximants (e.g. r – ɹ, s – ʃ);
- *debuccalisation*, i.e. reduction to a laryngeal consonant (e.g. t – ʔ, s – h);
- and, at its most extreme, complete *elision* (e.g. t – Ø).

2.1. Previous approaches to lenition

As a threshold matter, we must we must ask why this set of processes should be regarded as a unified phenomenon. First, lenition processes have a unified phonetic characterisation: they all involve reduction of the magnitude or duration of articulatory gestures.³ Second, these processes occur in substantially the same set of contexts cross-linguistically (most typically intervocalic and coda positions). Indeed, sometimes we find lenitional chain shifts in a single context within a given language (e.g. Danish t – d – ð, Bauer et al. 1980), or lenition patterns whereby consonants display *more extreme lenition* in faster/more casual speech (e.g. Florentine t – {θ, ʈ, ð, or Ø} depending on rate/register of speech). Clearly, then, an approach which treated lenition as a collection of unrelated processes would be missing significant generalisations.

Nevertheless, previous treatments of lenition have failed to offer an (empirically adequate) unified formal characterisation of lenition, or to account for the contexts in which lenition typically occurs. Let us briefly consider the two most standard approaches. First,

² This is not to suggest, however, that lenition is the only empirical domain where this sort of approach appears to make headway. For example, Jun 1995 accounts for a number of generalisations concerning place assimilation in consonant clusters in terms of a very similar constraint system.

³ Voicing lenition might appear to involve adjustment of laryngeal specification rather than reduction. Nevertheless, voicing in fact conforms to the gestural reduction characterisation, upon a closer phonetic examination: first, voiced obstruents are typically realised with a more reduced oral constriction than their voiceless counterparts; secondly, voicing lenition typically occurs in medial position, where reduction of an active devoicing (glottal abduction) gesture results in passive voicing (Westbury & Keating 1986).

autosegmental feature-spreading treatments have been proposed (e.g. Harris 1984, handling Spanish spirantisation as [+continuant] spreading, cf. Mascaró 1983, Jacobs & Wetzels 1988; and see Selkirk 1980, Mascaró 1987, Cho 1990, Lombardi 1991 for similar treatments of voicing). But feature-spreading cannot be extended to lenition generally, for degemination, debuccalisation, and elision can only be expressed in autosegmental theory as deletion or delinking of phonological material. Moreover, this approach fails to give a natural account of the most typical lenition context, viz. intervocalic position: it suffices to spread the relevant feature from *either* adjacent vowel, and so the role of the other vowel in conditioning the lenition is unexplained.

An alternative approach, often tentatively suggested (e.g. Foley 1977, Churma 1988, Clements 1990, Hock 1991, Ní Chiosáin 1991, Elmedlaoui 1993, Lavoie 1996), but rarely fleshed out in explicit analyses, is the notion of lenition as *sonority promotion*. But if we take the sonority scale (e.g. stops > voiceless fricatives > voiced fricatives > nasals > liquids > high vowels/glides > low vowels (Dell & Elmedlaoui 1985)) seriously as a characterisation of lenition, we incorrectly predict that fricatives ought to be able to lenite to nasals. Moreover, vowel reduction, which would appear to be the vocalic counterpart of consonant lenition, typically involves *raising* (and centralisation), e.g. a – ə (see Crosswhite 1999); but the higher the vowel, the *less* sonorous it is. Finally, the sonority promotion proposal says nothing, per se, about the contexts and conditions under which lenition naturally occurs.

2.2. A unified, effort-based approach

Rather, I propose that lenition is driven by a phonetic imperative: minimisation of articulatory effort. This approach thus crucially presupposes that phonological representations can include universally predictable phonetic properties — in particular, the *effort cost* associated with a given set of articulatory gestures. Language-specific lenition patterns arise from the effort minimisation constraint (which I style LAZY), interacting with a class of lenition-blocking constraints, within an Optimality Theoretic grammar. Spirantisation, for example, is analysed in terms of rankings where LAZY dominates PRES(continuant) (6); under the opposite ranking (b), spirantisation is blocked:

(6) a.	<table border="1" style="border-collapse: collapse; text-align: center;"> <tr> <th style="padding: 2px 10px;">/d/</th> <th style="padding: 2px 10px;">LAZY</th> <th style="padding: 2px 10px;">PRES(cont)</th> </tr> <tr> <td style="padding: 2px 10px;">d</td> <td style="padding: 2px 10px;">**!</td> <td style="padding: 2px 10px;"></td> </tr> <tr> <td style="padding: 2px 10px;">ð</td> <td style="padding: 2px 10px;">*</td> <td style="padding: 2px 10px;">*</td> </tr> </table>	/d/	LAZY	PRES(cont)	d	**!		ð	*	*	b.	<table border="1" style="border-collapse: collapse; text-align: center;"> <tr> <th style="padding: 2px 10px;">/d/</th> <th style="padding: 2px 10px;">PRES(cont)</th> <th style="padding: 2px 10px;">LAZY</th> </tr> <tr> <td style="padding: 2px 10px;">d</td> <td style="padding: 2px 10px;"></td> <td style="padding: 2px 10px;">**</td> </tr> <tr> <td style="padding: 2px 10px;">ð</td> <td style="padding: 2px 10px;">*!</td> <td style="padding: 2px 10px;">*</td> </tr> </table>	/d/	PRES(cont)	LAZY	d		**	ð	*!	*
/d/	LAZY	PRES(cont)																			
d	**!																				
ð	*	*																			
/d/	PRES(cont)	LAZY																			
d		**																			
ð	*!	*																			

(This assumes that stops, *ceteris paribus*, have a higher effort cost than continuants, due to the greater distance which the articulator must travel in the former.)

Trivially, the treatment of spirantisation above, in terms of conflict between LAZY and faithfulness, can be extended to all manner of lenition phenomena. The type of structural change occurring in a given language depends upon which of the lenition-blocking constraints, if any, are ranked below LAZY: if PRES(length), then degemination; if PRES(voiced), then voicing; if PRES(sonorant), then reduction of an obstruent to an approximant; if PRES(place features), then debuccalisation; if PRES(segment), then elision; if no PRESERVE constraint, then no lenition at all. Lenition thus receives a unified characterisation, under this approach, in terms of the ranking schema: LAZY » lenition-blocking constraint.

2.3. Contexts

Under the simple case of LAZY outranking some faithfulness constraint, such as PRES(cont), the result is context-free lenition, e.g. Berber (Saib 1977), in which all singleton obstruents are realised as fricatives, in all contexts. However, with a few enrichments, the theory can capture context-sensitive lenition patterns as well, in terms of the same basic ranking schema. First, restriction of lenition to coda and word-final position can be understood in terms of the impoverished perceptual cues to a consonant's identity in phonotactic positions where it lacks an audible release, see Steriade 1993, 1995, 1996; Jun 1995. The greater perceptibility of consonants in positions where their release is audible can be formally expressed by breaking up faithfulness constraints according to context: the more salient position corresponds to a universally higher-ranked faithfulness constraint, thus PRES(F/released position) » PRES(F), as motivated by Jun 1995, cf. Beckman 1997. Coda and word-final lenition can then be obtained by ranking LAZY between these:

(7)

	PRES(cont/ released)	LAZY	PRES(cont)
ak [̣] - ak		**!	
☞ ak [̣] - aɣ		*	*
ak [̣] ta - ak [̣] ta		**!	
☞ ak [̣] ta - aɣta		*	*
☞ aka - aka		**	
aka - aɣa	*!	*	*

Intuitively speaking, this treatment captures the insight that there is greater impetus to lenite in contexts where there is relatively little perceptual "bang" for the articulatory "buck."

Restriction of lenition to particular *places* of articulation may similarly be obtained in terms of context-sensitive faithfulness constraints: specifically, the operative context refers to specific place features. For example, the following ranking results in spirantisation of coda dorsal consonants, but not coronals or labials, as in Quechua.⁴

⁴Alternatively, such patterns may be obtained by allowing place-specific LAZY constraints: the ranking {LAZY_{cor}, LAZY_{lab}} » IDENT(cont) » LAZY_{dors} yields the same pattern. It is not clear to me that any empirical difference follows from this place-specific LAZY proposal. I am inclined to favour the context-sensitive faithfulness proposal, however, based on the intuition that effort is an indivisible notion, and therefore it does not make sense to suppose that languages might arbitrarily disfavour effort involving some particular articulator; whereas it does make sense to suppose that speakers of languages differentially attend to particular auditory cues (see section 1.3).

(8)

	PRES(cont/ released)	PRES(cont/ lab)	PRES(cont/ cor)	LAZY	PRES(cont /dors)
Coda position: ak ^ɾ - ak				**!	
☞ ak ^ɾ - a				*	*
ap ^ɾ - ap				**	
☞ ap ^ɾ - a		*!		*	*
at ^ɾ - at				**	
☞ at ^ɾ - a			*!	*	*
Onset position: ☞ pa, ta, ka - pa, ta, ka				**	
pa, ta, ka - ϕa, θa, xa	*!	*	*	*	*

Furthermore, intervocalic lenition receives a straightforward effort-based treatment. *Ceteris paribus*, the more open the flanking segments, the greater the displacement (hence effort) required to achieve a given degree of consonantal constriction. The primacy of intervocalic position as a context for lenition thus falls out from the natural assumption that the impetus to lenite more effortful gestures is stronger than the impetus to lenite easier gestures. The correctness of this displacement-based understanding of intervocalic position is supported by the existence of the following related contexts:

CASE 1: triggers include open C's as well as V's:

(9) Shina (Rajapurohit 1983)

- a. *Voiced stops spirantise in /N__V position:*
baβo 'father' səði: 'monkey' muɣur 'bowl'
- b. *and when flanked by liquids or vowels, i.e. /r__V position:*
darβak 'race' paɾða: 'veil' gurɣur 'churning rod'
- c. *but not elsewhere (e.g. preceded by another stop):*
ekbo 'alone' səkdər 'file (tool)'

CASE 2: triggers restricted to low V's:

(10) Middle Italian (Central dialects) (Grammont 1939)

- a. *k > g /N__V when either flanking vowel was low:*
laku > lago ('lake') mika > miga ('crumb')
- b. *but not elsewhere:*
amiku > amiko ('friend') kaeku > tʃieko ('blind')

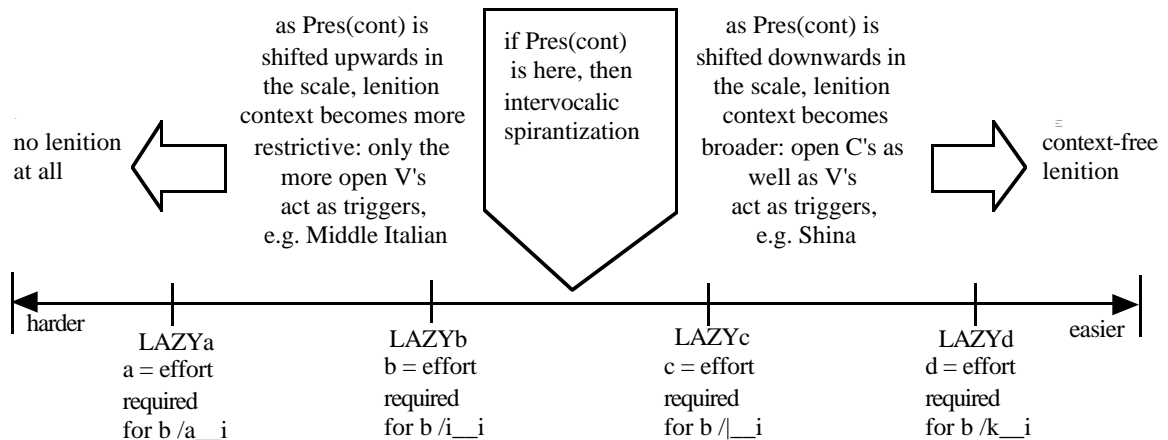
A formal treatment of intervocalic, and other displacement-based lenition contexts, simply requires decomposition of LAZY into a series of binary effort thresholds:

(11) LAZY "binarised":

... LAZY_{n+1} » LAZY_n » LAZY_{n-1} ...
where LAZY_n = "Do not expend effort *n*"

The intervocalic lenition context (and variations thereon) can now be obtained by interleaving PRES(cont) (or other lenition-blocking constraints) at particular points within the LAZY series:

(12)



2.4. Fortition constraints

Note, however, that for cases of *complementary distribution*, e.g. no word-initial fricatives, and no non-initial stops, the use of faithfulness constraints as lenition-blockers is insufficient.

(13)

	PRES(cont/#__)	LAZY	PRES(cont)
☞	a. #ka - #ka	**	
	#ka - #xa	*	*
☞	b. #aka - #aka	**!	
	#aka - #axa	*	*
☞	c. #xa - #ka	**	*
	#xa - #xa	*	
☞	d. #axa - #aka	**!	*
	#axa - #axa	*	

If, as in (13c), some word-initial obstruent is underlyingly [+cont] (and the OT tenet of Richness of the Base (Prince & Smolensky 1993, ch. 9) prevents us from excluding such an input), both faithfulness and LAZY favour the fricative candidate; thus it is impossible to rule out word-initial fricatives. An additional class of lenition-blocking constraints is therefore required: these must not only block lenition, but actively *induce fortition*, e.g. requiring word-initial obstruents to be realised as stops (*[+cont,-son]/#__).⁵ It seems plausible that these fortition constraints are, like the context-sensitive faithfulness constraints, grounded in perceptual considerations. For example, the release burst of a stop contains salient place of articulation cues (e.g. Wright 1996); thus, by militating in favour of consonants with a release burst, this constraint can be viewed as enhancing the perceptibility of the consonant; and the allocation of more robust cues to word-initial position may be viewed as reflecting the greater importance of word-initial consonants in lexical access. More generally, I will assume that the fortition constraints which we appeal

⁵ The ASPIRATE constraint introduced in section 1 falls into this fortition class.

to, for purposes of lenition typology, are of the form $*\alpha F/[D_]$, where D is some prosodic or morphological domain (including a stressed syllable), and αF refers to some feature specification which is less perceptually salient in the context $/[D_]$ than is $-\alpha F$.

2.5. Further generalisations

Furthermore, the effort-based approach to lenition straightforwardly accounts for a number of additional generalisations concerning lenition typology:

- Synchronic spirantisation processes never convert a stop to a sibilant or other strongly fricated continuant; rather, spirantisation typically results in weak fricatives or approximants, such as [ð] or [ɸ]. (This result is surprising, given the general observation that [s] and other strident fricatives are the unmarked fricatives.) Strident fricatives require a sustained interval of precise, close constriction: this is more effortful than a stop. Hence stop – strident fricative constitutes a net *increase* in effort; whereas stop – non-strident fricative involve a change to a less effortful, imprecise, acoustically weak partial constriction.
- Geminate stops never lenite, unless they concomitantly degeminate. (The phenomenon of geminate “inalterability” (cf. Guerssel 1977, Hayes 1986, Schein & Steriade 1986), reduces to this generalisation, to the extent that it does not reduce to language-specific blocking effects.) Geminate continuants require an even longer interval of precise, close constriction, a fortiori they are more effortful than a geminate stop; and voiced (obstruent) geminates are more effortful for aerodynamic reasons, cf. Ohala 1983). Consequently, geminate spirantisation or voicing would constitute an increase in effort; since lenition is, by hypothesis, driven by effort minimisation, such geminate lenition processes are ruled out universally.
- All else being equal, the faster or more casual the speech style, the more likely a given consonant is to undergo lenition. Greater effort (qua velocity) is required to achieve a given constriction in a shorter amount of time, hence there is greater impetus to lenite in fast speech. Register sensitivity can be captured in similar terms, using a register-adjusted effort cost. That is, a numerical index, inversely proportional to register, is added to the raw effort cost. Thus, a gesture which counts as sufficiently cost-effective in formal speech may be evaluated as too costly in casual speech, resulting in lenition.

For documentation and fuller discussion of these generalisations and my account thereof, and fuller discussion of the effort-based approach to lenition generally, see Kirchner 1998.

3. CASE STUDY: LENITION IN TÜMPISA SHOSHONE

To exemplify the effort-based approach to lenition in greater depth, we now examine the sound system of Tümpisa Shoshone, which includes spirantisation, voicing, nasal weakening, and elision processes.

3.1. Data

The data for this case study are drawn from Dayley's (1989) grammar of the Tümpisa⁶ (also known as Panamint) dialect of Shoshone, a Uto-Aztecan language spoken

⁶Note that Dayley's /ü/ represents a phonetically central unrounded vowel [i]; thus the word Tümpisa ('Death Valley') is pronounced [tĩmbiʃa].

in the region of the California-Nevada border.⁷ As a frame of reference, I present the following chart of consonant "phonemes" for Tümpisa Shoshone:

Table 1. Consonant "phonemes" of Tümpisa Shoshone.

labial	coronal	dorsal	labio-dorsal	laryngeal
p pp	t tt	k kk	k ^w kk ^w	ʔ
	ts tts			
	s			h
m mm	n nn	ŋ	ŋ ^w	

Note that here and throughout the description of consonant variation below, my use of phonemic terminology and notation (e.g. voiceless stop "phonemes" realised as voiced fricative "allophones") is purely descriptive; it does not imply the assumption that the surface fricatives are uniformly stops in underlying representation. Syllables are maximally CVVC, and the only permissible clusters are full geminates and homorganic nasal + stop clusters. All words end in vowels on the surface, though Dayley posits some word-final /h/'s and /n/'s in UR.

3.1.1. *Spirantisation.* In Tümpisa Shoshone (TS), stops occur in initial position (14a), as geminates (b), and following a homorganic nasal (c). Flaps (14b) and nonstrident fricatives (a) occur elsewhere.⁸

- (14) a. puhayānt̚i 'shaman'
 taβett̚j̚i 'sun'
 tuɣ^wānni 'night'
 tsiðooḥj̚i 'push'
 k̚immāy̚inn̚a 'to come here'
 k^wɪ̚j̚āã 'eagle'
- b. pariasipp̚i 'ice'
 uttūnn̚a 'to give'
 taβett̚j̚i 'sun'
 pūnikk̚a 'see, look at'
 ukk^w̚a 'when, if'
- c. taziūmbi 'star'
 indāw̚i̚j̚i 'your little brother'
 t̚ippij̚iφū̚ŋk̚i 'stinkbug'

⁷Other processes described by Dayley include optional devoicing of short vowels between voiceless consonants and in initial unstressed position; nasalisation of vowels adjacent to a nasal consonant; palatalisation of sibilants and nasals after front vowels; fronting of velars before front vowels; coalescence of /w + a/ to [o] or [u]; lowering of high vowels after [ʔ]; vowel rounding harmony; rounding of velars before round vowels; and a rather complex system of consonant gradation, involving morphemes which Dayley analyses as ending in abstract /n/ and /"/ (the latter indicating that the morpheme induces gemination of the initial consonant of the following morpheme).

⁸Dayley characterises the spirantisation/flapping environment as "intervocalic" (or more precisely, /V(h)_V). However, as the language's phonotactics permit no consonant clusters other than geminates and homorganic nasal + stop clusters, and (rarely) [hC] clusters, the spirantisation context reduces to the context-free characterisation above, subject to blocking in initial position and in full and partial geminates.

That is, [p,k,k^w] are in complementary distribution with [β,ɣ,ɣ^w], as [t] is with [ð] (after a front vowel) and with [r] (after a back vowel or /h/, the latter context illustrated by /tɪkkappih tukk^wan/ – [tɪkkappi_{ɔ̃} rukk^wan] 'under the food'). “Initial position” in Dayley’s description appears to mean utterance-initial, since spirantisation applies across word boundaries:

(15) pie ðuy^wānni ʒāãɣīnṅa 'it's already getting dark'

The distribution of sibilants is somewhat more complicated. On the one hand, the affricate /ts/ spirantises (to [z]), except in initial position and in full or partial geminates, just like the stops, e.g. /motso/ – [mōzō] ('whiskers'). But unlike the other fricatives [s] can occur in initial (16a) as well as medial (b) position.

(16) a. suřimṅi 'those'
b. pariasippṅi 'ice'

Nor is there a contrast between geminate and singleton [s], as there is in the stop and affricate series.

3.1.2. *Voicing and devoicing.* The distribution of voicing is also predictable. Stops are voiced following a nasal (14c); in initial position, and in geminates, however, they are voiceless (a,b). The fricatives resulting from spirantisation are voiced in most contexts. However, the underlying fricative /s/ (i.e. which does not derive by spirantisation from /ts/) is realised as voiceless in all contexts. Moreover, utterance-final vowels are optionally devoiced, in which case the preceding consonant (or typically the second half of a geminate nasal, e.g. suřimṅi) is devoiced as well (16).⁹ Furthermore, h + obstruent clusters coalesce to voiceless obstruents (17b). In these non-initial singleton obstruents which surface as voiceless, either due to final devoicing, or due to underlying /h/, spirantisation is optional; whereas flapping (17c) is obligatory, as is spirantisation in the non-devoiced case (14a).

(17) a. tahaβi ~ (φ/p)ṅi 'snow'
huβiarṅi ~ ... (x/k)ṅi 'sing'
peḏi ~ ... (θ/t)ṅi 'daughter'
mōzō ~ ... (z/ts)ṅi 'whiskers'
b. /ohpimpṅi/ – o(φ/p)ṅimbi 'mesquite tree'
/iattiah ka/ – iattia(x/k)ḡ
c. /tɪkkappih tukk^wan/ – [tɪkkappi_{ɔ̃}rukk^wan] 'under the food'

Note that the devoiced sibilant fricative [z̥], derived from /ts/, does not neutralise with /s/: Dayley describes [z̥] as more “lenis,” presumably meaning shorter, than [s].

3.1.3. *Nasal weakening.* The spirantisation pattern of obstruents is partially paralleled by nasals: a non-initial singleton labial nasal is realised as [w̃] (18a); and a non-initial singleton coronal nasal as [j] after a front vowel (b).

⁹This utterance-final devoicing optionally takes the form of glottalisation of the final vowel, described by Dayley as insertion of [ʔ] plus a voiceless echo vowel. Dayley notes that this devoicing-by-glottalisation is most common in uninflected nouns (in final position), and speculates that it may function as an allomorph of the absolutive suffix.

¹⁰In addition, /h + k^(w)/ sequences can debuccalise to [h^(w)], but this outcome is restricted to particular suffixes.

- (18) a. s̥h̥w̥ɔ̥ɔ̥ri 'ten'
 b. s̥ɛ̥j̥ū 'therefore'
 c. j̥i̥β̥ḁ̃ni 'autumn'

After a back vowel, Dayley transcribes the coronal nasal as [n], apparently without weakening (18c). The velar nasal [ŋ] likewise is not described as weakening (19a). Labiovelar [ŋ^w] does not weaken after a front vowel; however, after back vowels [ŋ^w] occurs in free variation with [w̃] (19b).

- (19) a. p̃ã̃ŋ̃e 'up'
 b. s̃ɔ̃(ŋ^w/w̃)ɔ̃ 'lungs'

Note that there are no geminate velar or labiovelar nasals, unlike the labial and coronal nasals. Also note that vowels are nasalised before and (to a lesser extent) after a nasal consonant.

3.1.4. *Elision of laryngeal consonants.* Dayley further describes an optional process of elision of intervocalic /h/ and /ʔ/, e.g. [po(ʔ)ittʃi] ('path'), [ta(h)aβi] ('snow'). However, as /ʔ/ *only* occurs in intervocalic word-medial position to begin with, the /ʔ/ elision can alternatively be viewed as context-free. The distribution of /h/ is somewhat broader. It can occur initially, as in [huβiariyi] ('sing'), and before a following glide, as in [tikkappih jãã] ('on the food'). As noted in section 3.1.2, h + obstruent clusters coalesce to a devoiced fricative or stop (or a devoiced flap, in the case of coronal stops). Thus, /h/ elision is restricted to intervocalic and pre-obstruent position.

3.2. *Analysis of Spirantisation and flapping*

3.2.1. *Basic spirantisation pattern.* In accordance with the effort-based approach outlined in section 2, the basic TS pattern of context-free spirantisation at all places of articulation, subject to blocking in utterance-initial position, follows from the following ranking:¹¹

¹¹Voicing / devoicing of the outputs is addressed in section 2.2 below.

(20)	non-initial:	*[+cont,-strid,+cons]/[___...]Utt	LAZY	PRES(cont)
	tapettʃi		**!	
☞	taβettʃi		*	*
	tsitooɰi		**!	
☞	tsiðooɰi		*	*
	puhakãntɨ		**!	
☞	puhayãntɨ		*	*
	tuk ^w ãnni		**!	
☞	tuy ^w ãnni		*	*
	initial:			
☞	puhayãntɨ		**	*
	βuhayãntɨ	*!	*	
☞	tuy ^w ãnni		**	*
	ðuy ^w ãnni	*!	*	
☞	kimmãɣĩnɲa		**	*
	ɣĩmmãɣĩnɲa	*!	*	
☞	k ^w ɨjãã		**	*
	ɣ ^w ɨjãã	*!	*	

(Here, and in tableaux below where the lenition-blocking constraints need not be interleaved within the LAZY series, I present LAZY as a single scalar constraint (with greater or lesser violations.) As discussed in section 1, since continuancy is allophonic in these obstruents, it is the constraint system which determines the surface value; even if we assume that the underlying specification is contrary to the surface value (as indicated in the above tableaux by assuming PRES(cont) violations even in the winning candidates), the non-initial singletons surface as continuants. The failure of these consonants to lenite further is captured by ranking faithfulness to other features above LAZY, e.g. PRES(cons).

(21)		PRES(cons)	LAZY
☞	p - β		*
	p - Ø	*!	

Moreover, the failure of /t/ to spirantise to [s], and the blocking of spirantisation in geminates and homorganic nasal + stop clusters (i.e. partial geminates) instantiate cross-linguistic generalisations, which follow from this effort-based approach, as mentioned in section 2.5.

3.2.2. *Sibilants.* The sibilant fricative /s/ occurs in initial position, hence the utterance-initial fortition constraint above does not prohibit strident continuants; but under the constraint hierarchy above, we incorrectly fail to block spirantisation of initial /ts/. I assume that the general property distinguishing /s/ from /ts/ and its allophones [ts,z,z.3,ʒ] is the shorter duration of strident energy in the latter. This seems plausible, in light of the general observation that voiced fricatives are typically shorter than voiceless (e.g. Narthey 1982), and Dayley's comment that [z,ʒ] are more "lenis" than [s,ʃ]; moreover, I observe in spectrograms of my own speech that the fricated portion of an sibilant affricate (or t + s cluster) is typically shorter than that of a fricative, presumably due to the more gradual onset of strident energy in the latter. Furthermore, a short strident fricative (with gradual onset of strident friction) is presumably perceptually weaker than an affricate (with abrupt onset of "full-strength" stridency, due to the sudden release of the preceding stop closure),

or than a longer strident fricative. I therefore posit a binary feature, [long stridency], which distinguishes the fortis strident fricative [s] ([+long strid]) from the lenis strident fricative [z] or [ʒ], and the strident affricate [ts] ([-long strid]), as well as the palatalised variants of all of the above (non-stridents are unspecified for this feature). The TS sibilant pattern now follows from undominated ranking of PRES(long strid), in combination with another utterance-initial fortition constraint, *[-long strid,+cont] / [_____]Utt, grounded in the relative perceptual weakness of the shorter non-affricate sibilants.

(22)

initial:	PRES(long strid) *[-long strid,+cont] / [_____]Utt	LAZY
☞ tsitooḥi – tsiḏooḥi		***
tsitooḥi – siḏooḥi	*!	**
tsitooḥi – ʒiḏooḥi	*!	*
senu – tsɛ̃jũ	*!	***
☞ senu – sɛ̃jũ		**
senu – ʒɛ̃jũ	*!	*
non-initial:		
motso – mōtsō		**!*
motso – mōsō	*!	**
☞ motso – mōzō		*
patiasippi – pariatsippi		***!
☞ patiasippi – pariasippi		**
patiasippi – pariazippi	*!	*

These constraints block initial /ts/ from spirantising, but do not block spirantisation of /ts/ medially, and permit /s/ to surface unchanged both initially and medially.

Finally, note that the absence of a geminate fricative [ss] is reflective of the higher effort cost, hence markedness, of geminate fricatives relative to stops (section 2.5), and follows from subordination of PRES(cont) to LAZY.

(23)

	LAZY	PRES(cont)
☞ ss – ss	**!	
ss – tts	*	*

Thus, even if an input contains a geminate sibilant fricative, it will neutralise to an affricate in all contexts (degeminated and deaffricated outputs are presumably ruled out by ranking of PRES(length) and PRES(strid) above LAZY).

3.2.3. *Variation with flapping.* A minor elaboration of this analysis further captures the variation between [ð] and [ʌ] as lenited allophones of /t/. Relative to a stop, a flap involves a reduction in magnitude, such that the active articulator makes the briefest of contacts with the passive articulator, while still maintaining non-continuity (see generally Inouye 1995). Presumably, coronal (specifically, apical) flaps are common, whereas non-coronal flaps are rare or unattested,¹² because of the greater stiffness of the coronal articulator, which allows it to reach its closure target and release the closure relatively quickly, without additional expenditure of energy.

¹²Margi presents the only known case of labial flaps (see Maddieson & Ladefoged 1996), and dorsal flaps are unattested.

(24)

	LAZY	PRES(son)
p, k	**	
b̥, ɡ̊ (extra-short)	***!	*
t	**!	
ɾ	*	*

I further hypothesise that in TS, the distribution of [ɾ] (after back vowels) vs. [ð] (after front vowels) is due to (LAZY-driven) coarticulation involving the tongue body. That is, in contexts where the non-continuant ([ɾ]) can be achieved without significant tongue body displacement, i.e. following a back vowel (Figure 1b), this is done. However, in a front vowel context, the tongue tip is closer to the dental region (Figure 1a); to achieve a flap, therefore, the tongue tip must either be dramatically retroflexed, or the tongue body must be retracted before the flap is made.

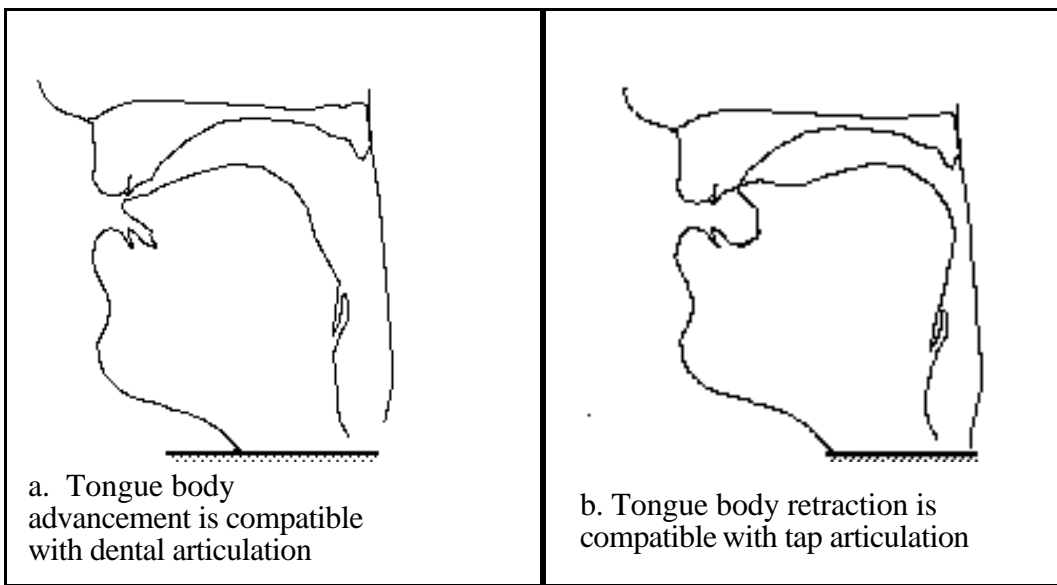


Figure 1. Flapping as coarticulatory retraction of the tongue tip.

(A dental flap is presumably not generally feasible, due to typical leakage of airflow through the teeth.) Hence, I assume that a flap is slightly more effortful following a front vowel than following a back vowel. Specifically, let x denote the minimum of effort required to achieve a flap following a front vowel, and y following a back vowel: for the foregoing phonetic reasons, $x > y$. The TS allophonic flapping pattern now follows from interleaving of a spirantisation-blocking fortition constraint, $*[+cont,-strid,+cons]$, between these effort thresholds:

(25)

	LAZY _x	*[+cont,-strid,+cons]	LAZY _y
at	*!		*
ar			*
að		*!	*
it	*!		*
ir	*!		*
ið		*	*

Analysis of (devoiced) flapping from coalescence of /h + t/ is deferred until after the general account of voicing and devoicing below.

3.3. Analysis of (De)voicing

3.3.1. *Basic pattern.* The TS context-free devoicing of (full) geminate obstruents is reflective of a cross-linguistic markedness generalisation, and follows from the general effort-based aerodynamic account of geminate devoicing/blocking of voicing alluded to in section 2.5. TS utterance-initial obstruent devoicing, and voicing in most other contexts, likewise follow from aerodynamic considerations. Obstruents passively devoice in utterance-initial position (Westbury & Keating 1986). Of course, as in the aerodynamic account of geminate devoicing, this aerodynamic state of affairs may be overcome by intercostal contraction (raising subglottal pressure) or various oral cavity expansion gestures, such as larynx lowering and pharynx expansion (lowering oral pressure); but these additional voicing-enabling gestures carry some additional effort cost. Hence,

(26) Utterance-initial voiced obstruent $>_{\text{effort}}$ Utterance-initial voiceless obstruent

Moreover, in utterance-medial position, (singleton) obstruents are passively voiced (Westbury & Keating 1986). Thus,

(27) Utterance-medial voiceless obstruent $>_{\text{effort}}$ Utterance-medial voiced obstruent

TS voicing allophony now follows from the following ranking:

(28)

	medial:	LAZY	PRES(voi)
	ta ϕ ettʃi	**!	
☞	taβettʃi	*	*
	tsi θ oohi	**!	
☞	tsi δ oohi	*	*
	puhaxãnti	**!	
☞	puhayãnti	*	*
	tux ^w ãnni	**!	
☞	tuy ^w ãnni	*	*
	initial:		
☞	puhayãnti	*	*
	buhayãnti	**!	
☞	tuy ^w ãnni	*	*
	duy ^w ãnni	**!	
☞	kimmã γ inna	*	*
	gimmã γ inna	**!	
☞	k ^w iãã	*	*
	g ^w iãã	**!	

3.3.2. *h + obstruent coalescence/devoicing.* This coalescence/devoicing resulting from /h + obstruent/ clusters can be obtained by disjunctively combining PRES(-voi) and PRES(aspiration) (cf. Kirchner 1996). Elision of the /h/ then follows from ranking plain PRES(asp) below LAZY:

(29)

	PRES(-voi) ∨ PRES(asp)	LAZY	PRES(asp)
→ ...hϕ...		***!	
→ ...ϕ...		**	*
→ ...β...	*!	*	*

That is, the [ϕ] candidate satisfies the disjunctive constraint, even though the aspiration noise is lost, because the voicelessness of the /h/ is preserved, shifted onto the following obstruent (satisfying PRES(-voi)); whereas the [β] candidate violates both.

3.3.3. *Final devoicing.* Finally, TS devoicing of utterance-final syllables may be attributed to abduction of the vocal folds, or increase in inspiratory force (causing subglottal pressure to drop off) (Westbury & Keating 1986), in anticipation of post-utterance breathing. Variable timing of these respiratory gestures relative to the end of the utterance is sufficient to account for the optionality of this process.

3.3.4. *Interaction with spirantisation and flapping.* Recall that spirantisation is optionally blocked in these devoiced obstruents. I attribute this blocking to a fortition constraint, *[+cont,-voi,{-strid∨-long strid}]. This constraint is presumably grounded in the observation that lack of modal voicing tends to obscure the formant transitions (cf. Silverman 1995) associated with these continuants, which are relatively acoustically weak, either because they lack strong friction, as in the nonstridents [ϕ,θ,x], or because the duration of this friction is brief, as in [z]. The TS optional blocking of spirantisation now follows from free ranking of this constraint with LAZY:

(30)

	*[+cont,-voi,{-strid∨-long strid}]	LAZY
→ tahap̥i		**
→ tahaϕ̥i	*	*
→ huβiariki		**
→ huβiarixi	*	*
→ mōts̥o		**
→ mōz̥o	*	*

An additional aspect of the post-/h/ context is that coronals lenite to a voiceless flap, rather than [ð] or [θ], even following a front vowel (cf. section 3.2.3). Presumably, the loss of the /h/ in this context results in some phonetic compensatory lengthening of the transition from the preceding vowel into the consonant, preserving something of the duration of the original /hC/ cluster. As a consequence, the tongue tip/tongue body ensemble have a longer time to achieve a non-continuant target.

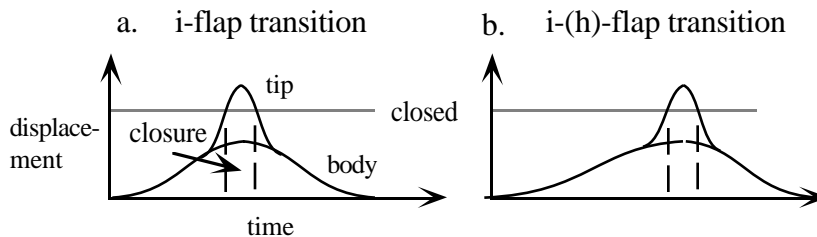


Figure 2. Schemata of tongue tip/tongue body ensemble displacement vs. time, without and with compensatorily attenuated transition.

Because the consonant gesture does not require as great a velocity to reach closure (Figure 2b) as it would in the non-attenuated case (a), less effort is required in (b) than in (a). Thus, if the effort required to achieve a flap in Figure 2a = x (as assumed in example (25), the effort required in Figure 2b (call it w) is somewhat less than x . Under the same ranking as posited in (25), with the additional assumption that $*[+cont,-strid,+cons] \gg LAZY_w$, we obtain the result that the coronal stop lenites to a flap, rather than a continuant, in the post-h context (31a), whereas it otherwise lenites to [ð] following a front vowel (b):

(31)

	PRES(cluster duration)	LAZY _x	*[+cont,-strid,+cons]	LAZY _w
a.	it – ir	*!		*
☞	it – ið		*	
b.	iht – i·ɾ			*
☞	iht – i·θ		*!	
	iht – iθ	*!		

3.4. Analysis of Nasal Weakening

3.4.1. *Basic Pattern.* The foregoing analysis of spirantisation can be extended to the nasal weakening facts, with minimal elaboration. Indeed, the general ranking $LAZY \gg PRES(cont)$, motivated above for TS, results in spirantisation of obstruents and nasals alike.¹³ TS differs from most other languages (e.g. Spanish, Harris 1969), in which spirantisation is restricted to oral noncontinuant, in that TS further subordinates the nasal fortition constraint, $*[+nas,+cont]$, to $LAZY$.

(32)

	LAZY	*[+nas,+cont]	PRES(cont)
☞	k ^w inãã	**!	
	k ^w ĩjãã	*	*
	s ^h mõõri	**!	
☞	s ^h wõõri	*	*

The blocking of weakening in full and partial geminate nasals follows from the same considerations as the blocking of spirantisation in geminate obstruents. Moreover, utterance-initial nasal weakening is blocked by the same fortition constraint which blocks obstruent spirantisation:

(33)

	*[+cont,-strid,+cons]/[_____]Utt	LAZY
☞	mõtsɔ	**
	wõtsɔ	*

3.4.2. *Voicing.* The failure of the nasals to devoice initially, as the obstruents do, is attributable to nasal venting of airflow, which prevents significant build-up of oral pressure; hence, initiation of voicing in nasals does not present the same aerodynamic problems as in oral stops. Nasals are therefore passively voiced in all contexts, modulo optional utterance-final devoicing, due to anticipatory glottal abduction, as discussed in section 3.3.3.

¹³Loss of closure in a nasal results in a nasalised approximant rather than a fricative (in the absence of dramatically increased subglottal pressure), due to the inhibiting effect of nasal venting on oral pressure, which is necessary to generate fricated airflow.

3.4.3. *Apparent place restrictions on nasal weakening.* Two facts, however, remain to be explained. First, according to Dayley's transcription, the coronal nasal surfaces as [n] after a back vowel, apparently failing to lenite. Given the reduction of the oral coronal stop to a flap in this context, and in light of the generally parallel behaviour of obstruent spirantisation and nasal weakening in TS, we would expect /n/ to reduce to a nasalised flap, [ɾ̃] in this context. However, without instrumental measurements of duration, it is difficult to distinguish [n] from [ɾ̃], since the other acoustic cues to the stop/flap distinction (e.g. presence of a burst) are absent in nasals.¹⁴ Therefore it seems plausible that these coronal nasals are actually flaps. Assuming this to be the case, the variation between [ɲ] and what Dayley transcribes as [n] follows from the same analysis as the variation between coronal fricatives and flaps, in section 3.2.3.

(34)

	LAZY _x	*[+cont,-strid,+cons]	LAZY _y
an	*!		*
aɾ̃			*
aɹ̃		*!	*
in	*!		*
iɾ̃	*!		*
iɹ̃		*	*

Second, according to Dayley, the velar nasal /ŋ/ never weakens, though velar stops spirantise. This fact might be attributed to the lowering of the velum during nasalisation, decreasing the distance which the tongue body must travel to achieve full closure. Blocking of /ŋ/-weakening would then follow from an interleaved ranking, where *u* denotes the effort required for [ŋ], and *v* denotes the (greater) effort required for a non-velar nasal.¹⁵

(35)

	LAZY _v	*[+cont,-strid,+cons]	LAZY _u
ŋ			*
ɳ̃		*!	
m	*!		*
ɱ̃		*	

However, the presence or absence of complete closure in a velar nasal is a subtle cue. Ohala (1975) (citing House 1957) observes that [ŋ] is acoustically quite close to nasalised vocoids:

[T]he velar nasal has primarily just a single resonating cavity with a small, perhaps negligible side-cavity, unlike other nasals, and thus negligible anti-resonances with large bandwidths and is more like that of a nasalised vowel than are those of any other nasal.

It is therefore plausible (again, notwithstanding Dayley's impressionistic transcription), that these velar nasals are, at least in some cases, a nasalised vocoid (presumably a nasalised dorsal glide, [ɳ̃]) rather than a non-continuant (cf. Trigo 1988 on the "placeless" behavior of many nasals which have been transcribed as [ŋ]).

¹⁴Particularly since the [n/ɾ̃] distinction is non-phonemic in English, Dayley's native language.

¹⁵This ranking is consistent with the previous tableaux, provided that *v* > *y* and *x* > *u*.

In fact, the variable weakening of the labiovelar nasal ($[N^w] \sim [w̃]$) suggests that both scenarios occur in TS. When complete velar closure is achieved, the nasal does not appear to weaken. Hence, $/ŋ/$ and $/ŋ^w/$ can surface unlenited. But when velar closure does not occur, due to contextual or pragmatic conditions which raise the effort cost of velar closure in a nasal (see the discussion of register-adjusted effort cost in section 2.5), the resulting continuants are heard (by Dayley) as $[ŋ]$ in the case of the plain velar (due to its confusibility with $[ũ]$), and as $[w̃]$ in the case of the labiovelar. Indeed the notion of contextual raising of the effort cost of velar closure allows us to understand why the variation in the realisation of $/N^w/$ appears to be limited to the context $/V_{+back}__$. Presumably, it is easier to achieve closure with the tongue body against the velum when the tongue body is already retracted due to the preceding vowel.

3.5. Analysis of Laryngeal elision

We have already accounted for obligatory elision of $/h/$ in pre-obstruent position (section 3.3.4). To account for its optional elision in intervocalic position, we simply need a context-sensitive version of the blocking constraint, PRES(aspiration): specifically, higher ranking for preservation of aspiration noise in contexts where it is followed by a more sonorous segment (see Bladon 1986, Silverman 1995 for the auditory basis for greater salience of quiet-loud vs. loud-quiet transition). Moreover, the variability of $/h/$ elision in intervocalic position, versus non-elision in pre-glide position, follows from interleaving of the context-sensitive faithfulness constraint within the LAZY series: specifically, between effort thresholds s (corresponding to $[h]$ in pre-glide position) and t (corresponding to $[h]$ in pre-vocalic position). (Because high tongue-body position tends, for aerodynamic reasons, to facilitate friction (see Ohala 1983), $s > t$.)

(36)

	LAZY _s	PRES(asp/___[-cons])	LAZY _t	PRES(asp)
...hφ...			*!	
☞ ...φ...				*
☞ ...VhV...	*		*	
☞ ...VV...		*		*
☞ ...Vhw...			*	
...Vw...		*!		*

Finally, context-free optional elision of $/ʔ/$ follows from free ranking of PRES(glottalisation) relative to LAZY.

(37)

	LAZY	PRES(glottalisation)
☞ ʔ - ʔ	*	
☞ ʔ - Ø		*

3.6. Summary

The TS lenition facts can thus be accounted for in terms of the following constraint hierarchy (in Hasse diagram form):

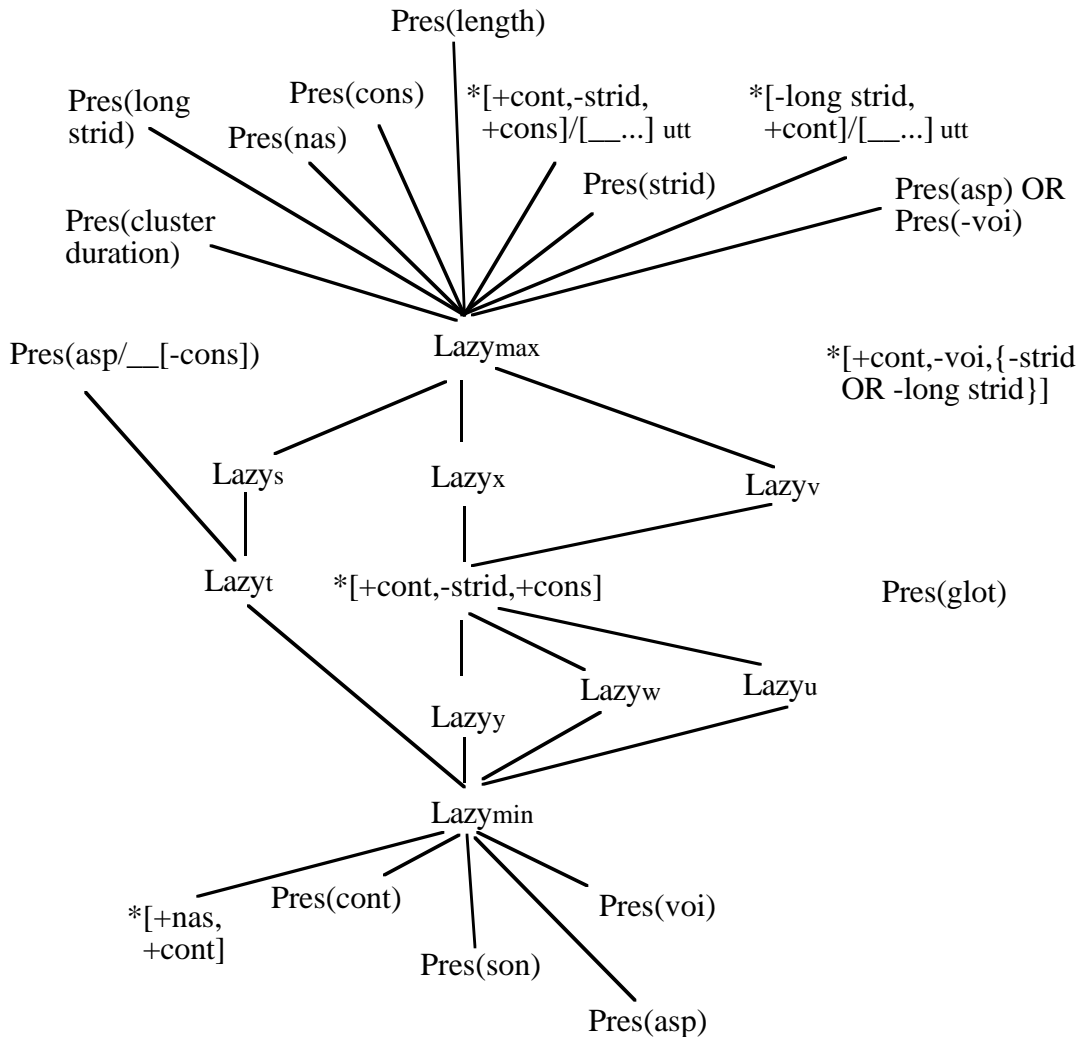


Figure 3. Hasse diagram of constraint hierarchy for Tümpisa Shoshone

(LAZY_{max} and LAZY_{min} in Figure 3 refer to the highest and lowest effort threshold, respectively, within the LAZY series: this corresponds to rankings in previous tableaux where LAZY (i.e. the whole series) either dominates or is dominated by some lenition-blocking constraint. Constraints which are not connected to the lattice in Figure 3 are freely ranked with respect to the other constraints.)

It is worthwhile to contrast the foregoing analysis of TS with conceivable rule-based alternatives, which permit no unified expression of the spirantisation operation (-cont - +cont, i.e. constriction reduction) and the flapping operation (-son - +son, i.e. temporal reduction).¹⁶ The effort-based analysis, on the other hand, is unified, in the sense that both the flapping process and the spirantisation process are driven by the same constraint, LAZY, and the choice between the spirantised and flapped outputs follows from a single, consistent constraint hierarchy.

¹⁶See generally Inouye 1995 for a review of the phonetic and phonological arguments against treatment of flaps as continuants (though Inouye does propose that flaps are [+cont] at their edges, i.e. tripartite contour segments, as noted in Chapter 1.

4. CONCLUSION

In the foregoing article, I have argued that, contrary to standard assumptions, an adequate formal treatment of phonological contrast within the OT framework does not require a restrictive inventory of distinctive features. Rather, phonological representation can include the entire sea of predictable or freely varying phonetic detail, including articulatory effort cost. Furthermore, by including effort cost in the representation, and allowing a set of phonological constraints which refer to effort cost, namely the LAZY series, we can devise a unified approach to lenition.

Moreover, the foregoing analysis of Tümpisa Shoshone demonstrates that this effort-based approach is not only capable of accounting for isolated typological generalizations, but can also offer a coherent and comprehensive analysis of the detailed lenition patterns of a particular language. It is worthwhile in particular to contrast this analysis with conceivable rule-based alternatives, which permit no unified expression of the spirantisation operation (-cont - +cont, i.e. constriction reduction) and the flapping operation (-son - +son, i.e. temporal reduction).⁷ The effort-based analysis, on the other hand, is unified, in the sense that both the flapping process and the spirantisation process (as well as all the other lenition processes in this language) are driven by the same constraint, LAZY (or particular thresholds thereof), and the choice between the spirantised and flapped outputs follows from a single, consistent constraint hierarchy.

REFERENCES

- Bauer, L., J. Dienhart, H. Hartvigson, & L. Jakobsen (1980) *American English Pronunciation: Supplement, Comparison with Danish*. Gyldendalsk Boghandel, Copenhagen.
- Bladon, A. (1986) *Phonetics for Hearers*, in G. McGregor (ed.) *Language for Hearers*, Oxford: Pergamon.
- Boersma, P. (1998) *Functional Phonology*, Doctoral Dissertation, Instituut voor Fonetische Wetenschappen.
- Cho, Y. (1990) *Parameters of Consonantal Assimilation*, Doctoral Dissertation, Stanford University.
- Churma, D. *On 'On Geminates,' ms.*, SUNY-Buffalo.
- Dayley, J. (1989) *Tümpisa (Panamint) Shoshone Grammar*, University of California Publications in Linguistics, vol. 115., UC Press.
- Dell, F. & M. Elmedlaoui (1985) *Syllabic Consonants and Syllabification in Imdlawn Tashlhiyt Berber*, *J. African Languages and Ling.* 7, 105-130.
- Flemming, E. (1995) *Auditory Features in Phonology*, Doctoral Dissertation, UCLA.
- Foley, J. (1977) *Foundations of Theoretical Phonology*, Cambridge University Press.

¹⁷See generally Inouye 1995 for a review of the phonetic and phonological arguments against treatment of flaps as continuants (though Inouye does propose that flaps are [+cont] at their edges, i.e. tripartite contour segments, as noted in Chapter 1.

- Giannelli, L. & L. Savoia (1979) *Indebolimento Consonantico in Toscana*, *Revista Italiana di Dialectologia* 2: 23-58.
- Guerssel, M. (1977) *Constraints on Phonological Rules*, *Ling. Analysis* 3:267-305.
- Harris, James (1969) *Spanish Phonology*, Cambridge Mass., MIT Press.
- Harris, James (1984) *La espirantización en castellano y su representación fonológica autosegmental*, in Bartra et al. (eds.) *Estudis Grammaticals*, Universitat Autònoma de Barcelona, Bellaterra.
- Hayes, B. (1986) "Inalterability in CV Phonology," *Language* 62:2, 321-352.
- Inouye, S. (1995) *Trills, Taps and Stops in Contrast and Variation*, Doctoral Dissertation, UCLA.
- Jun, J. (1995) *Perceptual and Articulatory Factors in Place Assimilation: an Optimality Theoretic Approach*, Doctoral Dissertation, UCLA.
- Kaun, A. (1994) *The Typology of Rounding Harmony*, Doctoral Dissertation, UCLA.
- Kirchner, R. (1996) *Synchronic Chain Shifts in Optimality Theory*, *Linguistic Inquiry* 27:2, 341-350 (1996).
- Kirchner, R. (1997) *Contrastiveness and Faithfulness*, *Phonology* 14:1, 83-111.
- Kirchner, R. (1998) *An Effort-Based Approach to Consonant Lenition*, Doctoral Dissertation, UCLA (available on Rutgers Optimality Archive, <http://www.ruccs.rutgers.edu/ROA>).
- Mascaró, J. (1983) *Continuant Spreading in Basque, Catalan, and Spanish*, in M. Aronoff and R. Oehrle (eds.), *Language Sound Structure*, Cambridge, MIT Press.
- Mascaró, J. (1987) *A Reduction and Spreading Theory of Voicing and Other Sound Effects*, ms. Universitat Autònoma de Barcelona.
- McCarthy, J. and A. Prince (1995) *Faithfulness and Reduplicative Identity*, Ms., University of Massachusetts, Amherst, and Rutgers University, New Brunswick, New Jersey.
- Nartey, J. (1982) *On Fricative Phones and Phonemes: Measuring the Phonetic Differences Within and Between Languages*, *UCLA Working Papers in Phonetics* 55.
- Ohala, J. (1975) *Phonetic Explanations for Nasal Sound Patterns*, in C. Ferguson, L. Hyman, J. Ohala (eds.) *Nasalfest: Papers from a symposium on nasals and nasalization*, Stanford: Language Universals Project, 289-316.
- Ohala, J. (1983) *The Origin of Sound Patterns in Vocal Tract Constraints*, in P. MacNeilage (ed.) *The Production of Speech*, 189-216, New York, Springer-Verlag.
- Pierrehumbert, J. and D. Talkin (1992) *Lenition of /h/ and Glottal Stop*, in Docherty and Ladd (1992), 90-117.

- Prince, A. & P. Smolensky (1993) *Optimality Theory: Constraint Interaction in Generative Grammar*, ms. Rutgers University, University of Colorado.
- Schein, B. and D. Steriade (1986) On Geminate, *Linguistic Inquiry* 17:4, 691-744.
- Shryock, A. (1995) *Investigating Laryngeal Contrasts: An Acoustic Study of the Consonants of Musey*, Doctoral Dissertation, UCLA.
- Silverman, D. (1995) *Phasing and Recoverability*, Doctoral Dissertation, UCLA.
- Westbury, J. & P. Keating (1986) On the Naturalness of Stop Consonant Voicing, *Journal of Linguistics*, 22, 145-166.