Title
Positional Neutralization: A Phonologization Approach to Typological Patterns

Permalink
https://escholarship.org/uc/item/82v7q3d2

Author
Barnes, Jonathan

Publication Date
2002
Positional Neutralization:
A Phonologization Approach to Typological Patterns

by

Jonathan Allen Barnes

B.A. (Columbia University) 1992
M.A. (University of California, Berkeley) 1995
M.A. (University of California, Berkeley) 1997

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy
in
Linguistics
in the
GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:
Professor Sharon Inkelas, Chair
Professor Andrew Garrett
Professor Larry M. Hyman
Professor Alan Timberlake

Fall 2002
Positional Neutralization:
A Phonologization Approach to Typological Patterns

© 2002

by

Jonathan Allen Barnes
Abstract

Positional Neutralization: A Phonologization Approach to Typological Patterns

by

Jonathan Allen Barnes

Doctor of Philosophy in Linguistics
University of California, Berkeley
Professor Sharon Inkelas, Chair

This study investigates the typology of patterns of positional neutralization, a term referring to systems in which, from a given set of oppositions, one structural position licenses a wider array of contrasts than another. Patterns of positional neutralization of vowel contrasts are surveyed in three pairs of strong and weak positions: stressed/unstressed, final/non-final, and initial/non-initial. For each pair, regularities involving the phonetic content of neutralization patterns are accounted for using a phonologization-based approach to typological patterns.

In the phonologization model, phonetics influences phonology only by providing the gradient inputs out of which categorical patterns are created by phonologization. Perceptually ambiguous language-specific phonetic patterns are reinterpreted by listeners, producing new phonological representations of the relevant strings. Once phonologized, these patterns cease to be dependent on phonetic factors for their existence. Phonological
patterns are natural because the majority of them descend directly from phonetic patterns. This same phonetic grounding would obtain with or without specific restrictions in the phonology mandating that it be so.

Chapter two examines unstressed vowel reduction, of which the vast majority of patterns are based on the neutralization of vowel height contrasts, the phonologized result of perceptual difficulties produced by duration-dependent undershoot. Chapter three looks at final syllables, in which both strong and weak licensing is attested. This ambiguity is attributed to the phonetic complexity of final position: lengthening and gestural enhancement increase the phonetic prominence of domain-final material, while drops in F0 and intensity, along with devoicing or non-modal phonation, do the opposite. Chapter four investigates initial syllables, showing that most reported examples of strong licensing for vowels here can be attributed to past or present fixed initial stress. Additionally, an experimental study involving Turkish isolates a process of small-scale phonetic lengthening of initial-syllable vowels. A diachronic scenario implicating this process in the development of vowel harmony patterns is presented. Chapter five concludes the study with a brief comparison of split and integrated models of the phonetics-phonology interface and with discussion of the roles of phonologization and Universal Grammar in accounting for patterns of phonological typology.
Table of Contents

Chapter 1. Introduction 1

1.1. Phonetics and phonology in recent approaches to positional neutralization 4

1.1.1. Pure Prominence 5

1.1.1.1. Phonetic arbitrariness in Pure Prominence models 9

1.1.2. Phonetically-driven Phonology 11

1.1.3. Integrated Phonetics and Phonology 12

1.1.4. Neo-Grounded Phonology 13

1.2. Goals 14

1.3. Phonetics and Phonology in the Phonologization Approach 16

1.4. Categories, neutralization and the phonologization process 20

1.5. Organization 29

Chapter 2. Stressed syllables and unstressed vowel reduction 34

2.1. Unstressed vowel reduction: patterns of neutralization 36

2.1.1. Vowel height contrasts 37

2.1.2. Nasalization 49

2.1.3. Quantity 49

2.2. The Phonetics of Vowel Reduction 52

2.3. Case Studies 57

2.3.1. The phonologization of phonetic vowel reduction: Bulgarian 57
2.3.2. Categorical vowel reduction: Belorussian 65

2.3.3. A mixed system: Brazilian Portuguese 67

2.4. A UG-based approach to UVR typology: Crosswhite (2001) 72

2.5. Vowel reduction in Russian 87

2.5.1. Facts 88

2.5.2. Experimental investigation of Russian UVR 95

2.6. Phonetic Determinism in UVR systems 106

2.6.1. Vowel Reduction in Seediq 108

2.7. Summary 113

Chapter 3. Final Syllables 117

3.1. Final syllable prominence 119

3.1.1 Phonetic prominence 119

3.1.2. Psycholinguistic prominence 124

3.2. Phonological strength effects in final position 126

3.2.1. Tone in final position 127

3.2.2. Vocalic strength effects in final position 128

3.2.2.1. Hausa 129

3.2.2.2. Pasiego Spanish 133

3.2.3. Final syllable resistance to assimilation and reduction 143

3.2.3.1. The phonetic roots of Final Resistance 153

3.2.3.2. Syllable structure effects on Final Resistance 161
3.2.3.3. Final Strength and psycholinguistic prominence 163

3.3. Why Final Strength is different from Stressed Syllable Strength 166

3.4. Categorical and gradient effects in Final Strength systems 172

3.5. On assumptions concerning licensing capacity and phonetic prominence 181

3.6. Final Weakening 190

3.6.1. Devoicing 194

3.6.2. Glottalization 213

3.6.3. Mixed Systems 226

3.6.4. Nasalization 231

3.6.5. Final vowel reduction and loss of contrasts 244

3.6.5.1. Lowering 244

3.6.5.2. Final reduction 247

3.6.5.3. The phonetics of final syllable reduction 253

3.7. Vowel Quantity in final syllables 262

3.7.1. Neutralization of quantity contrasts in final position 262

3.7.2. Final short vowel avoidance 274

3.8. Summary 276

Chapter 4. Initial Syllables 281

4.1. Recent work involving the phonology of the initial syllable 283

4.1.2. Previous proposals for the motivation of PN in stressed and initial syllables 284

4.1.3. Agenda for this chapter 288

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2. Licensing asymmetries in initial position</td>
<td>289</td>
</tr>
<tr>
<td>4.3. The Phonetics of Initial Position</td>
<td>293</td>
</tr>
<tr>
<td>4.4. Initial-syllable strength effects</td>
<td>303</td>
</tr>
<tr>
<td>4.4.1. Initial syllables in languages with fixed initial stress</td>
<td>304</td>
</tr>
<tr>
<td>4.4.2. Other sources of duration in initial syllables</td>
<td>310</td>
</tr>
<tr>
<td>4.5. Vowel harmony and initial syllables</td>
<td>314</td>
</tr>
<tr>
<td>4.5.1. Initial syllables in English</td>
<td>315</td>
</tr>
<tr>
<td>4.5.1.1. Methods</td>
<td>315</td>
</tr>
<tr>
<td>4.5.1.2. Results</td>
<td>317</td>
</tr>
<tr>
<td>4.5.2. Initial syllables in Turkish</td>
<td>318</td>
</tr>
<tr>
<td>4.5.2.1. Methods</td>
<td>320</td>
</tr>
<tr>
<td>4.5.2.2. Results</td>
<td>323</td>
</tr>
<tr>
<td>4.5.3. Consequences of domain-initial strengthening:</td>
<td>328</td>
</tr>
<tr>
<td>The Emergence of Turkic Palatal Harmony</td>
<td>328</td>
</tr>
<tr>
<td>4.5.3.1. Does vowel harmony come from vowel reduction?</td>
<td>329</td>
</tr>
<tr>
<td>4.5.4. Bantu</td>
<td>343</td>
</tr>
<tr>
<td>4.6. Summary</td>
<td>348</td>
</tr>
<tr>
<td>Chapter 5. Conclusions</td>
<td>353</td>
</tr>
<tr>
<td>5.1. Phonologization and Universal Grammar in phonological typology</td>
<td>353</td>
</tr>
<tr>
<td>5.2. On the integration of phonetics and phonology</td>
<td>358</td>
</tr>
<tr>
<td>5.3. Does &quot;anything go&quot; in phonology?</td>
<td>366</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Acknowledgements

The thanks I owe to my committee are on a scale so vast I’m not really sure where to begin. Over the last however-long-it-has-been, they have all shared with me enormous kindness, patience and wisdom. Alan Timberlake shared his knowledge of everything Slavic and his keen thinking at crucial moments throughout the development of this work. Larry Hyman was generous with his ideas, as well as with his experience and his endless enthusiasm. (How many discussions with him have begun, “Too bad you don’t have a chapter on...”!) Andrew Garrett has been helpful most recently in keeping a tight rein both on my prose style and on my driving. Over the long term what I have appreciated most has been his remarkable gift for seeing exactly what (if anything) is interesting and important about virtually any issue. Additionally, he has provided inspiration by example, both in his teaching and in his research endeavors. Sharon Inkelas has been the perfect advisor in so many ways – not only because of the lavish Turkish feasts at her house over the years, but also because of the tremendous clarity and insight of her thought. Without her constant help and guidance I could never have succeeded in completing this project.

Outside my committee, many people have been subjected to far more of this dissertation than I suspect they ever bargained for. Dasha Kavitskaya listened with patience and understanding to more than her share of half-baked theories and long-
winded lamentations. She's been a fine colleague and a true friend this whole time. Eddie
Cushman has been a source of real support for me all the while too, as has Suzanne
Wertheim, who has helped more than I could hope to thank her for. Reg, Rosie, and
Orangik (So orange!) were faithful companions as well.

I am grateful to John Ohala for teaching the course that got me interested in
phonetics and phonology to begin with, and for much advice over the years too. I profited
greatly from discussions and correspondence with Juliette Blevins, Robert Blust, Bruce
Hayes, Ian Maddieson and Donca Steriade at various points in the development of my
work. I'm grateful to Ronelle Alexander for starting me off on things Bulgarian and
South Slavic. I'd also like to thank Jim Matisoff in this year of his retirement for the word
"rhinoglottophilia", and also for the advice and encouragement he gave me at the
beginning of my career in linguistics.

Above all I thank my family for being there for me. I think you suffered through
the tougher parts of all this at least as much as I did, and it really meant a lot to me. This
dissertation is dedicated to the memory of my grandmother, Elizabeth M. Lee.
Chapter 1. Introduction

In this study I investigate the typology and implementation of positional neutralization of contrasts in vowel inventories, with an eye both to accounting in a meaningful way for regularities in that typology, and to elucidating, on the basis of the phenomena observed, the nature of the relationship between phonetics and phonology more generally. Most current approaches to positional neutralization (PN) attribute the emergence of crosslinguistic regularities in patterns of positional neutralization to specific restrictions imposed on possible phonological systems by Universal Grammar. I argue here that the role of the phonological grammar in generating the typology of PN systems, particularly as concerns matters of phonetic naturalness or grounding, is far less central than has previously been assumed. Rather, a diachronic phonologization approach to PN yields far more accurate typological predictions as to which systems should and should not be attested, while at the same time providing realistic phonetic explanations for the existence of those patterns. The phonologization approach thus liberates the categorical phonology from the need to reproduce in synchrony the phonetic naturalness sound patterns will acquire in diachrony one way or another.

meaning that the contrast in question is neutralized whenever one of the terms involved comes to occupy a particular structural position in the word or phrase. This Trubetzkoy opposed to “contextually conditioned” neutralization, in which the collapse of contrasts is triggered by, for example, segmental environment through processes such as assimilation, dissimilation and the like. I will use the term positional neutralization throughout this work in its maximally generic sense, referring to any instance of an asymmetrical capacity of two positions (or sets of positions) in the representation to license phonological contrasts, such that one set of structural positions licenses a larger array of contrasts than another. Specifically, one set of positions, termed “weak”, allows realization of only a subset of the range of contrasts available in another set of positions, termed “strong”. In this sense positional neutralization refers not only to instances in which contrasts are lost through mergers in weak positions, but also to cases in which licensing asymmetries arise through, for example, the emergence of new contrasts in strong positions.

In my treatment of PN systems, I defend an approach to the phonetics-phonology interface recognizing a complete separation between a gradient phonetics and an abstract categorical phonology in a manner similar to that advocated in Keating (1996, inter alia). Functioning as they do with different sets of primitives, the only link between these two systems comes in diachrony, as naturally arising language-specific phonetic patterns are
divorced from their phonetic origins and made phonological. Apparent restrictions on possible systems or principles of “phonetic grounding” operative in the phonology itself prove to be chimerical and unnecessary. Phonological patterns are grounded in phonetics by virtue of the fact that they originate in the phonetics in the vast majority of cases. Once phonologized, however, the relationship between two or more neutralizing entities is functionally arbitrary and no longer responsive to the phonetic forces which gave rise to the pattern of neutralization in the first place. Patterns appearing “unnatural” or phonetically unexpected in the phonology are the result of additional historical circumstance such as morphological changes motivated by factors such as paradigm uniformity, or subsequent sound changes which may obscure the phonetic origins of a process, or the appropriateness of a process to its environment. This view of the role of phonologization in determining the phonetic shape of phonological patterns has its roots in the listener-oriented approach to sound change developed by Ohala (1981, 1993a), and the work on phonologization of Hyman (e.g. 1977). More recent relatives of this approach are find Blevins and Garrett (1998), Garrett and Blevins (to appear), and the work of Blevins (in prep.) under the rubric of Evolutionary Phonology. Similar views are advanced in Kavitskaya (2001) as well.

The alienation described of a pattern of neutralization from its phonetic roots creates a problem for the phonologist seeking to understand the nature of the processes
involved, but *in no way* makes the pattern of neutralization any less viable from the point of view of synchronic implementation in the phonology. This is to be expected under the approach I advance in this study, according to which the phonetics-phonology interface literally *is* phonologization, such that the former influences the latter only through this medium.

1.1. Phonetics and phonology in recent approaches to positional neutralization

neutralization such that some combination of these features (i.e. voicing, aspiration, glottalization) can be freely contrasted in syllable onsets, but see their ability to contrast neutralized in syllable codas (see Steriade 1997 for details and analysis). Phonetic facts are often enough implicated more or less uncontroversially in the explanation of the diachronic development of such systems. In coda position, for example, the lack of a stop burst or CV transitions following the consonant may make the features in question more difficult to perceive, and hence more prone to effacement. C-place features are known to suffer similarly under the same conditions. Less obvious, however, is the extent to which this phonetic information is necessary or desirable in a synchronic model of positional neutralization.

1.1.1. Pure Prominence

Some approaches to positional neutralization are largely unconcerned with the phonetic motivations for the alternations they model (Beckman 1998, Zoll 1997, inter alia). While these Pure Prominence models differ from one another in substantive and principled ways, they share the basic assumption that positional licensing restrictions are best expressed in the grammar through constraints which have reference to a fixed set of phonological features and positions. Strong and weak positions are essentially listed as such, and are freely combinable with phonological features to produce constraints.
generating the necessary alternations or regularities. The Pure Prominence approach sees strong or weak licensing capacity as an inherent abstract property of a given position supplied by Universal Grammar, irrespective of language-specific phonetic details. In this sense it finds an ancestor in approaches to certain tonal and segmental phenomena in Bantu languages advocated by, e.g. Goldsmith (1982) or Hyman (1987), in which the designation of a particular syllable or mora with a grid mark or accent meant not the presence of metrical structure or anything interpretable as phonetic stress, but rather an abstract higher degree of relative phonological prominence, a prominence which might be expressed in increased licensing capacity for consonants and vowels or special behavior of some kind in the system of tone assignment, or some combination of all these. This approach, of course, need not necessarily bear the Universalist baggage encumbering later OT-approaches to positional prominence. The fact that the first mora of the stem-initial syllable in Kukuya bears an accent (reflecting the fact that the verb-stem initial syllable in Kukuya has, among other things, greater licensing capacity for both onset consonants and vowels than other syllables - Hyman 1987, Paulian 1974) may be a fact about Kukuya alone, leaving open the possibility that some other language might single out this very position as the locus of particular relative weakness of licensing capacity. Still though, the fact that the same set of positions (initial syllable, stressed syllable, root) appear again and again as strong licensers throughout the languages of the world, has led
Pure Prominence advocates to assume that the "strength" or "weakness" of particular structural positions are properties of those positions supplied in Universal Grammar.

A constraint system constructed on the principles of the Pure Prominence approach might look as follows in (1) below. This system has the effect of excluding mid vowels from unstressed syllables, a pattern typical of many languages with contrast-neutralizing vowel reduction.

(1) Neutralization of a Vowel Height Contrast in Unstressed Syllables
   a. Ident[hi]/ο  >>  *MidV  >>  Ident[hi]
   b. *MidV/unstressed ο  >>  Ident[hi]  >>  *Mid

(1a) accomplishes the vowel height neutralization using Positional Faithfulness constraints as proposed by Beckman (1998). Here, a general constraint banning mid vowels (*MidV) is ranked higher than a general faithfulness constraint mandating faithful output realization of the input feature [hi]. Were this the extent of the constraint system, the grammar would generate a language with no surface mid vowels at all. The reverse ranking, by contrast, would brook no alteration of underlying [hi] specifications despite a general dispreference for mid vowels, the latter being exposed only, for example, in situations of epenthesis, where the lack of underlying feature specifications makes the Ident constraint irrelevant, and allows markedness free rein in determining of the identity
of the new vowel. In the case of positional neutralization, however, the presence of the higher-ranked positional faithfulness constraint (Ident[hi]/⊙) mandates faithful realization of the feature [hi] specifically in stressed syllables, and has the effect of allowing input mid vowels located therein alone to surface, while all in all other cases they are barred from the output representation. This is illustrated below in the tableau in (2)\(^1\).

\[
\begin{array}{|c|c|c|c|}
\hline
/\text{C}e\text{C}/ & \text{Ident}[hi]/\text{C}/ & *\text{Mid} & \text{Ident}[hi] \\
\hline
Ce \text{C} \text{C} & * & * & \text{C} \text{C} \text{C} \\
\hline
\text{C} \text{C} \text{C} & * & * & \text{C} \text{C} \text{C} \\
\hline
\end{array}
\]

(1b) on the other hand is a system of positional markedness constraints of the type proposed in Zoll 1997 which accomplishes the same banishment of mid vowels from unstressed syllables. In this system, a general markedness constraint against mid vowels is outranked by a general faithfulness constraint for the feature [hi]. The higher-ranking positional markedness constraint *MidV/unstressed σ, however, has the effect of leaving intact only input mid vowels located in stressed syllables.

\(^1\)Additional constraints necessary to ensure raising rather than lowering of unstressed mid vowels are not shown. Specifically, Ident[lo] would have to outrank Ident[hi] to make a change of [−hi] to [+hi] more acceptable than the analogous alteration for [lo]
1.1.1.1. Phonetic arbitrariness in Pure Prominence models

Otherwise significant differences between these two approaches will not be of concern here\(^2\). What is noteworthy in this context rather is the arbitrary relationship between the positions “stressed syllable” or “unstressed syllable” and the features [hi] and [lo] defining mid vowels. As it happens, precisely this combination of positions and features is necessary with great frequency crosslinguistically, and thus raises few eyebrows in its formalization as above. But as far as the phonology is concerned, there is no reason why these statements should be preferred over the combination of the same positions with any other sets of features, e.g. *[anterior]/unstressed \(\sigma\). Models not assuming universality of positional strength or weakness, furthermore, such as the accent approach to stem-initial prominence in Kukuya, are in principle free to designate any position strong or weak in this manner, regardless of crosslinguistic concerns of phonetic plausibility. Any one combination of feature with position is considered just as well-formed from the point of view of the phonology as any other combination of feature with

\(^2\) Zoll’s reason for advocating the Positional Markedness approach for certain cases was due to the existence of patterns of positional neutralization in which material banned from weak positions migrates to the strong position and is realized there. The core concept of Positional Faithfulness, that it is more important not to alter input representations in strong positions than elsewhere cannot accommodate this pattern, since in moving the marked features from weak position to strong it has precisely the effect of altering the input specifications of the strong position. This is precisely the kind of issue which this dissertation argues should concern phonologists modeling positional neutralization in the grammar.
position, whether or not there is any reason to suppose that that feature is in any way related to that position phonologically in any language. For the case of laryngeal neutralization cited above, the relevant constraints would simply combine the relevant feature (e.g. [+constricted glottis]) with the relevant position (here, coda) in either a faithfulness or a markedness constraint to generate the desired neutralization pattern. It is not clear in these approaches what we should make of the fact that the grammar would be just as happy formally with the same constraint with “unstressed syllable” substituted for “coda”, both being weak positions.

For better or worse, then, there is a fairly obvious way in which this approach is missing certain generalizations. Specifically, it is manifestly not the case that all features are equally relevant or active phonologically in all positions. For example, languages do not typically ban all laryngeal contrasts are typically not banned from all and only positions within the stressed syllable, and a theory which allows for the simple expression of any number of such unattested and phonetically-unexpected systems is simply deficient in its ability to account for the crosslinguistic regularities attested in the typology of positional neutralization patterns.
1.1.2. Phonetically-driven Phonology

Observations like these form the basis of the influential Licensing-by-Cue theory advanced by Steriade (1997) and implemented in numerous works of other authors adopting versions of this approach, such as Bradley 2001, Côté 2000, Crosswhite 2001, and Zhang 2001. Steriade observes that in positional neutralization patterns, the same features appear correlated again and again crosslinguistically with the same positions, and argues furthermore, that this correlation follows from the specific phonetic characteristics of each position. More precisely, Steriade’s claim is that features are licensed preferentially in positions in which phonetic conditions make them maximally perceptually robust, and are likewise eschewed in positions where they would be less robust perceptually, and hence easily overlooked. It is not then the position itself which licenses or bans features, but rather the concrete phonetic cues which are important for those features’ perception. For example, potentially hard-to-perceive laryngeal contrasts, such as those involving voice, aspiration or glottality could be licensed on stops only in the presence of release bursts and following CV transitions, wherever those may happen to occur in a given language. Likewise certain vowel height contrasts, such as mid vs. high or low, could be permitted exclusively on vowels with sufficient phonetic duration for their accurate perception. That voiced, aspirated, or glottalized consonants happen to contrast in one or another structural position within the syllable, or that the vowels with
inadequate durations are often phonologically unstressed is on Steriade’s view irrelevant.

This is the point of greatest contrast with the Pure Prominence approach to PN. If it turns out that not only stressed vowels but, for example, phrase-final vowels as well have enough duration to license mid vowels in some languages, then Steriade’s approach is doubly vindicated, in that it avoids the disjunctive specification of environment which would otherwise be necessary in the phonetics-free models described above.

1.1.3. Integrated Phonetics and Phonology

The integrated models of phonetics and phonology of Flemming (passim) and Kirchner (e.g. 1998) share this goal of deriving phonological patterns of neutralization directly from phonetic factors, though are not committed specifically to perceptual cues as the basis of all licensing statements. In these models the boundary between phonetics and phonology is erased entirely, allowing any phonological process to refer directly to an unlimited array of phonetic information, whether acoustic or articulatory in nature. Flemming (2001), for example, derives patterns of vowel reduction from the decreased duration of unstressed syllables resulting in an unacceptable amount of effort necessary to reach articulatory targets for the production of non-high vowels. This class of approaches shares with the Licensing-by-Cue model the ability to achieve improved accuracy of typological prediction in comparison with the pure prominence models detailed above.
since it too brings the phonetic forces which shape phonological patterns diachronically
directly into their implementations in synchrony. By linking the licensing patterns
directly to the phonetic effects which cause them, these approaches avoid the vast
overgeneration characteristic of the phonetics-free, abstract phonological model.

1.1.4. Neo-Grounded Phonology

Attempts to chart a third path in accounting for positional neutralization patterns
include the approaches of Hayes (1997) and Smith (2002), both of which owe a great deal
to the Grounded Phonology approach of Archangeli and Pulleyblank (1994). According
to Smith, for example, phonological positions and features are combined as before to
form PN-inducing constraints, only now, before incorporation into the grammar, they are
subject to screening by a set of phonetically-sensitive substantive filters which endorse
constraints reflecting articulatory or perceptual reality in some way and reject
combinations of features and positions not grounded in this manner. These filters are said
to constitute a sort of “meta-grammar” of constraint construction. This approach holds
the specter of phonetic detail in phonology at bay while incorporating some of the
restrictiveness of the Licensing-by-Cue model into the theory.
1.2. Goals

This study aims to shed light on the controversies concerning both the implementation of positional neutralization and the role of phonetics in phonology more generally. It is devoted to specifically to positional neutralization patterns affecting contrasts in vowel inventories, though in some areas where relevant consonants will be treated as well. I have selected for investigation three commonly cited pairs of strong and weak positions: stressed vs. unstressed syllables, initial vs. non-initial syllables, and final vs. non-final syllables. For each pair of positions, I conducted an typological survey of the licensing asymmetries found in several hundred languages, the results of which I report in each chapter, using some of the clearer cases to illustrate general patterns.

The question to be answered again is why the neutralization of certain contrasts should be common in a given pair of positions while other patterns are rare or unattested, or more generally, how can we best account for the existence and nature of typological regularities in the patterning of sounds. In each case I show how phonetic characteristics commonly associated with the position in question can explain why one pattern of neutralizations should be attested regularly while others seem not to occur. In the case of

---

1 The first two of these are uncontroversially included in most lists of strong/weak pairs, while the last is a matter of some disagreement. Steriade (1994) singles final position out as strong, while Zhang (2001) and Smith (2002) note that the status of final syllables is somewhat unclear. Final position is also traditionally cited as a monolithically weak position (Hock 1999).
stressed syllables, where the phonetic factors involved in the development of licensing asymmetries are relatively unambiguous (dramatic durational asymmetries between stressed and unstressed syllables in the languages in question), the phonological neutralization patterns found there are likewise fairly unequivocal. In the case of final syllables, however, where the phonetic situation is more complicated and variable, the phonological status of the final syllable with respect to the licensing of contrast is just as complex, making final syllables strong for some contrasts in some languages, and weak for the same or other contrasts in other languages. The ambiguous status of the final syllable presents a strong argument against the notion of strength or weakness of licensing potential as inherently determined primitives in Universal Grammar. Whether phonological strong or weak licensing effects are ultimately phonologized in a given position for a given feature in a given language will depend on the language-specific phonetic characteristics expressed in the relevant position.

The recurrence of certain patterns in the phonology is entirely a function of the recurrence in the phonetics of the input patterns to the phonologization thereof. It is at this phonetic level then where the task of explaining the origins of these potentially universal patterns becomes meaningful\(^4\). The recurrence of the phonetic patterns together

\(^4\) I refer here only to aspects of sound patterns in which issues of naturalness or phonetic content are concerned of course. I do not mean to suggest that the phonological grammar does not or cannot make
with the reinterpretative mechanism of phonologization, on the other hand, insures that
certain patterns will be attested in certain positions to the exclusion of others in the
phonology whether these patterns are specifically derived by the phonological grammar
or not.

The role of the phonetics in determining the typology of positional neutralization
patterns now becomes clear. The link between phonetic reality and the phonological
process, however, I argue is severed at the moment of phonologization in a manner which
is not compatible with the direct connection between these two assumed in the Licensing-
by-Cue and other "direct phonetics" approaches described above. The sporadic surfacing
of phonetically unmotivated neutralization patterns through the agency of layered sound
change or the operation of analogy underscores the indifference of the phonological
grammar to the phonetic identity of entities it manipulates in patterns of positional
neutralization.

1.3. Phonetics and Phonology in the Phonologization Approach

The phonologization approach to phonological typology I advance in this study
rests on the following assumptions concerning the relationship between phonetics and
demands of its own on the processes it implements. Only that the existence of such restrictions should be
demonstrated to be necessary, rather than assumed a priori.
phonology: phonetics and phonology are separated by a formal split of essentially the
type proposed by Keating (1988b, 1996, inter alia), and assumed with some variations by
Cohn (1990), Zsiga (1993) and many since. While these models differ in some respects,
crucial and common to all and adopted here is the view the phonetic component operates
on gradient, physical representations containing quantitative specifications for each
dimension, while the phonology is categorical, operating on abstract symbolic
representations, interpreted by the phonetics, but not necessarily having phonetic content
themselves. This interpretation process provides contextually-appropriate phonetic targets
for each feature specification in the phonology, between which a one to one relationship
is of course not assumed (a given phonological feature may receive target specifications
along a series of phonetic dimensions). In many ways this view of the phonetics-
phonology interface resembles more traditional approaches to phonetics and phonology,
which distinguish phonological processes from processes which are “just phonetic
implementation”. It differs from traditional approaches, however, in removing phonetic
content from phonology altogether, freeing the phonology from the responsibility of
reflecting naturalness in any fashion, allowing it to be truly abstract and symbolic. It also
removes from the purview of phonology the implementation of a great number of
gradient, phonetic processes previously thought to apply within that component. The
phonetics thus has increased responsibilities, generating as it must not just universal
aspects of implementation, but a broad array of language-specific patterns which cannot be considered categorical (phonological).

I adopt, for purposes of modeling the effect of the phonetic component in creating the vowel realization patterns which precede the phonologization of merger, Keating's (1996) view of phonetic target windows. Here Keating uses target windows of varying widths to model phonetic underspecification as a matter of degree, a refinement of her earlier all-or-nothing approach to the subject (Keating 1988a). Rather than assuming that there either is a target specified for a given phonetic feature or there is not (in which case values must be interpolated as a path between specifications for surrounding segments), specifications of varying degrees of strictness can be modeled by assuming narrower or broader target windows as the case warrants. Keating (1996: 274) describes these target windows as "ranges of permitted values", where a narrow window represents a precise target with little room for contextual variation, while a wide window indicates a more loosely specified target, such that more contextual variation in its realization is tolerated. The widest possible window corresponds to phonetic underspecification, where targets must simply be interpolated on the basis of the specifications of neighboring target values along the relevant dimensions. Thus, in cases where the F1 value for the low vowel /a/ is seen not to vary significantly with the duration of the vowel in question (despite the increased effort required to reach the more open target articulations of non-high vowels
within periods of shorter duration), we can assume a narrow target window for F1 values. Where F1 lowers significantly as /a/ raises toward [ə] in contexts of reduced vowel duration such as unstressed syllables in UVR languages, a wide target window for the vowel height dimension models the increased variability of the specification, such that contextual factors like duration can have a stronger influence over the outcome the articulation.

I will assume the existence of both acoustic and articulatory target specifications, though in some approaches, such Articulatory Phonology (e.g. Browman and Goldstein *passim*, Zsiga 1993, Byrd 1996), targets are exclusively gestural in nature. I will also assume that target windows may be resized according to structural position or prosodic environment. The width of the target window itself is meant to encode variability by context, such that this may seem unnecessary. On the other hand, it may be necessary to assume in certain cases of positional allophony target windows of differing widths in different circumstances, such that while a vowel under stress, for example, shows little variation regardless of contextual changes in speech rate or segmental environment, an unstressed vowel may be allowed to vary more under identical circumstances. Such a possibility is mentioned by Keating (1996: 275-276) and assumed by Guenther (1995). The case of vowel reduction in Russian described in chapter two may ultimately warrant
such a treatment, though further detailed phonetic investigation of contextual variation will be necessary before this is entirely clear.

1.4. Categories, neutralization and the phonologization process

As outlined above, while the influence of phonology on phonetics takes place in the synchronic interpretation of categorical phonological representations as implementable sets of phonetic targets, in the approach assumed here the influence of phonetics on phonology is diachronic in nature, the result of the process of phonologization changing phonological representations according to reinterpretations of phonetic patterns. Before outlining the phonologization process, however, it is necessary first to provide further details concerning the relationship between phonological categories and phonetic targets in the phonologization model assumed here.

The first possibility for the relationship between the realizations of a single phonological element in weak and strong positions involves a single wide target specification for the phonetic feature in question. Thus, in a system where a single phonological entity /e/ is realized as [e] under stress, but in unstressed syllables raises toward [i] as a function of the duration accorded it, we can imagine a single wide target window for all realizations of /e/. In stressed syllables, assuming durational targets to be consistently high, the pressure from shortened durations necessary to condition raising
would never arise, allowing values for the specification of the height of /e/ to be realized toward the center of target window for that dimension (whether conceived as a set of articulatory parameters corresponding to openness of the upper vocal tract or as acoustic targets for F1). In environments such as unstressed syllables, however, where duration is characteristically low, pressure to undershoot the target would increase, resulting in values closer to lower end of the F1 target window. This situation would imply, however, that any instance of reduced duration would result in raising, while any instance of longer duration would see none. We would predict then that in fast speech stressed /e/ too would raise toward [i], while in slower speech unstressed /e/ should fail to raise. This gradience of application of reduction is certainly characteristic of the vowels of unstressed syllables in what I will argue are unphonologized, gradient systems of UVR such as the reduction of [a] to schwa in Russian non-first-pretonic unstressed syllables. It is less clear, however, that the same is true of the realization of stressed vowels.

If it is the case that in the UVR language in question stressed /e/ never raises significantly toward /i/, even in cases where increased speech rate lowers its duration substantially, the above picture must be revised somewhat. In this case it would be necessary to assume a narrow target window for parameters connected with the realization of vowel height in stressed syllables, thereby allowing less contextual variation. In unstressed syllables, however, the target window for the same phonological
/e/ would be expanded, such that durational changes (for example) would cause significant variation in the realization of the vowel along the height dimension. Shorter durations would mandate strong raising toward [i], while longer durations would necessitate less, possibly no reduction.

At no point in this discussion so far has the concept of contrast neutralization played any role. It is nonetheless commonplace in the literature to see discussion of vowels nearly neutralizing, or sometimes neutralizing, or gradiently neutralizing. For the purposes of this study such characterizations are held to be meaningless. Neutralization here refers solely to a phonological process whereby two abstract categories are merged into one under certain conditions, resulting in a loss of contrast between the two. In a situation such as that described above, in which we find duration-dependent reduction of /e/ toward [i] in unstressed syllables, it is quite possible, and indeed often the case in systems of positional neutralization (such as the Bulgarian UVR of /o/ toward [u] described in chapter 2) that due to frequent contextual pressures the statistical distributions of the realizations of two independent phonological entities such as /i/ and /e/ will overlap significantly along some, perhaps all dimensions. This would mean in effect that some percentage of the realizations of both phonological entities would be phonetically ambiguous, the relevant tokens having a certain probability of belonging to the distribution of either the one or the other phonological category. Two categories may
overlap in this way in a given set of positions (as Bulgarian /o/ and /u/ in unstressed syllables), however, without implying that the contrast between the two has been neutralized. Indeed, we would speak of neutralization of the contrast only in such case as the distributions of the two entities could be shown to be the same, such that no significant differences emerge in statistical comparisons of the two. Assuming that identical distributions imply identity of the relevant phonetic target windows, and assuming that a single array of phonetic targets for all dimensions in a given context must correspond to a single array of feature specifications for the segment in question, identity of realizational distributions for two underlying segments implies the phonological neutralization of the contrast between those two segments in the position in question\( ^5 \). If it can be determined, thus, that /e/ and /i/ in the unstressed syllables of a given language have identical phonetic distributions, the contrast between the two must be assumed to be neutralized, implying an operation within the categorical phonology of some process changing the underlying feature specifications of /i/, /e/, or both, as the pattern would have it. The realization of /e/ and /i/ identically within a distribution characteristic of [i] implies change of /e/ to /i/ in the phonological grammar\( ^6 \).

---

\(^5\) This means that two phonological entities with non-identical feature specifications in a given context must not have identical distributions of their phonetic realizations.

\(^6\) See Myers (2000) for illuminating discussion of similar issues.
We could equally imagine, however, a situation in which a single phonological entity, /a/ for example, had a window of realizations in stressed syllables varying narrowly around [a], and a window of realizations in unstressed syllables varying equally narrowly around [ə], such that in neither structural position would we find significant variation in the realization of the vowel /a/, but in the two positions the narrow target windows for the relevant phonetic dimensions would be completely distinct. In such a system stressed /a/ would be [a], while unstressed /a/ would be [ə] regardless of perturbations in the durations of the vowel in question. It is common in the literature to find allophonic processes such as this one, in which the two allophones vary little and have little or no overlap, referred to as “categorical” processes, as opposed to gradient processes in which a continuum of realizations is attested between the two extremes, most likely as function of contextual factors such as duration. This study will avoid this usage completely. Allophonic processes in which allophones have distinct distributions with little overlap will be assumed to have one narrowly specified phonetic target in one structural position and another narrowly specified phonetic target in the other. Gradient processes describing continua of realizations between two points will be assumed to be the result of loose specification of a single target for both positions, or potentially even partially overlapping but separate target windows in the positions in question. The term
"categorical" will be reserved solely for processes affecting the phonological category membership of the entities in question.

At this point, however, the distinction between the phonetic and the phonological becomes more tenuous. What exactly is the difference between a phonetic process applying to a single phonological entity, the output of which is two distinct, narrowly-specified phonetic distributions, and a phonological process changing the categorical feature specification of a given phonological entity in a given context such that the two resulting phonological representations are realized as two distinct, narrowly-specified phonetic distributions? In other words, how can we distinguish /a/ -> /a/ -[a] in stressed syllables and /a/ -> /a/ -[a] in unstressed from /a/ -> /a/ -[a] in stressed syllables and /a/ -> /a/ -[a] in unstressed? If /a/ and /a/ contrast underlying and have identical distributions in the syllables in question, the process must be assumed to be phonologically neutralizing. If there is no underlying /a/, however, the situation becomes ambiguous, and is not, I argue, resolvable through an appeal to phonetic distributions. In such situations we must consider the process to involve changes in phonological feature specifications only in such case as the output specification can be shown to be phonologically relevant to some other process. Where there is no evidence that the features in question are phonologically relevant (for the linguist and the learner alike), there is no reason to
assume that any change in the category membership of the phonological entity in question has taken place. If /a/, for example, is realized as something like [uː] in unstressed syllables and the process is non-neutralizing, we can assume either that the phonetic targets for /a/ in unstressed syllables are such that it is realized phonetically as [uː], or it might be that a phonological change to the input /a/ takes place such that it goes from [+low], [-high] to [-low], [+high], and is realized as [uː].

Evidence that the latter is the correct analysis might be susceptibility of the vowel in question to some phonological process targeting segments specified [+high], or equally failure to undergo processes targeting all segments specified [+low]. In the absence of such phonological evidence, however, there is no reason for the linguist or the native speaker/learner to assume that any categorical phonological process has taken place. A single phonological entity can then correspond to one set of phonetic targets or two distinct sets. Two phonological entities corresponding to the same set of phonetic targets, however, implies neutralization.

A situation in which two phonological entities overlap in a significant percentage of their realizations may not constitute phonological neutralization, then, but it is certainly a situation ripe for reinterpretation as such. Where the phonetic distinction between two categories becomes ambiguous in a large enough percentage of the tokens
of each in a particular structural position, that distinction becomes unstable, and the likelihood of the phonologization of the phonetic pattern as phonological positional neutralization increases.

The notion of sound change as reinterpretation based on the listener's misperception of the speaker's intentions has its source in the work of John Ohala (passim, but especially 1981 and 1993a.). On Ohala's conception of sound change, ambiguous phonetic realizations carry with them the possibility of changes in phonological representations. Briefly, a speaker utters a string /AXC/ - [AXC]. If the phonetic realization here is ambiguous, such that the distinction between it and a realization [AYC] of string /AYC/ is not perceptually robust, a listener may hear the token of [X], but misperceive it as [Y] due to their perceptual similarity. Believing the speaker to have intended to utter [Y] as a realization of /Y/, the listener then sets up a potential alternate realization of the string in question as /AYC/. In this particular context then, the contrast between /X/ and /Y/ has been neutralized. A simple example would be the development of umlaut-like vowel fronting such that /uCi/ becomes /yCi/ (or yC). In Ohala's schema, the speaker, with a UR of /uCi/, produces the coarticulated sequence [uCi] with significant fronting of the /u/. The listener then fails to attribute the fronting he or she hears to the influence of the following front vowel, hearing instead [y], and believing this to be the speaker's intended pronunciation, constructs a new representation for the form in question: /y/. If the language in question already contrasts /u/ and /y/ underlyingly, this results in the neutralization of the contrast between these in the position in question.

27
Hyman (1976) presents a model of phonologization in which he distinguishes three distinct stages of development. Phonologization in this view is conceived of as the intrinsic becoming extrinsic, such that automatic contextual variation is reinterpreted as intentional on the part of the speaker. Hyman distinguishes two distinct processes, the first of which he calls phonologization, and the second phonemicization. His example is consonantally-conditioned tone split, whereby naturally occurring perturbations of F0 after voiced and voiceless obstruents are reinterpreted as intentional, such that speakers begin to use tone contours as cues to voicing, exaggerating the range of the perturbations in the process such that they can no longer be attributed to universals of phonetic implementation. In Keating’s terms, however, this is the development of language-specific phonetic patterns. According to the model laid out above, while the exaggerated F0 perturbations are clearly intentional (targeted), there is still no reason to believe that they are phonological (meaning themselves encoded as distinct phonological categories. That they function as cues to the distinction between other phonological categories is clear enough). Hyman’s next stage, phonemicization, comes when the distinction between the voiced and voiceless consonants which originally conditioned the pitch differences is lost, yielding a contrast between segmentally identical strings differing only in their tonal contours. I will use the term phonologization throughout to mean specifically the innovation of changes to phonological representations, whether these result in the neutralization of contrasts or not. Hyman’s first stage of phonologization I will treat as the alteration or addition of phonetic targets to existing phonological representations.
In this study I advance the view that typological regularities involving the content of phonological patterns such as positional neutralization are best accounted for through an understanding of the phonologization process by means of which those patterns arise from their phonetic antecedents. I also argue that both the phonologization process and the synchronic implementation of the phonetic and phonological patterns at either end of that process are best understood in a split model of phonetics/phonology such as that described above.

1.5. Organization

The rest of this study is organized as follows: Chapter 2 presents the results of a survey of licensing asymmetries found between stressed and unstressed syllables. These effects are well-known in the literature as unstressed vowel reduction (UVR). The typological profile of UVR is quite unambiguous. In virtually all cases neutralization of contrasts occurs along the dimension of vowel height. While other features, such as frontness/backness or rounding are often involved as well, cases in which a system of unstressed vowel reduction is unambiguously based on the neutralization of palatality, rounding or ATR contrasts are exceedingly rare, if not utterly unattested\(^7\). The neutralization of vowel height in UVR systems has commonly been attributed to the

\(^7\) In the sense in which a UVR system reducing five vowels [i, e, a, o, u] to three vowels [i, a, u] is unambiguously based on the neutralization of height contrasts, or palatal harmonies are unambiguously based on the neutralization of contrasts of frontness or backness.
difficulty of reaching articulatory targets for non-high vowels in the shortened durations allowed vowels in unstressed syllables in most of the languages in question (Flemming 2001). Compression of the vowel space along this dimension due to undershoot of targets, together with a shorter window within which to correctly identify the unstressed vowel leads to the possibility of misperceptions by listeners leading to reinterpretation of speakers’ intended pronunciations. In this way the neutralization of height contrasts is phonologized. Nearly identical patterns of vowel reduction are presented in their phonologized and unphonologized states, in some cases in different positions within a single language.

In contrast to the relatively simple typological patterning found in stressed and unstressed syllables, Chapter 3 shows the array of effects found in final syllables to be substantially less uniform. There is robust crosslinguistic attestation of patterns of final syllable, and particularly final vowel resistance to neutralization processes for which the vowel of the final would otherwise be a legitimate target. The existence of these final resistance effects is attributed to the well-known phonetic pattern of domain-final lengthening, together potentially with articulatory strengthening of the type documented in pre-boundary elements in English by Cho (2001). The phonetics of domain-final position, however, unlike that of the stressed syllable, is not uniformly prominence-enhancing. Together with final lengthening and strengthening, final syllables, especially
above the level of the word, see dramatic drops in subglottal pressure. The significantly
lower intensity and pitch of the domain-final material, often accompanied by developing
non-standard phonation types can lead to phonologization of, among other things, a
pattern of weak licensing potential for the final syllable, such that in some languages
neutralizations take place in final position alone. The range of final syllable
neutralizations is catalogued and phonetic explanations for each are discussed. The
diversity of patterns found in domain-final syllables is a problem for an approach
assuming the inherent strength or weakness or structural positions to be encoded in UG.
An approach deriving the typology of positional neutralization from the phonologization
of common phonetic patterns, on the other hand, predicts precisely this ambiguous
behavior on the part of final syllables.

Chapter 4 presents an analysis of licensing asymmetries between initial and non-
initial syllables, the first regularity of which is the surprising rarity of unambiguous cases
of the positional neutralization of vowel contrasts in initial syllables. The majority of
cases presented as such in the literature turn out also to have (or have had at the crucial
moment in the past) fixed initial stress as well. It is therefore not possible in these cases
to say with certainty whether the licensing asymmetries in question are the result of the
initiality of the syllable or the stressedness of the vowel. The relatively small durational
asymmetries characteristic of fixed stress systems crosslinguistically would be unlikely to
result in the types of dramatic reduction processes discussed in the chapter 2. I argue, however, that smaller durational asymmetries of this sort, together with vowel-to-vowel coarticulation in languages with a particular morphological structure could result in the phonologization of precisely the types of harmony processes so frequently proceeding (synchronously or diachronically) from initial syllables. I also present the results of an experimental study showing that under certain circumstances even in the absence of initial stress a small durational asymmetry between the vowels of initial syllables and those of following syllables may be produced by the process of initial strengthening well-documented already for domain-initial consonants. Here too then the typology of positional neutralization effects falls out directly from the phonologization approach.

In chapter 5 I present my conclusions together with a discussion of the enterprise of phonological typology more generally. Here I compare the phonologization approach to typological patterning to approaches relying on restrictions on possible systems encoded in UG. Also discussed are predictions made by the Licensing-by-Cue and Direct Phonetics approaches to PN in light of the empirical findings presented in chapter two through four. I conclude that the synchronic connection these draw between the phonological regularities and the phonetic patterns which give rise to them is too immediate to account for certain aspects of the behavior of these systems both synchronically and diachronically. While typological patterns of PN are thus best
accounted for with reference to the phonetics, ironically enough it is probably the phonetics-free phonological approaches, overgeneration and all, which draw the most accurate picture of categorical PN in synchrony.
Chapter 2. Stressed syllables and unstressed vowel reduction

This chapter is an examination of the licensing asymmetries attested crosslinguistically between stressed and unstressed syllables. Unstressed vowel reduction (hereafter UVR), the topic of inquiry for most of this chapter, is perhaps the best known of the licensing asymmetries to be treated in this dissertation. Section 2.1 presents an overview of the typology of unstressed vowel reduction, identifying the main patterns a theory of vowel reduction must be able to account for. Specifically, it is shown that the overwhelming majority of UVR patterns are based on the elimination of height contrasts from unstressed syllables. Nasalization and quantity contrasts can also be involved. Systems based on the elimination of other contrasts, however, such as palatality, roundness, and ATR/RTR/Pharyngealization, features commonly operative in systems of vowel harmony, appear never to arise. Section 2.2 introduces the primary phonetic tendencies which give rise to vowel reduction patterns and shows how the phonologization model uses these to account straightforwardly for the typological patterns identified here.

Most treatments of vowel reduction make a strict division between “phonological vowel reduction”, the neutralization of phonological contrasts outside of stressed syllables, and “phonetic vowel reduction”, usually held to be a less dramatic process of centralization or raising of non-high vowels as a result of articulatory undershoot of the
sort demonstrated in Lindblom (1963) and discussed below. I maintain the distinction between the phonological and the phonetic here, arguing in fact that a number of UVR patterns commonly identified as phonological in the literature are in fact phonetic in nature. I demonstrate that these phonetic patterns are no less dramatic in their effect than their phonological counterparts. Furthermore, I argue that it is these phonetic patterns, through the medium of phonologization as described in Chapter 1, which account for the fact that most (but not all) categorical patterns appear phonetically motivated. The phonologization approach to typological regularities, I argue, gives us a broader and deeper understanding of the reasons for the existence of those regularities than do synchronic models with phonetic motivations built into them. Section 2.3 presents three case studies of unstressed vowel reduction systems which illustrate both the phonetic and phonological patterns of UVR, as well as the process of phonologization whereby the one is converted into the other.

Section 2.4 introduces an alternative approach to vowel reduction, that of Crosswhite (2001). This is a phonetically-driven Optimality-Theoretic study set in the Licensing-by-Cue framework advocated by Steriade (1994). This model is evaluated in light of the typological and phonetic findings presented here. Section 2.5 presents a case study of vowel reduction in Contemporary Standard Russian, a system which presents an interesting test case for the comparison of the Licensing-by-Cue and phonologization-
based approaches. On the basis of experimental findings, I propose a new interpretation of Russian vowel reduction based on the distinction between phonetic and phonological processes outlined above. Section 2.6 discusses the ramifications of the phonologization model of UVR typology for the question of phonetic determinism in phonology.

2.1. Unstressed vowel reduction: patterns of neutralization

Unstressed vowel reduction is a relatively familiar phenomenon in the linguistic literature, in part surely due to its robust attestation in such thoroughly studied language families as Romance and Slavic. A typological investigation of vowel reduction systems yields the following clear and striking results: The vast majority of licensing asymmetries between stressed and unstressed syllables in the languages of the world involve the neutralize of contrasts of vowel height, nasalization, or quantity. Neutralizations of features along any other dimensions, to the extent that they are at all are spottily attested and usually secondary to the neutralization of height contrasts in the same system. The enormous typological gaps this generalization leaves, including the absence of any systems of UVR based on neutralizations of palatality, ATR, or rounding contrasts, precisely the features most commonly involved in vowel harmony systems, must be accounted for in any theory of vowel reduction. As we shall see below, these gaps are
predictable according to the phonologization approach, while in other theories they remain mysterious.

2.1.1. Vowel height contrasts

Turning first to the neutralization of height contrasts, these are far and away the most commonly-effaced contrasts in systems of unstressed vowel reduction, forming the basis of virtually all the well-known Romance and Slavic systems noted above. A commonly cited example comes from Central Eastern Catalan, shown in (3).


<table>
<thead>
<tr>
<th>Stressed</th>
<th>Unstressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>i</td>
</tr>
<tr>
<td>e, e, a</td>
<td>e</td>
</tr>
<tr>
<td>u, o, o</td>
<td>u</td>
</tr>
<tr>
<td>říw 'river'</td>
<td>řiwéť 'river, dim.'</td>
</tr>
<tr>
<td>néw 'snow'</td>
<td>nawéť 'snow, dim.'</td>
</tr>
<tr>
<td>méř 'honey'</td>
<td>méléť 'honey, dim.'</td>
</tr>
<tr>
<td>pálő 'shovel'</td>
<td>poléť 'shovel, dim.'</td>
</tr>
<tr>
<td>řőů 'wheel'</td>
<td>řůďěť 'wheel, dim.'</td>
</tr>
<tr>
<td>móna 'monkey, fem'</td>
<td>munéť 'monkey, fem. dim.'</td>
</tr>
<tr>
<td>kůřć 'cure'</td>
<td>kuréť 'cure, dim.'</td>
</tr>
</tbody>
</table>
In Central Eastern Catalan, stressed syllables license contrasts within a seven-vowel inventory /i, e, ɛ, a, ɔ, o, u/, while unstressed syllables license contrasts among only three vowel qualities, [i, ɔ, u]. Systems with these particular inventories in stressed and unstressed positions are fairly common (particularly within Romance). We find the same seven-to-five inventory reduction in stressed and unstressed syllables in Berguener Romansh (Kamprath 1987: 52-54), posttonic syllables in Brazilian Portuguese ((Major 1981, 1985, Nobre and Ingemann 1987, Simões 1991, Wetzels 1992), and unstressed syllables in South Russian dialects (Flier 1978). Equally common are reductions from seven-vowels to five as in Standard Italian (Vincent 1988b), or Loniu (Austronesian - Admiralty Islands, Hamel 1994), and of five vowels to three, as in Eastern Bulgarian (Stojkov 1993), Standard Belorussian (Czekman and Smułkowa 1988), or posttonic syllables in Seediq (Austronesian - Atayalic, (Asai 1953, Li 1977, Li 1980, Holmer 1996). The fact that these same canonical inventories appear again and again in the strong and weak positions of systems of unstressed vowel reduction is good news for dispersion-based theories of vowel inventory construction (Liljencrants and Lindblom 1972, Flemming 1995, 2001). The same perceptual demands for maximal dispersion of vowels

---

38

---

Actually six vowels to three. See below.
within the vowel space appear to be operative in constructing inventories of unstressed vocalism as have been argued to govern selection of full inventories (Flemming 2001).

Interestingly, however, while stressed and unstressed inventories are remarkably steady across UVR systems, the mapping relationships between the vowels of the two inventories vary wildly from language to language. Thus, while in the Eastern Central Catalan seven-to-five reduction, /e, e, a/ merge as [ə], /u, o, ɔ/ merge as [u], and /i/ is realized as [i], in Berguener Romansh /i, e/ merge as [i], /u, o/ merge as /u/, and /e, ɔ, a/ are realized as [ə], and in Brazilian Portuguese posttonic syllables /i, e, e/ merge as [i], /u, o, ɔ/ merge as [u], and /a/ is realized as [ə]. By the same token, while in the Standard Italian seven-to-five reduction /e, e/ merge as [e], /o, ɔ/ merge as [o], and [a] is raised slightly toward schwa, in Loniu unstressed syllables /e, ɔ, a/ are realized as [ə] while /i, e, o, u/ stay more or less in place. Finally, in the five/six-to-three systems cited above, in Eastern Bulgarian /i, e/ merge as [i], /u, o/ merge as [u], and /a, a/ merge as [ə], while in Belorussian /e, a, o/ merge as [a–ə] leaving /i, u/ in place, and in Seediq post tonic /i, a/ are not neutralized, while /e, o, u/ merge as [u]. Partial, sometimes asymmetrical variants of these basic system types are also well-attested, all of which indicates the central role played by language-specific phonetic patterns of unstressed vowel realization in determining which vowels will ultimately undergo phonologized merger. Crucially,
however, in all cases systems are oriented around the elimination of height contrasts in
unstressed syllables.

Conspicuous in their absence are the unstressed vowel reduction analogues of the
common vowel harmony systems of the world. Specifically, UVR systems based on the
elimination of contrasts in frontness/backness, rounding, or ATR/RTR/pharyngealization
are spottily attested at best, if not utterly non-existent. There are, of course, plentiful
instances in which a front/back contrast is neutralized along with a height contrast, such
as in mergers of /a/ and /e/ as [ə] (as Steriade 1994: 11 notes for Catalan where the
resulting system is [i, u, ə]), but this could hardly be interpreted as a manifestation of the
need to eliminate front/back contrasts from weak positions.

Aside from this, we do find neutralization of front/back contrasts among low
vowels, as in Chamorro (Topping 1969, Steriade 1994, Crosswhite 2001), where in
unstressed syllables /æ/ and /a/ merge as [a]. Given the relatively small acoustic distance
between front and back low vowels, and the corresponding dispersion-based preference
against this contrast in smaller vowel inventories, it is unsurprising that specifically this
front/back pair should neutralize before others. The system of unstressed vocalism in
Chamorro is [i, u, a].

For further reductions of front/back contrasts to occur, it is necessary for the
unstressed inventory to fall below three vowels. Thus, in pretonic syllables in certain
southern Italian dialects such as that of Bari (Loporcaro 1997: 340-341), the only vowel contrast remaining is between unstressed [a] and [ə]. This pattern is also in accord with the crosslinguistic typology of vowel inventories, in which, alongside the famed vertical vowel inventories of Marshallese or Kabardian, we find no instances of "horizontal inventories", the only other option for an inventory smaller than /i, u, a/ but retaining some contrast still. The final option of course is the neutralization of all vowel contrasts, as in Saanich (Straits Salish, Montler 1986, Urbanczyk 1999), where nearly all unstressed vowels are realized as schwa.

In section 2.2 below I will argue that the predominance of height-contrast neutralizations in UVR systems follows directly from the phonetic characteristics of the systems in question. The virtual non-existence of systems based on the neutralization of front/back, round, or ATR contrasts is furthermore predicted under the phonologization approach, since except in the most extreme cases of durational contraction of unstressed syllables, these contrasts are not threatened phonetically to the extent that height contrasts are. While the phonologization approach predicts that the UVR analogues of palatal, rounding, or ATR harmonies should be rare, by no means does it preclude their existence entirely. This is because while the phonologization approach accounts for typological regularities in the patterning of PN systems through an understanding of the phonetic circumstances which give rise to them, a central tenet of the approach is that once

---

9 Another example of this neutralization pattern, in its slightly idealized form outlined in Steriade 1994, is that of word-medial short vowels in Hausa. See section 3.2.2.1 below for details.
phonologized, the developmental logic of the systems in question ceases to be phonetic in nature. Instead, it is constrained by other factors, what we might call grammatical, cognitive, or systemic.

Thus, UVR systems based on the neutralization of, e.g. palatality contrasts should not arise naturally through phonologization, since the phonetic patterns which would create such systems seem not to arise. Other diachronic developments, however, not constrained by phonetic factors, could ultimately conspire to produce inventories of this sort, just as I will argue below that other phonetically unmotivated correspondences can arise through analogical extensions or layered sound changes. In contrast to “direct phonetics” models of positional neutralization, however, in which the phonetic naturalness of the majority of extant phonological patterns is derived online in the synchronic phonological grammar, the phonologization model predicts only the improbability of the genesis of unnatural patterns. Should they ultimately come to exist by whatever circuitous path, however, the abstract, categorical phonology assumed here, devoid as it is of restrictions on the phonetic content of its patterns, should have no trouble implementing the typologically disfavored patterns. The case of the vowel systems of the Ob-Ugrian languages (Finno-Ugric) Khanty and Mansi provide evidence that this is so. In the Sosva dialect of Mansi, for example, in the non-initial syllable only /i, e, a, ə/ are contrasted (Honti 1988: 335), meaning that front and back vowels are contrasted only in the initial, stressed syllable. This system, predicted not to arise as the direct result of the phonologization of phonetic patterns, developed from an earlier system in which non-initial vowels were subject to palatal harmony, such that while both front and back vowels occurred, their appearance was conditioned by the frontness or
backness of the vowel of the initial syllable. In many (but not all) Ob-Ugrian dialects it is
the case that the ancestral harmony system has broken down in some way.

In Sosva Mansi, then, even prior to the merger of the front and back non-initial
syllable vowels which created the appearance of a system of UVR, the contrast between
the front and back was already neutralized. This preexisting lack of phonological contrast
along the front/back dimension may have predisposed the language to undergo the
reductions it did further along in its history. I know of no instance of the attestation of a
system similar to this one in which a healthy contrast between front and back vowels was
simply lost to unstressed vowel reduction in the same manner height contrasts are lost in
the development of canonical UVR systems. All this makes Sosva Mansi a typological
rarity, but there is no reason to think that this crosslinguistic fact in any way impedes the
functioning of the system in synchrony. This is unproblematic for the phonologization
approach to typology. Theories seeking to derive the phonetic “motivation” of
typological patterns through the existence of phonetic restrictions in phonological
Universal Grammar, on the other hand, in order to derive the near exceptionless height-
first rule for UVR systems must assume that the UG itself has something against systems
like Sosva Mansi. But while systems of this sort may be disadvantaged in evolutionary
terms, for the Mansi, presumably, they are acceptable enough.

Something similar to the Sosva Mansi system occurs in some dialects of Khanty,
such as Tromagan (Abondolo 1998). Additionally, in most Khanty dialects (as noted by
Steriade 1994), no round vowels occur outside initial syllables, and vowel harmony has
been lost, creating a system in which UVR neutralizes specifically round contrasts.

43

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
another rare configuration\textsuperscript{10}. The Proto-Ob-Ugrian vowel inventory for non-initial syllables contained front and back harmonic variants of ii, aa, and \v (Honti 1988: 331). Proto-Uralic, in fact, is reconstructed with palatal harmony affecting non-initial-syllable vowels of only two heights, written as *[i, i, ä, å] (Sammallahti 1988: 481). Discounting the harmonization, this system is identical to the two vowel system attested in cases of UVR such as the pretonic syllables of Bari Italian (given as [ə, a] by Loporcaro), in which remaining front/back contrasts are neutralized while a single height contrast remains (the next step being the reduction of that last height contrast, yielding schwa). The lack of round vowels in Khanty non-initial syllables is thus a continuation from this earlier state, one which is both attested and expected in the typology of UVR systems. It does not, therefore represent a legitimate instance of a (phonetically-driven) UVR system based on the elimination of roundness contrasts in unstressed syllables. Again, though, this is no reason not to consider it such in synchrony. Ob-Ugrarian vowel systems are fascinating and richly deserving of increased attention by phonologists.

Another putative instance of a UVR system based on the elimination of palatality contrasts are the Upper and Lower Carniolan dialects of Slovenian as described by Lencek (1982: 146-151) and cited in Crosswhite (2001). In these dialects, which contrast short and long vowels, the short high vowels /i/ and /u/ fall together with underlying /a/  

\textsuperscript{10} Proto-Ob-Ugrian, like Proto-Uralic, is reconstructed with fixed initial stress (Sammallahti 1988). Khanty, however, appears to have at least some instances of stress attraction heavy peninitial syllables in an otherwise fixed initial system (Honti 1988). This may mean that synchronically harmony-free dialects of Khanty should not be considered systems of UVR, though at the relevant historical period at least they may have been.
(in Upper Carniolan underlying schwa has a tendency to delete altogether). Crucially, though, these mergers are not related to the placement of stress, but are rather unconditioned mergers of these three short vowels in both stressed and unstressed positions. Among Lencek's examples are even such stressed monosyllables as Upper Carniolan [nët] (cf. Standard Slovene [nit]) and Lower Carniolan [sët] (cf. Standard Slovene [sit]). It is true, of course, that unstressed syllables share with short vowels the property that they are the shorter of two sets of vowels in the system, and hence both may be subject to similar pressures. But they differ in important respects as well, perhaps most importantly in that unconditioned mergers in the short (or long) vowel inventories are not examples of positional neutralization. They take place across the boards, regardless of the structural position in which they are located. In positional neutralization systems such as UVR, the two opposing sets of vowels, stressed and unstressed, are by definition in complementary distribution. Long and short vowels in, e.g., Slovene, are a different matter. They may both occur in stressed and unstressed syllables, and because of the potential for paradigmatic contrast of this sort (unavailable in UVR systems), they may exert pressure on one another in the manner that high vowels and mid vowels of the same quantity may also, potentially occasioning chain shifts and the like. That such a merger should take place in the short vowel inventory of Slovene dialects is interesting\footnote{Doubly so in fact for this being the second occurrence of this sound change in the history of the language, the first instance being in common with Serbo-Croatian the merger of the overshort Late Common Slavic Jer vowels (originally short /i/ and /u/), yielding Slovenian /a/ and SCR. /a/.}, and the fact that similar mergers seem absent from UVR systems points out a potential difference in the relationship between long and short and stressed and unstressed vowel inventories,
but this is just one of many differences we do not yet fully understand (see Steriade 1994: 14 and 22 for some notes here) in the development of these systems.

One clear difference here is that long vowel inventories are routinely both larger and smaller than short vowel inventories, while in UVR systems stressed inventories are rarely if ever smaller than unstressed. Languages with contradictory patterns from this point of view listed in Maddieson (1984) include Telefol (long /iː, eː, aː, ɔː, uː/, short /i, a, u/) and Chipewyan (long /iː, aː, uː/, short /i, e, a, o, u/. Obviously in each instance radically different factors may have contributed to the development of the systems in question. The point is that the relationship between long and short inventories is of a different nature altogether, and that phenomena affecting this relationship and that of stressed and unstressed vowels in UVR systems may be similar, but are not directly comparable. That stress shifts in Slovene may result in categorical alternations between contrastively long high vowels and [a] (formerly a short high vowel) indicates the disconnect between potential synchronic alternations and the phonetic patterns commonly giving rise to vowel reduction, but does not necessarily add to the list of patterns that phonetic motivations need account for.

A more convincing instance of high vowel front/back neutralization in unstressed syllables only comes from the history of Welsh (Bosch 1996, Williams 1989). In Northern Welsh there is a synchronic pattern whereby final [i, u] alternate with penultimate [a]. Synchronically this case is complicated by the fact that stress in Northern Welsh is penultimate, making this an instance of non-augmenting reduction in strong
position. Additionally, there are lexical exceptions to the reduction of [i]. It appears that non-reducing [i] is actually original, while reducing [i] is derived from an earlier /y/, which later merged with /i/. The original reduction, then, was of /y/ and /u/ to [ə]. The case is complicated, and in any event not an instance of the elimination of all front/back contrasts even among the high vowels (since original /i/ and /i/ remain contrastive in non-final syllables).

Tamil (Beckman 1998) presents a case in which no round vowels occur outside the stressed initial syllable (see chapter 4 for a discussion of stress in Tamil and Dravidian at large. Why Beckman chooses to treat Tamil as an instance of initial prominence rather than UVR is not obvious), contrary to the generalization made here. In the dialect Beckman treats, of the five underlying short vowels /i, e, a, o, u/, only /i, a, u/ appear contrastively outside the initial syllable (Beckman 1998: 87). Additionally, in non-initial syllables, the three peripheral vowels receive the allophonic realizations [i, ɔ, ɯ]. It is thus not so much the case that the system avoids round/unround contrasts in unstressed syllables, but rather that the allophonic unrounding of short /u/ and the predictable UVR elimination of the mid vowels conspire to produce a system in which no

---

12 This stress, however, is not duration-cued (Williams 1989). In fact, Welsh final syllables are routinely longer than stressed penults. This is a result of the fact that stress was previously final, but upon retraction to the penult, left behind traces of that earlier stage in the form of increased duration.
round vowels appear. It is potentially relevant here too that it is short /u/ in particular which is affected in this way. Even in the initial syllable the short peripheral vowels undergo some qualitative phonetic reduction, being realized as [ɪ, ə, ʊ] respectively.

The special vulnerability of the short high vowels in systems with long/short contrasts deserves serious further investigation. One potential avenue of exploration would focus on the durational characteristics of these, particularly in connection with the well-known crosslinguistic tendency for short high vowels to be subject to devoicing. Another direction might explore the qualitative issues, in particular the perceptual effects of the laxing short high vowels often undergo, presumably to keep them distinct from the long high vowels in the event of shorter phonetic realizations for these latter. The vulnerability of -ATR high vowels in ATR systems may be of relevance here as well.

The generalization, in any event, is clear: While UVR systems eliminating height contrasts from unstressed syllables are abundant throughout the languages of the world, systems of UVR eliminating contrasts in palatality, rounding, or ATR/RTR or Pharyngealization in a reduction equivalent to the commonly attested harmony systems appear to be completely unattested. While the list of height-neutralizing systems is obviously in no way meant to be exhaustive, the list of potential exceptions is. A central claim of this chapter is that this hole in the typology is no accidental gap, but rather
follows from the phonetic characteristics of UVR systems in a manner consistent with the
phonologization approach to typological patterns. Details follow in 2.2.

2.1.2. Nasalization

Another contrast found neutralized in unstressed syllables is vowel nasalization. In Copala Trique, for example, nasal and oral vowels contrast only in the stressed syllable (Hollenbach 1977, Steriade 1994). The same is true in Guarani (Beckman 1998: 158). Reduced durations in unstressed syllables may go some way toward explaining the neutralization of the contrast between nasal and oral vowels outside stressed syllables. It is well-known that, crosslinguistically, nasalized vowels tend to be significantly longer than their oral counterparts (Whalen and Beddor 1989). If longer duration is needed for accurate perception of the contrast, its neutralization in shorter unstressed syllables is understandable. Stressed vowels may also be subject to less coarticulation than unstressed vowels, increasing the likelihood that assimilations would flow from the former to the latter, rather than vice versa, as argued by Majors (1998) in her treatment of stress-based harmony systems.

2.1.3. Quantity

Quantity contrasts behave somewhat schizophrenically with respect to stressed syllables in the languages of the world. It is a commonplace that certain languages require stressed vowels to be bimoraic, lengthening them when necessary, and thus
neutralizing the contrast between stressed and unstressed syllables. This type of process is what Smith (2002) calls augmentation, and cites as a characteristic of stressed syllables. The idea here is that it is more important to enhance a stressed syllable’s phonetic prominence, ensuring perception of the placement of stress than it is to maintain segmental contrasts within the stressed syllable. It does seem clear that some languages value the heaviness of a stressed syllable over the preservation of the contrast between monomoraic and bimoraic vowels. From the phonologization perspective, we would expect these to be languages in which a vowel quantity contrast was forced to coexist with strongly duration-cued stress. The phonetic lengthening of short vowels under stress (common in many languages without long/short contrasts, such as Russian) would ultimately make the contrast difficult to perceive, eventually resulting in misperceptions and concomitant reanalysis of the lengthened short vowel as an intended distinctively long vowel. This phonologization, of course, neutralizes the contrast. It is not surprising, giving that stressed syllables are, if anything, a lengthening environment, that the output of this neutralization should be a long vowel. See the discussion of vowel length contrasts in final syllables in Chapter 3, however, for a more complicated case.

In addition to languages neutralizing quantity contrasts in stressed syllables, however, there are also attested a fair variety of languages which neutralize length contrasts in all syllables except the stressed syllable. Kolami (Emeneau 1961: 6-7) and
several other Dravidian languages are examples. Here the initial syllable is stressed, and long and short vowels do not contrast in non-initial syllables. In Warumungu (Bosch 1991), as in Dravidian, stress is fixed initial, and contrastively long vowels occur only here. In Jamul Tiipay (Miller 2001), vowels contrast long and short only in the stressed and immediately pre-stress syllables, with all vowels outside these positions being short. Also in Burushaski (Anderson 1997), vowel length contrasts exist only in stressed syllables, with all unstressed syllables licensing only short vowels (though this case is not without complications). What these patterns suggest is an origin in languages in which unstressed syllables were at least somewhat restricted durationally, enough to cause sufficient shortening of unstressed long vowels that the contrast should be lost. Stressed syllables, however, cannot have been expanded durationally to a great extent for the quantity contrast to remain intact. This prosodic profile actually fits the Dravidian languages remarkably well. Many Dravidian languages are described as having some degree of initial stress, or some immediate derivative, but this stress, as in Tamil (Schiffman 1999, Bosch 1991:183), is not reported to be marked by a great durational asymmetry. Durational curtailment of non-initial syllables, however, is demonstrated in the frequent patterns of syncope of non-initial vowels in various patterns throughout the histories of various Dravidian languages, e.g. Kannada (Schiffman 1983) Kurux (Pfeiffer 1972, Gordon 1976) and Malto (Gordon 1976). In Toda, in fact, a language which also
underwent massive syncope in non-initial syllables, Emeneau describes a length contrast with a durational ration of about 1:2 in initial syllables. Toda also has strong initial-syllable stress. In non-initial syllables, by contrast, Emeneau characterizes the long-short distinction as more between "short and less short", with the distinction being variable and often difficult to perceive. This leads Emeneau to speculate that the contrast may be in the process of disappearing.

The preceding sections have shown that the only clearly attested patterns of neutralization involving stressed and unstressed syllables are those affecting primarily height, nasalization, or quantity. For each of these features, duration is either a well-known cue for the contrast in question, or in the case of quantity, potentially the only cue for the contrast in question. Patterns of UVR conspicuously not attested include those based on contrasts of palatality, roundness, and ATR, all commonly active in harmony systems (see chapter 4), though not known to be strongly cued by duration. The following section explores the phonetic basis of UVR in more detail.

2.2. The Phonetics of Vowel Reduction

The fact that unstressed vowel reduction should so definitively prefer targeting height contrasts is perfectly comprehensible in light of the phonetic environment in which it generally arises. The correlation between the presence of UVR in a language and duration as a primary correlate of lexical stress has been widely recognized for decades (see, e.g. Lehiste 1970). Generally speaking, UVR appears in languages with a large
durational asymmetry between stressed and unstressed syllables, such that unstressed syllables undergo significant durational contraction relative to a substantially longer stressed syllable, particularly under increased rate of speech.

Flemming (to appear) follows Lindblom (1963) and others in attributing neutralizations of height contrasts in durationally impoverished syllables to the comparatively longer duration needed to achieve the open upper vocal tract required to produce a higher F1. One way of adapting to the durational limitations of unstressed syllables would be to speed up articulatory movements sufficiently so that the more open target articulations of non-high vowels could be reached in the time allotted them. Flemming attributes the failure of a language to implement this alternative to a desire to minimize articulatory effort where possible. The other option, short of expending that additional effort or violating restrictions on the durations of unstressed vowels is to articulate the vowel in such a way that it falls short of its articulatory targets.

In this new, more crowded vowel space, of course, perceptual problems produced by spectral overlap in the realizations of unstressed vowels, are compounded by the shorter time the listener has to correctly identify a comparatively indistinct vowel. This situation causes frequent misperceptions, which can ultimately result in the loss of the contrasts in question. This reinterpretation is the phonologization process which produces systems of unstressed vowel reduction. While section 2.1.1. demonstrated the enormous crosslinguistic variation found in the merger patterns between the stressed and unstressed vowels of otherwise identical inventories, the merger patterns attested are not completely

53
arbitrary. Rather, each set of mergers represents plausible groupings of vowels according to perceptual proximity to one another cause by one or another language-specific pattern of shrinking of the vowel space due to durational pressures on articulatory targets. Which is to say, patterns of merger reflect the directions in which vowels were being shifted under durational pressure in unstressed syllables. This can be seen in action in the case studies of Brazilian Portuguese and Bulgarian below, in which different dialects show phonologized and unphonologized versions of the same patterns.

Since it is chiefly the implementation of vowel height contrasts which is hindered by the decrease in duration of the unstressed syllable, it is to be expected that unstressed vowel reduction would primarily impact contrasts of vowel height, leaving other contrasts, such as those not found to be active in systems of UVR for the most part unscathed. The phonologization approach thus accounts for the fact that the contrasts most often involved in vowel reduction systems are precisely those which would be most severely impacted by serious durational impoverishment. Recent UG-based approaches to vowel reduction, on the hand, have not even remarked on the existence of the typological generalization.

\[\text{\cite{Crosswhite}}\text{ points out that merger patterns are not predictable from the synchronic positioning of the vowels relative to one another in the stressed syllables of the languages in question. This is certainly true, and underscores the extent to which synchronic patterns of PN are independent from the phonetic factors which created them. To understand the development of a given UVR pattern, though, it is necessary to consider not the phonetic realizations of the stressed counterparts of the reducing vowels (since these by virtue of those very realizations are not in fact subject to mergers), but to the realizations of the unstressed vowels themselves immediately prior to the phonologization of the synchronic neutralizing patterns.}\]
Another way in which the phonetics can create the input to phonologizations resulting in stressed-unstressed syllable licensing asymmetries is through the durational and articulatory enhancement of the stressed syllables. Cho (2001) presents articulatory evidence from English that the presence of pitch accents on vowels results in their strengthening not only in duration and resistance to vowel-to-vowel coarticulation, but also in what Cho calls sonority enhancement, whereby lip and jaw position are both more open, and in many cases tongue position is lower as well (but see de Jong 1995 for somewhat contradictory results). These enhancements, taken together, can account for lengthenings, lowerings, and diphthongizations of stressed vowels, potentially resulting in the emergence of new contrasts in those positions (assuming that some comparable stressed vowels fail to undergo the process, or enter the language only after the process has taken place).

That it is specifically duration or lack thereof, and not "stressedness" per se, that is responsible for the licensing asymmetries found in stressed and unstressed syllables is underscored by cases in which vowels of one or another degree of stress are nonetheless subject to reduction due to the failure of that stress to impart any additional duration to the vowel in the syllable it is realized on. Thus, in Italian, only vowels under primary stress are immune to vowel reduction, while secondarily-stressed vowels undergo the neutralization of unstressed high and low mid vowels to [e] and [o] just as would an unstressed vowel (Crosswhite 2001: 206). Experimental evidence from Farnetani and Kori (1990) provides an explanation for this fact, in that vowels with secondary stress are seen not to differ significantly from unstressed vowels in duration, while the vowels of stressed syllables are regularly longer than both.
It should be noted also, though, that while durational impoverishment is strongly correlated with vowel reduction, it is not necessarily the case that additional duration alone will rescue an otherwise-reducing unstressed vowel from its fate. This is the conclusion reached by Johnson and Martin (2001) in their experimental study of phonetic vowel reduction in Creek, where the additional duration supplied to final vowels by phrase-final lengthening does not cause them to resist the centralization characteristic of unstressed syllables in the language. Nord (1987) reaches similar conclusions for Swedish. The role of final lengthening in inhibiting UVR in many languages (along with its failure to do so in others) is treated in detail below in Chapter 3.

The phonologization approach advanced in this study derives the typological characteristics of UVR systems from language-specific patterns of phonetic undershoot developing in unstressed syllables under articulatory pressure cause by durational impoverishment. We should therefore expect to find in the languages of the world plentiful instances of patterns of phonetic vowel reduction in which vowels of different heights come to approximate one another's positions in the vowel space, but have yet to undergo phonologization as categorical UVR. Such systems are in fact well-attested. In fact, in addition to clear cases of phonetic vowel reduction and phonological vowel reduction, we also find systems in which both patterns are attested together. The following sections present case studies of each type.
2.3. Case Studies

2.3.1. The phonologization of phonetic vowel reduction: Bulgarian

The first case study I present is that of unstressed vowel reduction in Bulgarian. Bulgarian is an especially valuable example of UVR in the continuum of its dialects from west to east illustrates eloquently the way in which gradient phonetic patterns over time become phonologized, yielding categorical patterns of neutralization. Some of the westernmost dialects of Bulgarian lack qualitative vowel reduction more or less completely. Unstressed non-high vowels are realized with qualities not overly different from those of stressed non-high vowels. Other western dialects, such as that of the Sofia region, have varying degrees of phonetic raising of non-high vowels, without, however, neutralizing the relevant contrasts (at least in the case of the mid and high vowels). Further east, the phonetic raising pattern has undergone phonologization, such that Eastern Bulgarian has a fully developed system of categorical UVR. Details are as follows.

Bulgarian has six vowels in stressed syllables: /i, e, a, â, o, u/. In unstressed syllables in Eastern Bulgarian, the mid vowels raise to neutralize with the high vowels as [i] and [u]. The low vowel /a/ raises to merge with /â/ as [a]. I use the standard transcription symbol ‘â’ to transcribe the non-low central vowel of Bulgarian. Many
scholars transcribe it as /a/, which unfortunately gives the impression that this vowel is qualitatively identical to the unstressed vowel representing the merger of /a/ and /a/ in most dialects. Crosswhite (2001), for example, assumes on this basis that stressed /a/ is featureless, with its realization interpolated between the gestural targets of adjacent segments. In fact, however, the stressed realization of this vowel and the unstressed vowel representing the merger of /a/ and /a/ are not identical in quality. The phonetician and dialectologist Vladimir Zhobov of Kliment Oxridski University in Sofia describes it (personal communication) more in the direction of cardinal vowel 15 ([ɤ]). The Bulgarian Academy of Sciences grammar edited by Tilkov (1982: 50) shows a clear degree of constriction in the velar region present in the case of stressed /a/ and absent for its unstressed counterpart. Of the unstressed vowel relative to the stressed, Tilkov says the unstressed version “is characterized by a greater openness of the oral cavity, such that the tongue has a lower position, and thus a greater degree of closure in the pharyngeal cavity” (Tilkov 1982: 50). Of its acoustic characteristics, he continues, “this change in articulatory configuration is reflected spectrally through a concentration of acoustic energy through raising of the frequencies of the first formant, and a corresponding lowering of the frequencies of the second formant - the characteristics of a more open vocalization”14. The stressed vowel is very explicitly lowered to merge with unstressed

14 The original reads as follows: The unstressed vowel “se xarakterizira s po-goljama otvorenost na ustnata kuxina, tâj kato ezikât ima po-nisko položenie, i sâotvetno s po-goljama zatvorenc na gårlenata kuxina”. And further, “V akustichen plan tazi promjana v konfiguratsijata se otrazjava vârxi spektâra v 58
/a/. This merger makes sense as the result of centralization, whereby less jaw opening for /a/ and less velar constriction for /â/ brought the two close enough together that misinterpretations occurred, resulting in loss of the contrast, and neutralization as actual [ə].

The rise of the Eastern Bulgarian system becomes clear when viewed in conjunction with material from other Bulgarian dialects, such as those underlying the standard language. In Eastern Bulgarian UVR, there are three categorical neutralizations, yielding a phonological three vowel unstressed inventory. In the standard language, and most of the western and southern part of the country, however, this is not the case. Wood and Pettersson (1988) characterized the situation in general terms as follows: “The actual occurrence of reduction in everyday speech is subject to normative, social, dialect and morphological constraints. Unstressed /a/ is reduced in all dialects. Complete reduction of both /e/ and /o/ is limited to eastern dialects, but they are reduced to a varying extent elsewhere, depending on the informality of the situation. The reduction of /o/ is common in the Sofia dialect, but reduction of /e/ is less usual there (Wood and Pettersson 1988: 240). In Pettersson and Wood (1987), they provide more details, but decide to describe all the systems in terms of a single phonetic tendency: “However, we prefer to see the

59

sâsredotochavane na akustichna energija chrez povishavane na chestotite na pârviya formant, i sâotvetno chrez ponizhavane na chestotite na vtoriya formant - ukazanie za po-otvoren glasezh” (Tilkov 1982: 50).
regional distribution as a case of varying generalization of a single process, a reflection of
how the Bulgarian speaker's willingness or unwillingness to make certain reductions is
subject to a complex of stylistic, sociolinguistic and dialect forces (P & W 1987: 268).

This holistic approach to the dialect continuum and social fabric is fine of course, for
certain aims, but is unacceptable as the basis for a model of the grammar of a single
speaker of a single dialect, if that speaker's dialect simply lacks certain of the categorical
reductions represented in the holistic take. Crosswhite (to appear), who describes
"Bulgarian" with the three neutralizing reductions characteristic only of Eastern
Bulgarian dialects, notes in a footnote that "In the pronunciation norm of Sofia and other
western areas of Bulgaria, vowel reduction is weaker than in the eastern areas, or even
entirely absent. Since Sofia pronunciation defines the standard, many vowel-reducing
speakers attempt to suppress vowel neutralizations when speaking in formal registers. In
particular, suppression of /e/ > [i] is quite common, and the lack of this suppression is
rather stigmatized" (Crosswhite, to appear: 51). The general point concerning greater
degrees of reduction in informal speech and lesser degrees in formal speech is correct, as

15 Neutralizing reduction of all vowels, though /e/ in particular, as well as automatic palatalization of
consonants before front vowels are indeed features of eastern dialects subject to parody in other parts of the
country, as in the mocking phrase [as sam ud daì'eku, i guvo're a mì'eku], "I am from far away, and speak
with automatic palatalization of consonants before front vowels" (lit. 'softly').
is the suppression of the reduction of /e/ by speakers of eastern dialects wishing to avoid
detection. But speakers of dialects such as that of Sofia and virtually the entire west and
south of the country are not “suppressing” reduction of /e/ or categorical reduction of /o/
under some sort of social pressure to conform to the literary standard. They are speaking
their native dialects, which happen not to include these features in their grammars. A
speaker whose dialect is close to the standard in this respect (e.g. a Sofian) will at least in
informal speech merge /a/ completely with /á/. The same speaker may even realize some
percentage of utterances of unstressed /o/ in informal speech as overlapping more or less
completely with unstressed /u/. But that speaker will not, no matter how informal, lazy,
drunk, or socially maladjusted, suddenly adopt categorical merger of /e/ and /i/ in
unstressed syllables, because no register of this dialect contains such a process in its
phonological or phonetic grammar.

Crosswhite’s description of neutralizing reduction in Bulgarian is attributed to the
results of an experimental study on the speech of two Bulgarians from villages near Sofia
presented in Wood and Pettersson (1988). Describing these results, she writes:

Pettersson and Wood started their investigation by first verifying the existence of
acoustically neutralizing vowel reduction using spectrographic evidence - the formant
frequencies measured for unstressed /e,o,a/ were found to coincide with those
measured for the vowels /i,u,a/, respectively. Since there was no acoustic difference
in vowel quality reflecting these underlying vowel contrasts, we can state that Bulgarian vowel reduction is acoustically neutralizing” (Crosswhite 2001: 42).

But this is simply false, both as a characterization of vowel reduction in the Sofia dialect, and as a characterization of the experimental findings of Pettersson and Wood (1987). What Pettersson and Wood actually report is the opposite\(^{16}\): On pages 244-247 they show scatterplots of their results for unstressed vowels spoken in isolation, and also in frame sentences. Unsurprisingly, there is a tendency toward more reduction in the frame sentences than in forms spoken in isolation. For one speaker in the isolated words even unstressed /a/ and /â/ form clearly distinct, if overlapping, distributions. Since they provide no statistical analysis, it is hard to see whether otherwise /a/ and /â/ form one distribution or two. /o/ and /u/ overlap for both speakers, but in both contexts many of the realizations of /o/ fall well outside the distribution for /u/. Again, statistical analysis would say for certain. For /œ/ and /ũ/, however, there is very little overlap for either speaker at either rate, to say nothing of a single distribution. Most importantly, to the

---

\(^{16}\) It may be that the confusion is in their use of the term neutralizing, which they employ in several places in an articulatory sense, meaning “becoming more neutral”. For example, on p. 240, of the failure of mid and low vowels to merge completely in many dialects they say the following: “... selective relaxation of articulatory control and neutralization of components of the unreduced vowel would diminish articulatory precision and lead to a spectral transition between unreduced and fully reduced forms.” Describing as they are a gradient “transition” from one vowel to another, they clearly do not mean neutralization in the sense of Aufhebung. They then continue, “Alternatively (italsics mine), switching between complete configurations, by substituting [i] for [e] etc. in the underlying motor programme would lead to discrete shifts between reduced and non-reduced forms”, viz. actual neutralization.
question of phonological neutralization and categorical merger the authors say directly, evaluating their results (in the context of an earlier study by Lehiste and Popov): “Lehiste and Popov summarize the vowel spectra as means, which gives the impression that vowel reduction in Bulgarian is discrete (Though the means of Lehiste and Popov also reflect almost no reduction of /e/ - JB). But the raw data recorded in Figs 1-4 indicate that reduction is gradual” (Wood and Pettersson 1988: 251). Also on pages 258-259: “The two informants were typical of the reported tendency in their dialect not to reduce unstressed /e/ as much as /a, o/. One informant tended to be more formal when reading isolated words. ... Reduction was gradual rather than discrete, suggesting that an underlying articulatory plan for a complete rendering is executed with varying degrees of attention, depending on which components are neutralized and how far”. This is all fairly unambiguous.

So vowel reduction in Bulgarian, except in its eastern dialects (specifically the Balkan dialects, so-called, meaning the mountains, not the peninsula), is not neutralizing. It is gradient and apparently rate-dependent. Intriguingly, Tilkov (1982: 46), characterizing the likelihood and degree of application of this reduction in the standard language, states that unstressed vowels reduce the least in the first pretonic syllable (what he calls the first degree of reduction) and more so in non-first pretonic syllables and posttonic syllables. This is the same phonetic trend as is found in East Slavic and discussed below in the context of Russian and Belorussian, only apparently far less
dramatic. Tilkov provides no quantitative demonstration of this trend, unfortunately\(^{17}\).

Seen from a dialectological perspective, of course, the different Bulgarian dialects do represent a single whole with respect to UVR: They preserve a number of stages along a possible progression of phonologizations of gradient phonetic patterns. In the west, nearer to the borders with the Former Yugoslavia, dialects have little if any unstressed vowel reduction (missing entirely in standard Macedonian and Serbo-Croatian, present to some extent in certain dialects nearer to Bulgaria). In Sofia there is gradient phonetic raising of /a/ and /o/, possibly with the former phonologized as a merger for some speakers. Little or no raising of /e/, however, is attested. Variants on this situation are found across the country (as well as pockets of interesting innovations and/or archaisms which are beyond the scope of this study). Finally, in the eastern dialects, the phonetic reduction of /e, a, o/ has become dramatic enough and consistent enough that it has been phonologized, resulting in categorical phonological mergers of those vowels with /i, u, a/.

\(^{17}\) Still though, a weak, gradient version of the East Slavic pattern present in Bulgarian, where in most dialects (with certain exceptions in the Rhodopes) the directions of neutralization are completely different from those of East Slavic would be a fascinating testament to the staying power of minute aspects of the phonetic detail of a given prosodic system to survive the complete destruction and transformation of that system and to linger in a context in which they are otherwise wholly alien and phonologically unexpected.
2.3.2. Categorical vowel reduction: Belorussian

An example of phonologized, categorical unstressed vowel reduction comes from Belorussian (Czekman and Smutkowa 1988). In Belorussian, five vowels, /i, e, a, o, u/, in stressed syllables are reduced to three vowels, [i, a, u], in unstressed syllables. Specifically, in unstressed syllables /e, a, o/ merge, all being realized as [a] or [e], depending on the position of the vowel relative to the stress. Crucially, in Belorussian, as in Eastern Bulgarian, vowel reduction is neutralizing. It takes place whenever the non-high vowels occur in an unstressed syllable, regardless of the phonetic duration of the nucleus vowel. There is also no question of gradience. The mid vowels are simply realized as the contextually appropriate allophones of /a/. Reduced tokens in which, for example, /o/ is more [a]-like or more [o]-like depending on its duration are not found.

The diachronic situation in East Slavic is complex and will not be resolved definitively here. It is possible, however, that the mergers of /e/, /o/ and /a/ never took the route of gradual phonetic approximation to eventual phonologization of merger in the first place. To dramatically distill the opposing points of view on the matter, one account holds that southern East Slavic dialects lost the distinction between /a/ and /o/ in unstressed syllables as a result of vowel reduction. Evidence of this includes the relatively late appearance of orthographic evidence of the merger in East Slavic documents (14th century - Vlasto 1985: 316-317). Another view holds that this merger
(known as ‘akanje’, ‘[a]-saying’, in the Slavic literature), is actually of more ancient provenance, being not a reduction of unstressed vowels at all, but rather a failure of the contrast between /a/ and /o/ ever to emerge outside stressed syllables. Stressed [a] and [o] in the relevant dialects are ultimately derived from Late Common Slavic long /a:/ and short /a/ respectively. The rephonologization of this earlier quantity distinction as one of quality took place at a period in which similar changes were affecting the rest of the vowel system as well, essentially the breakdown of the Common Slavic prosodic system. While in most areas the split of the long and short low vowels took place in all syllables, irrespective of the placement of stress, it has been argued, most famously no doubt by Shevelov (1964), that in the relevant dialects of East Slavic the split occurred only under stress, while in unstressed syllables the quantity difference collapsed to yield a single [a]-like vowel of one or another degree of reduction18.

18 Crosswhite (2001) presents a similar account of the mergers of the tense and lax mid vowels in Italian (which if true would hold for all other Romance languages with this distinction, such as Brazilian Portuguese). The distinction between tense and lax mid vowels appears to have been neutralized already in Vulgar Latin, as they fail to show distinct reflexes in any daughter languages (Hall 1976). Whether this merger took place before or after the relevant contrasts shifted from being quantity-based to quality-based (/e, o/ vs. /ɛ, ɔ/ instead of /e, o/: e/ vs. /ɛ, ɔ/) is not entirely clear. This shift must have been underway by the time some of the early inscriptions at Pompeii, in which we find attested the merger of [i] (← Latin short /i/) and [e] (← Latin long /e:/) as [ɛ] (Vincent 1988: 33). A similar merger of the corresponding back vowels takes place later, attested in all of Western Romance, but not in Romanian. These mergers of original short and long vowels of different (but adjacent) heights makes no sense if the vowel system is still quantity-based, but if durational differences were not as significant by the time of the merger, with distinctions based already primarily on quality, then these sound changes are perfectly comprehensible.
Whatever the case diachronically, however, the synchronic status of the system as one of categorical UVR eliminating contrastive mid vowels from unstressed syllables is clear enough. While the neutralization is not conditioned in synchrony by the phonetic properties of the host syllable, incidentally, the realization of the allophone of /a/ resulting from the merger is subject to some duration-dependent variation. In the first pretonic syllable, the vowels of which are characteristically longer than other unstressed syllables in Russian and Belorussian, /a/ is realized as [a]. In non-first-pretonic unstressed syllables, however, contrary to the claims of many descriptions, it is realized somewhat differently. Czekman and Smułkowa (1988: 226) symbolize the vowel in question as Greek ‘α’, describing it as weaker and higher than stressed [a], and coming about via the reduction of [a].

2.3.3. A mixed system: Brazilian Portuguese

Brazilian Portuguese is a system in which both phonologized and unphonologized vowel reduction patterns are in evidence, though for some speakers the phonetic pattern appears already to have become categorical. The facts of Brazilian Portuguese vowel reduction are generally described as follows (Major 1981, 1985, Nobre and Ingemann 1987, Simões 1991, Wetzels 1992): the stressed vowel inventory is /i, e, e, a, ə, o, u/. In pretonic syllables, however, the contrast between high mid and low mid (or tense and lax mid) is neutralized, yielding [e, o]. In posttonic syllables, the mid vowels
are eliminated entirely and /a/ is raised to schwa, yielding the inventory [i, u, ə]. The neutralization occurring as a result of what is often called the first degree of reduction, the prohibition on realization of a contrast between /e, o/ and /e, ə/ in any unstressed syllable, is categorical (Major 1985). No amount of additional duration, emphasis, or formality of register will cause it to reemerge in syllables where it is neutralized. It must, therefore, be represented by the collapse of the two independent entities into a single category in the phonology. The second degree of reduction (occurring in posttonic syllables), however, is more ambiguous.

Major notes that for many speakers degree two reduction of /e, o/ to [i, u] does not take place in phrase-final open syllables. He demonstrates that these syllables are also subject to substantial phrase-final lengthening, which exempts the degree 2 vowels from reduction, making them durationally (and spectrally) equivalent to the vowels of pretonic syllables. Furthermore, the realization of the pretonic syllables as [e, o], which Nobre and Ingemann show to be somewhat reduced qualitatively in comparison with their stressed counterparts in any circumstance, can in certain speech styles undergo further phonetic reduction. Major notes that while in the "normal" register pretonic mid vowels are realized as (slightly reduced) [e, o], in "casual" speech these vowels are reduced further.
still, being realized instead as [i, u] also\(^\text{19}\). Major hypothesizes that the reason for the difference in register is one of speech rate, which is to say, the vowels pronounced at the faster rate have shorter durations. The reduction of [e, o] to [i, u], being duration-dependent, applies pretonically only when those vowels are shortened sufficiently to condition it. Posttonic vowels, on the other hand, are routinely short enough, unless altered by phrase-final lengthening. Other processes are also sensitive to speech rate in this way: in normal speech, nasalized mid vowels are never reduced, but in fast speech in posttonic position only they are raised to merge with the high nasalized vowels. Major attributes their failure to raise in normal speech to the fact that, as is typical crosslinguistically, the nasalized vowels are phonetically longer than their oral counterparts all things being equal, such that only a substantially faster speech rate would contract them sufficiently to begin to challenge the articulation of their target heights.

While degree one reduction is irreversible and categorical, degree two reduction, the raising of mid vowels clearly demonstrates gradient properties. The duration-dependence of the process clearly indicates that it is not yet independent of phonetic

\(^{19}\) On normal and casual, Major writes "NOR is the natural speech used in settings which vary from slightly informal to formal - such as a lecture, a newscast, or consultation with one's colleagues; it is the style which the layman considers good or correct pronunciation. CAS is used in very informal, casual, or intimate settings, e.g. conversation between good friends and lovers. The layman often considers this to be incorrect and sloppy" (Major 1985: 265). long. Pretonics in the same experiment were in all cases longer than these posttonics, but shorter than the tonics.
factors in the way that degree one reduction appears to be. Approaches treating both reduction processes in Brazilian Portuguese as equivalently phonological have no means of characterizing the failure of reduction in phrase-final vowels and the fast-speech reduction of pretonic vowels as a single process. By assuming the gradient, duration-dependent raising of Brazilian Portuguese mid vowels to be carried out entirely in the phonetics, we gain a more accurate, unified picture of the processes at work here. We need only assume that mid vowel raising is a function of phonetic vowel duration, and that the language-specific phonetics assigns different durations to syllables according to their position in the word relative to the primary stress, their position within the phrase, and rate of speech, none of which should be controversial.

As a final note, Major points out that not all speakers allow for the non-reduction of mid vowels in phrase-final syllables despite the additional duration supplied by phrase-final lengthening. He hypothesizes that for these speakers posttonic raising has become "lexicalized", or phonologized in the terms applied here. For speakers who have phonologized this merger, the raising becomes categorical, sensitive now only to the structural description of its host syllable, rather than to factors like the additional duration supplied by phrase-final lengthening.

The previous sections have presented examples of similar UVR systems on both ends of the phonologization process, demonstrating how patterns of vowel realization
produced by phonetic durational pressures are converted into categorical neutralization in the phonology through phonologization. Since durational pressures impact the realizations of some features more severely than others, and vowel height in particular, vowels come to approximate one another's positions creating perceptual ambiguity only along those dimensions. Since phonetic ambiguities arise primarily in the dimension of vowel height, it is precisely height contrasts which will be lost to the phonologization of UVR. Since contrasts in frontness/backness, rounding, or ATR are not dependent on longer durations articulatorily to the same extent as height contrasts are, we would not expect vowels to develop ambiguous realizations to the same extent, except under more extreme durational pressures. In such situations though, height contrasts will be abandoned as well, yielding the typological picture presented above in 2.1, whereby, for example, front/back contrasts may be lost in addition to height contrasts in unstressed syllables, but will not be lost instead of them. That other duration-sensitive contrasts such as those of nasalization and quantity should also be sensitive to the stressedness of the syllable in which they appear is predicted by the phonologization model in the same way. As the cases of Sosva Mansi and Tromagan Khanty illustrate, however, the phonologization approach predicts only the evolutionary improbability of the disfavored UVR patterns. Because of its diachronic take on typological patterning, however, should patterns of this sort nonetheless come to exist through channels other than straight
phonologization of duration-dependent phonetic patterns, the phonologization model in no way prejudices the phonology against their implementation.

2.4. A UG-based approach to UVR typology: Crosswhite (2001)

Crosswhite (2001) presents the most comprehensive account of vowel reduction patterns to date in the generative tradition. An Optimality-theoretic approach to unstressed vowel reduction, this study seeks to account both for the synchronic implementation of UVR systems in the constraint-based phonology, but also through the reranking of constraints or constraint-types, to produce a factorial typology generating all and only attested patterns of vowel reduction. The assumption that the phonological grammar must be responsible both for modeling individual competence and for producing a full accounting of crosslinguistic phonological typology is central to the program of Optimality Theory (Prince and Smolensky 1993). It is also, I would argue, a serious mistake. In the following sections I introduce the account of phonological UVR presented in Crosswhite (2001), and compare its typological predictions to those of the phonologization model.

In addition to contributing a comprehensive typological survey of vowel reduction patterns in the languages of the world, Crosswhite (2001) seeks, on the basis of evidence accrued therein, to divide vowel reduction systems into two major types. These types are
said to differ both in their phonetic motivations and formal implementations, and the
distinction between them for Crosswhite represents the primary typological
generalization to be made for UVR systems. The two types are named contrast-enhancing
and prominence-reducing UVR.

Contrast-enhancing vowel reduction is a neutralization process whereby
"undesirable or perceptually challenging vowel contrasts are limited to stressed position"
(Crosswhite 2001: 22). In this way it is a direct descendent of Steriade’s Licensing-by-
Cue approach to Positional Neutralization. The idea is that contrasts are deployed by
speakers only in positions in which the phonetic cues which make those contrasts
perceptually robust are present and strong. For vowel-quality contrasts, Steriade (1994)
identifies duration as the phonetic cue most indispensable for accurate perception. In
positions of lesser duration, then, only the most easily perceptible vowel contrasts will be
licensed, while in, e.g., stressed syllables, which are in many languages associated with
additional duration, the entire inventory of contrasts may be realized. Crosswhite follows
Steriade in this, presenting arguments showing why specifically mid vowels are targeted
by so many reduction systems, and why the resulting systems of unstressed vocalism are
so often /i, u, a/. The arguments are familiar from the phonetic literature on crosslinguistic
regularities in the shape of vowel inventories: she cites specifically optimal dispersion in
the vowel space (Liljencrants and Lindblom 1972, Flemming 1995), the quantal stability
73
of the corner vowels, meaning the tendency to keep stable acoustic realizations even with a degree of articulatory variation (Stevens 1986), and the tendency for certain pairs of formants to approach and enhance one another in each of the corner vowels (Stevens 1986), increasing their perceptual salience. Because of this perceptual salience, the vowels most favored by contrast-enhancing UVR will be peripheral [i, u, a] 20.

Crosswhite formalizes these observations in OT using Licensing constraints à la Steriade, formulated as follows:

(5) \( LiC-Q/\beta: \) The vowel quality \( Q \) is only licensed in context \( \beta \).
Where \( Q = \) any vowel quality or a natural vowel class
\( \beta = \) any context that enhances the perception of \( Q \)

In the case of UVR, a typical contrast-enhancing constraint is (6):

(6) \( LiC-\)Noncorner/\)Stress: Noncorner vowels are licensed only under stress.

The effect of this constraint is to ban the realization of mid vowels categorically from unstressed syllables in a manner quite like that illustrated in chapter one for Positional Markedness constraints. An example of a system produced by the interaction of this

20 Varying rankings of potential Licensing constraints with Faithfulness constraints will see that this is not always the system contrast-enhancing UVR ultimately generates. Different outputs depend on language-specific constraint rankings.
constraint with general Faithfulness constraints on the relevant features is Belorussian. As described above, Belorussian contrasts five vowels /i, e, a, o, u/ under stress, but in unstressed syllables, /e, o/ lower and merge with /a/, yielding [a]. Another example Crosswhite cites of contrast-enhancing UVR comes from Standard Italian, in which a seven-vowel stressed system /i, e, ë, a, ò, o, u/ is reduced to five vowels [i, e, a, o, u] in unstressed syllables. Most important here is the fact that contrast-enhancing UVR is argued specifically not to be a product of duration-dependent undershoot of the sort described above. It is rather a strictly perceptual phenomenon whereby naturally difficult contrasts are eliminated from positions of lesser phonetic prominence. The main evidence that this is so is the putative appearance of true corner vowels in contrast-reducing systems. Thus, in Belorussian and Italian, unstressed /a/ is not raised as dramatically toward schwa as it is in, for example, Bulgarian.

The immediate problem with contrast-enhancing UVR lies in the typological predictions it makes. Contrast enhancing reduction is said to take place as a response to an imperative to eliminate of “perceptually difficult” contrasts from positions in which a lack of the necessary cues (here, duration) would lessen the likelihood of their correct parsing by the listener. For this reason, it prefers peripheral vowels, explaining why Belorussian leaves /a/ as [a], while Bulgarian raises it to schwa. The problem comes in the definition of what constitutes a perceptually-difficult vowel contrast? Crosswhite
argues that mid vowels are less robust than corner vowels perceptually, and hence prone
to effacement, and this is surely true. But previous accounts of vowel neutralization in
phonetically-driven phonology have given a very different answer to this question.

Steriade in particular, in her seminal 1994 paper on positional neutralization, cites
the work of Kaun (1993) and Suomi (1983, 1984) on the typology and phonetic
underpinnings of vowel harmony systems. Steriade divides harmony systems into those
which are articulatorily-driven, and those which are perceptually-driven. Perceptually-
driven harmony systems include those based on front/back, round, and ATR distinctions.
It is this which motivates the dictum “bad vowels spread” (Steriade 1994: 23). Kaun
(1993: 73-80) discusses the relative perceptibility of F2-based distinctions vs. F1 based,
concluding that F1 based distinctions are perceptually more robust (probably due to the
higher intensity of F1). Kaun cites as evidence for this the existence of vowel systems
allowing contrasts only along the F1 dimension (the famous “vertical” vowel systems of
Kabardian, Marshallese, and others), and the corresponding non-existence of “horizontal
systems”. Steriade supports this view, noting an insight of Gorecka’s that “all cases of
wide-spread harmony (round, ATR, front) involve featural contrasts that are relatively
hard to identify; on the other hand, height, the most salient vocalic distinction, rarely, if
ever, leads to across-the-board harmony” (Steriade 1994: 23).
If this is correct, then, and height truly is the most salient vocalic distinction, and if it is correct that “contrast-enhancing” vowel reduction is motivated by the need to remove perceptually-difficult contrasts from weak positions, then it is utterly perplexing why all vowel reduction systems, contrast-enhancing or otherwise, should focus so consistently on height. It is even worse for this theory that instances of UVR systems based on precisely the contrasts Steriade identifies as most difficult perceptually (front/back, round, ATR) should be all but completely unattested in the languages of the world. Steriade 1994 actually notes the absence of UVR systems based on the elimination of roundness or ATR contrasts, but judges it to be an accidental gap. As argued above, however, given the phonetic factors underlying vowel reduction, there is nothing accidental about it. In sum, either Steriade and Kaun are wrong, and height contrasts are actually less robust perceptually than contrasts of palatality, roundness, and ATR, or there is more to vowel reduction than just the elimination of difficult contrasts from contexts where duration is insufficient for their accurate perception. Crosswhite’s treatment of vowel reduction does not account for the fact that the virtually all UVR systems deal in neutralizations of either height, nasalization or quantity. By contrast, this striking pattern is a primary prediction of the phonologization approach to the typology of UVR systems.
The second type of vowel reduction system Crosswhite introduces, prominence-reducing UVR, is of a different character altogether. While the purpose of contrast-enhancing reduction is to remove difficult-to-perceive contrasts from positions in which they are likely to be misperceived (such that there is little point troubling oneself to deploy them in the first place), the point of prominence reduction is said to be the removal of “particularly loud and lengthy vowel qualities” from unstressed syllables. The basic idea behind prominence reduction is that prominent segments should be aligned to prominent positions, while non-prominent segments should be aligned to non-prominent positions. Crosswhite conceives of prominent segments as those which are more sonorous, where sonority for her a measure derived from a combination of phonetic factors, including at least duration and low-frequency amplitude (2001: 37). For the purposes of vowel reduction, this measure characterizes differences in vowel height, with the caveat that unstressed schwa, conceived of as a placeless, targetless vowel, is less sonorous even than the high vowels. Prominent positions for our purposes here are stressed syllables. Crosswhite is careful in all this, incidentally, to distinguish prominence reduction from the gestural undershoot of Lindblom (1963), discussed above. Briefly, her reasons are this: “...in vowel undershoot, decreased articulation time leads to a change in vowel quality, and can be traced to a desire to avoid effortful articulations (i.e. ones in which articulator movement must be fast). In contrast, in prominence-reducing vowel
reduction, a change in vowel quality leads to a decrease in articulation time" (p. 46). Her argument is with the employment of articulatory undershoot to derive categorical UVR patterns synchronically. In OT terms, prominence reduction is implemented in a manner similar to that which Prince and Smolensky (1993) use to derive the role of sonority in syllabification syllable. Crosswhite assumes two "phonetic" scales, one of accentual prominence, the other of vowel prominence, as shown in (7)\textsuperscript{21}:

(7) Phonetic Scales for prominence-reducing UVR

a. Accentual Prominence:
   \[ \text{stressed} \overset{\text{prom}}{\rightarrow} \text{unstressed} \]

b. Vowel Prominence
   \[
   \begin{align*}
   a & \overset{\text{prom}}{\rightarrow} \varepsilon, \varepsilon \overset{\text{prom}}{\rightarrow} e, o \overset{\text{prom}}{\rightarrow} i, u \overset{\text{prom}}{\rightarrow} e
   
   \end{align*}
   \]

The first scale encodes the fact that stressed syllables are more prominent than unstressed, while the second encodes the familiar Sonority Hierarchy. These scales are then "crossed" to produce the relevant constraint families, resulting in fixed rankings of markedness constraints such as that shown in (8):

\[ \text{The scales assumed by P \& S were Syllabic Prominence, distinguishing the Peak from the Margin, and Segmental Prominence, based as here on sonority.} \]

\[ \text{79} \]
Prominence reduction family of constraints

*Unstressed/a >> *Unstressed/e,e >> *Unstressed/e,o >> *Unstressed/i,u >>*Unstressed/a

The result is a set of constraints banning high sonority vowels from unstressed positions in a ranking from most sonorous to least. Crossing the scales in the other direction (producing, e.g., *stressed/a) would create constraints that could be used to attract stress to sonorous vowels, or to raise the sonority of stressed vowels, à la Kenstowicz (1994). Crosswhite exemplifies prominence-reducing UVR with Bulgarian, presented, as noted above, in the form of its eastern dialects, in which neutralizations are in fact categorical. See section 2.3.1 for discussion of the crucial differences between systems of UVR in Bulgarian dialects.

Bulgarian, as presented by Crosswhite, has six contrastive vowels under stress: /i, e, a, a, u/. In unstressed syllables, however, the mid vowels are raised to merge with the high vowels as [i, u], and the low vowel /a/ is raised to merge with the mid central vowel as /a/\(^{22}\). In other words, constraints against prominent (non-high) vowels in unstressed positions are

---

\(^{22}\) It is interesting to note that most of the dialects which actually have this system of vowel reduction also have automatic palatalization of consonants before front vowels, which may not be an accidental correlation in light of the reduction of /e/ to [i] present here and absent or weak in other dialects. There are, however, dialects with more pronounced palatalization than in these and only weaker, non-neutralizing reduction of mid vowels: the correlation is thus not automatic.
syllables interact with Faithfulness constraints to cause the raising of /e, a, o/ to merge with /i, ə, u/.

Crucially, the way in which this system differs from that of Belorussian (aside from the obvious raising vs. lowering of the mid vowels), is in the realization of the low vowel as schwa. The fact that the high sonority vowel [a] is not allowed in non-prominent positions is what marks Bulgarian as prominence-reducing rather than contrast-enhancing.

Crosswhite notes, importantly, that it is not always possible to tell a prominence-reducing UVR system from a contrast-enhancing system on the surface. Many of their effects can be the same; for instance, both systems disfavor, and have a tendency to eliminate, mid vowels. Contrast enhancement does so because they are non-peripheral, and hence less perceptually robust, and prominence reduction does so because they are fairly sonorous. As noted, in a comparison of Belorussian and Bulgarian, the only diagnostic for the underlying formal nature and phonetic motivation for the vowel reduction pattern instantiated in each is the presence or absence of schwa. This for Crosswhite is a shibboleth. Schwa is extremely low in sonority, making it good for prominence reduction. That Bulgarian UVR produces a schwa means that it is prominence-reducing. Schwa, on the other hand, is no good for contrast enhancement,
which favors peripheral vowels. The fact that Belorussian has [a] in unstressed syllables therefore means that it is contrast-enhancing.

That the one system raises mid vowels and the other lowers them is irrelevant, since in both UVR types either pattern can be generated by simple reranking of Faithfulness constraints. A system with raising of mid vowels and no raising of the low vowel, however, would be completely ambiguous, since reduction of /a/ to schwa can actually be blocked in a prominence-reducing system by ranking the relevant Faithfulness constraints (Max [+low]) above even *Unstressed/a. The possibility of such cases in which it is a priori impossible to determine, presumably for linguist and learner alike, which grammatical mechanism is responsible for generating such a pattern is troubling. This is not, however, the only instance in which the line between the two types of reduction systems becomes blurred.

Firstly, contrast-enhancing UVR is not so simple in its implementation as it might seem. In the cases of Standard Italian, Belorussian, the pretonic syllables of Brazilian Portuguese and the immediate pretonic syllable of Russian (introduced below), the low vowel, designated as the diagnostic for system type, was said to be peripheral, and was transcribed [a]. This, however, turns out to be a substantial idealization of the phonetic facts.
In Standard Italian, for example, where all unstressed syllables witness a merger of the tense and lax mid vowels, the vowel transcribed [a] is in fact substantially raised in most unstressed syllables. Farnetani and Vayra (1991) provide experimental evidence showing that F1 values for the unstressed vowels of two (northern) speakers of Standard Italian are on average decreased by 28% (974 Hz vs. 700 Hz) for one speaker and 18% (669 Hz vs. 548 Hz) for another. Furthermore, this experiment found a high degree of correlation between vowel duration and F1 values (for the two speakers r(133) = 0.783 and r(130) = 0.774 respectively. Rather than implementing a maximally (or even steadily) peripheral vowel, then, here the realization of /a/ is a function of its duration. The reason it is not as high as Bulgarian schwa (for example) is therefore not because it is specified [+low], but because the pressure of the duration decrease is usually just not extreme enough to force it to raising the full distance to schwa. In a case such as this, it is not even clearly meaningful to draw a distinction between “schwa” and “a”, unless we wish to define an arbitrarily-selected F1 value as the border, beyond which a system ceases to be contrast-enhancing, and becomes prominence-reducing instead. The fact that the realization of /a/ is gradiently determined by duration is even worse, of course, since even such an arbitrarily-selected F1 boundary has no meaning if F1 of the vowel in question is so loosely targeted as to be essentially a function of another property, such as duration.
The same problem arises for the pretonic syllables of Brazilian Portuguese, which Crosswhite argues to undergo contrast-enhancing reduction. Nobre and Ingemann (1987) give experimental results showing the mean F1 of stressed and pretonic /a/ for four Brazilian Portuguese speakers to differ by nearly 100 Hz (stressed /a/ - 684 Hz, pretonic /a/ 585 Hz). Flemming (2001) notes of the tense and lax mid vowels that the neutralized realizations are centralized in the F2 dimension, while in F1 falling in between the stressed realizations of the tense and lax phones. To this may be added the facts observed by Major concerning the tendency for the second degree of reduction (raising of non-high vowels) to apply in pretonic syllables at faster rates of speech.

Again, since the low vowel of pretonic syllables is realized somewhere between /a/ and [a] as a function of duration, and the mid vowels apparently behave similarly, it is unclear how to draw a clear distinction between this type of system and one like Bulgarian. I submit that the difference between these systems is not one of formal implementation or phonetic motivation at all, but rather merely one of degree. Durational pressures in unstressed syllables in Italian and pretonic syllables in Brazilian Portuguese have simply not been severe enough to condition consistent drastic raising of the non-high vowels. In the posttonic syllables of Brazilian Portuguese, which Crosswhite treats

---

23 As noted above, it is not entirely clear whether the contrast between today's tense and lax mid vowels was eliminated in Vulgar Latin unstressed syllables at a time when it was still primarily a quantity contrast, or whether it had already taken on the form it has in the stressed syllables of the modern languages. Either way, however, insufficient duration to maintain the contrast is clearly the driving force behind the merger.
as prominence-reducing because of the raising of the non-high vowels, as discussed
above, for some speakers this pattern is simply not phonological in the first place, while
for others durations have been low enough to cause significant raising consistently
enough that the phonetic pattern has become phonologized. If there is a serious
dichotomy to be found here, it is between the phonologized and the unphonologized
versions of a variety of processes all ultimately arising from the same phonetic pressures.

The Belorussian situation is somewhat more complex, due to the unresolved
questions discussed above surrounding the diachronic source of the mergers of the mid
and low vowels. To review, the problem was that it is not clear whether mid and low
vowels once contrasted in unstressed syllables in East Slavic, but later fell together, or
whether the contrast between the mid and the low vowels only ever emerged in stressed
syllables to begin with, such that diachronically at least, no vowel reduction per se ever
took place.

Recall that in Belorussian immediate pretonic syllables /e, a, o/ are realized as
[a], while in other unstressed syllables the neutralization is the same, though the resulting
[a] undergoes some qualitative reduction, causing Czekman and Smulkowa (1988) to
symbolize it as [α], perhaps IPA [e]. Here again the ambiguity in the system arises: Since
some raising toward schwa does in fact occur, should this be considered contrast-
enhancing or prominence-reducing vowel reduction? Has non-first pretonic /a/-/o/-/e/
been raised enough to count as raised, or is it still peripheral, despite the raising?

85
Of course, the problem still remains that the Belorussian first pretonic syllable does, according to Czekman and Smulkowa, actually contain phonetic [a], so at least there the contrast-enhancement analysis seems solid. For the phonologization approach, however, this could be a problem, since if the low vowel is not forced by durational pressures to raise, becoming perceptually ambiguous with the reduced variants of the mid vowels, then it is not clear why neutralization takes place to begin with.

If it is correct that, however, that there never was a contrast between mid and low vowels in the unstressed syllables of East Slavic, then this issue disappears. Even under the opposite hypothesis, though, it is still possible that the unreduced [a] of the Belorussian pretonic syllable is a secondary development. Note that the realization of [a]−[e] in the first pretonic syllable varies substantially in both Russian and Belorussian dialects, in some being more like a stressed /a/, and in others being reduced to one extent or another. Again there is much dispute as to the original quality of the system, but most scholars assume that either the one (uniform pretonic [a]) or the other (uniform pretonic reduced [â]) was original (see Chekmonas 1987 for argumentation). Timberlake (1993: 435) presents a scenario in which other, more complicated dialectal reduction patterns can be plausibly derived if we assume the original first pretonic syllable to have contained an ‘a’ with some degree of reduction and shortening (he writes [d̂i]). If this is
the case, such that lengthening and lowering of the outcome of the merger of the mid and low vowels in all first pretonic syllables is a secondary development in Standard Belorussian, then the neutralization makes sense even as a result of ordinary duration-driven articulatory UVR.

2.5. Vowel reduction in Russian

The system of UVR in contemporary standard Russian is an excellent test case for the two approaches. Crosswhite treats Russian (like Brazilian Portuguese) as displaying two degrees of vowel reduction. The first and less extreme of these she analyzes as contrast-reducing, while the second, more extreme degree of reduction she treats as prominence-reducing. The treatment of Russian as having two distinct degrees of vowel reduction is in fact the standard treatment in the Slavic grammatical tradition. Generally, however, the difference is treated as one of degree, rather than one kind. Crosswhite’s innovation is to treat the two degrees of vowel reduction differing not merely in the vowels they allow expressed, but in the very substance, both in terms of formal implementation and phonetic motivation of the reduction process itself.

In what follows, I will lay out the facts of the Russian case. After brief discussion of Crosswhite’s approach to it, I will present a different account of Russian vowel reduction. Specifically, on the basis of experimental results to be presented below, I will
argue that there is only one phonological vowel reduction process in Russian: degree one reduction. Degree two reduction (as in Brazilian Portuguese for some speakers) is an unphonologized process of duration-dependent raising applying to all unstressed syllables, affecting those with shorter durational targets more dramatically than those with longer targets.

2.5.1. Facts

Contemporary Standard Russian (CSR) is usually described as having two "degrees" of vowel reduction. Degree 1 reduction applies to the first pretonic syllable (but see below for details), while Degree 2 applies elsewhere. The generalization usually cited is this: /a/ and /o/ are categorically neutralized in all unstressed syllables. In the first (immediate) pretonic syllable, both vowels are realized as [a]. In other pretonic syllables farther from the stress, and in all posttonics, they are realized as [ə]. In all unstressed syllables, /e/ is raised to merge with /i/. /a/ and /o/ following a palatalized consonant reduce according to the same pattern as /e/. There are further details concerning /e/, palatalization, and /i/ which will not concern us here. This is exemplified in (9).
(9) UVR in Contemporary Standard Russian

<table>
<thead>
<tr>
<th>Stressed</th>
<th>Degree 1 reduction</th>
<th>Degree 2 reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>молодёст⁴</td>
<td>youth</td>
<td>молодёжъ⁴</td>
</tr>
<tr>
<td>боль⁴</td>
<td>pain</td>
<td>боль⁴</td>
</tr>
<tr>
<td>старый</td>
<td>old (adj.)</td>
<td>старик</td>
</tr>
<tr>
<td>разум</td>
<td>reason</td>
<td>разумно</td>
</tr>
</tbody>
</table>

Under Crosswhite's analysis, the Degree 1 reduction occurring in the first pretonic syllable is contrast-enhancing reduction, since mid vowels are eliminated, but the low vowel is realized peripherally, rather than as schwa. Degree 2, exemplified above in the second pretonic syllable does reduce /a/ and /о/ to [ə], and therefore is analyzed as prominence-reducing. Note that the reduction of the result of the merger of /a/ and /о/ from [a] to [ə] is non-neutralizing. The merger takes place in all unstressed syllables, with different realizations depending on the position of the vowel relative to the stressed syllable. No additional mergers take place in Degree 2 reduction.

The details of the case, however, are more complicated than the picture sketched above would lead one to believe. The first complication concerns the realization of what is transcribed [a] in the first pretonic syllable. While speakers of some dialects, at least in certain circumstances, realize a pretonic [a] with little if any reduction⁴, standard

---

⁴ A "drawled", strong [a] under a rising pitch curve in certain contexts is a stereotypical feature of Muscovite speech.
handbooks describe a first pretonic realization of this /a/ (<- /a/, /o/), which seems to have undergone some degree of qualitative phonetic reduction. Though not reduced as much as they would be in non-first-pretonic syllables, it is not clear that the reduction they do undergo is completely different in kind from that more dramatic raising pattern. Sources are vague as to mean formant frequencies. Matusevich transcribes first pretonic /a/-/o/ as [ʌ], which is not, however, meant as the IPA value of the same symbol. Articulatorily, compared to /a/, he writes that “the tongue ... is pulled slightly back, while (importantly) the tension of the organs of speech is reduced” (Matusevich 1976: 100). It is substantially shorter than stressed /a/, with a mean duration given as 90 ms. to the 196 ms. of its stressed counterpart. It is said to be perceptually “further back and more closed than a stressed ‘a’, less tense and shorter”. Avanesov (1968) says pretonic /a, o/ is “not always identical, but generally produced with a bit less lowering of the lower jaw, and therefore not as wide opening of the mouth, as for a stressed [a]” (Avanesov 1968: 51). He cautions, however, that it must nonetheless remain “a-like” in character, i.e. not raised all the way to schwa.

25 Matusevich (1976) gives “F1 around 700 Hz and F2 around 1000” for first pretonics and stressed vowels, while claiming that the former is perceptually (‘na vosprijatie’) “more back and more closed” than its stressed counterpart, which seems odd. However, Matusevic also lists F1 of the non-first-pretonic unstressed [a] as “around 700”, suggesting that these figures may not be completely reliable.
What is clear from this is that the vowel in question is subject to some phonetic reduction, a fact which, as in the case of Italian and Brazilian Portuguese, questions the extent to which a clear distinction can be drawn between the reduction in question and the reduction found in, e.g. Eastern Bulgarian.

There are further complicating factors in the realization of degree two reduction as well. Firstly, in absolute word-initial position /a/ and /o/ are not in fact reduced to schwa, but rather are realized as the same sort of [a] as is found in the first pretonic syllable, which I will transcribe from here on as [u]. This is not unexpected, given the correlation between reduction and duration we have seen in action elsewhere. Vowels in onsetless initial syllables are seen in a variety of languages to be longer than their counterparts in initial syllables which have onsets. As discussed in 3.2.3., absolute word-initial vowels in Nawuri (Casali 1995, Kirchner 1998) fail to undergo an assimilation process involving roundness of which they would otherwise be targets. Long vowels and vowels in phrase-final position likewise avoid this assimilation, all three being associated with additional duration. In Tamil, likewise, absolute word-initial vowels have been shown experimentally to be longer than vowels in word-initial syllables with onsets (Balasubramanian 1981). In Luganda, the contrast between long and short vowels is neutralized in onsetless word-initial syllables. Additionally, only three vowels contrast in this position (in fact, initially in any morpheme): /e, o, a/ (Hubbard 1994: 161-165). While

91

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Hubbard's durational measurements for vowels in this position are equivocal as to the relative durations of these vowels, in at least certain cases they are longer than they would be in internal positions. A similar example operating at the phrase-level comes from Runyambo, in which initial /i, u/ are lowered to [e, o] after pause (Larry Hyman, p.c.). This lowering of the high vowels to of the quantity distinction is most likely the result of boundary-adjacent strengthening of the sort documented by Cho (2001), which is discussed at length in Chapter 3. It is possible that this effect is a result of the tendency for domain-initial segments to be strengthened both in duration and magnitude of gestures. See chapter 4 for discussion of this effect.

Unsurprisingly in light of the reduction data, vowels in absolute initial position in Russian are durationally enhanced as well. Phonetically, then, there is plenty of crosslinguistic evidence for the durational enhancement of absolute word-initial vowels, and the failure of these to reduce to schwa in Russian makes perfect sense if reduction to schwa is a function of vowel duration. Crosswhite, for whom reduction to schwa in Russian represents the inability of non-first-pretonic vowels to bear moras, derives this non-reduction with a constraint mandating the alignment of a mora to the word-edge.

---

26 Phonetic lowering substantial enough to lead to reinterpretation of high vowels as mid in particular would be a consequence of what Cho calls “sonority enhancement” and documents as part of the complex of strengthening effects found in English in boundary-adjacent and accented syllables.
Reduction to schwa also fails to apply in situations of hiatus before [a]. Thus, we find [saatnejneije] (*[saatnejneije]), 'relationship'. This too makes sense under the assumption that reduction to schwa in Russian is not carried out in the categorical phonology at all, but rather is an instance of gradient, duration-dependent phonetic raising, just like the less dramatic effect described above in, e.g., Italian. If the problem with producing [a] in an unstressed syllable with consonants on either side is lack of sufficient time for the jaw-lowering gesture to reach target, the realization of two identical vowels with no intervening consonant would solve this problem handily, giving ample time for the gestural target of the low vowel to be reach. This exception to the reduction of /a/-/o/ to schwa thus follows from the same duration-based phonetic principles as failure to reduce in absolute initial position: reduction occurs where phonetic durations are substantially impoverished, and fails where additional duration is realized. Crosswhite derives the failure of /a/-/o/ to reduce to schwa in hiatus to a constraint *Hiatus([a], [ə]), which states that the featureless vowel [ə] may not occur in hiatus with [a].

---

27 Crosswhite, for whom reduction to schwa in Russian represents the inability of non-first-pretonic vowels to bear moras, derives this non-reduction with a constraint mandating the alignment of a mora to the word-edge.
In addition to all this, there is one additional position in which non-first-pretonic /a/ and /o/ fail to reduce to schwa not treated by Crosswhite. In phrase-final open syllables, /a/ and /o/ are realized as [a], just as they are in the first pretonic syllable. This realization is said to depend on style, speech rate, and to be gradient, sometimes producing a vowel in between Degree 1 and Degree 2 reduction (Matusević 1976: 102, Zlatoustova 1962: 109-139). Non-reduction only occurs when the vowel is final in the phrase, with phrase-internal word-final vowels reduced to schwa along with the other non-first-pretonic unstressed vowels. In section 3.2.3, I attribute this rate-dependent failure to reduce to the additional duration provided by phrase-final lengthening, robustly attested in Russian. Again, failure to reduce to schwa is predictable according to the phonetic duration of the vowel.

Each instance of the failure of reduction to schwa in non-first-pretonic syllables is explicable as a function of the phonetic duration assigned to the vowels in those syllables. The additional duration in each of the above cases is furthermore either part of a well-documented crosslinguistic trend (initial strengthening, final lengthening), or simply falls out from the nature of the sequence being realized ([a.a]). There is nothing unique to Russian phonetics or phonology about any of these cases. The phonetic account of Degree 2 reduction presents a unified picture of all the cases of the failure of reduction, deriving them all from additional duration assigned in the phonetics.

The following can be said thus far: Degree 1 reduction, the merger of /o/ and /a/ in unstressed syllables is phonological. It is neither duration-dependent nor gradient in its application. Degree 2 reduction, the realization of /a/ as [ə], however, is as yet
unphonologized. The facts cited above concerning the failure of reduction to schwa suggest that it is duration-dependent. There is also no independent phonological evidence that reduction to schwa must take place in the phonology rather than in the phonetics (i.e. there are no other phonological processes which interact with reduction to demonstrate that schwa must be featureless, or even that it must not be specified [+low]. The next section presents experimental results which confirm this picture of Russian UVR.

2.5.2. Experimental investigation of Russian UVR

Not only does reduction to schwa fail to occur in certain positions with characteristically longer vowel durations, it turns out that this reduction applies to a greater or lesser degree in all non-first-pretonic syllables as a function of the duration of the vowel. In virtually all normal speech /a/-/o/ in positions subject to Degree 2 reduction is indeed realized closer to schwa than it would be if it were located in the first pretonic syllable. It reduces far less, though, in slower speech than in faster, and in deliberate, hyperarticulated speech it is virtually identical to the corresponding first pretonic. Thus, while reduction to schwa can be undone through hyperarticulation, as noted above, no
amount of hyperarticulation or emphasis can restore the contrast between /a/ and /o/, the neutralization of which is truly categorical.

The results of a small experimental study I have carried out illustrates the gradient, rate-dependent nature of the reduction to schwa. To test the hypothesis, I recorded one native speaker of Russian (born and raised in Leningrad) reading a series of sentences designed for this purpose. These frame sentences were of the form “The word X translates easily”, (Slovo X perevoditsja legko), such that the pretonic syllables were some distance from both utterance boundaries. Each token replacing “X” in the sentences was a real-word trisyllable of Russian, with either medial or final stress (the pretonic syllables should in all cases escape the effect of final lengthening applied at any prosodic phrase-boundary marked before the verb). The vowels in target first and second pretonic syllables were either /a/ or /o/ underlingly. Syllables were open and had onsets. The segmental context was non-nasal (following nasality has a clear effect of raising F1 in low vowels in Russian, as pilot versions of this experiment demonstrated). Sentences were presented in randomized blocks of ten. The speaker was asked to read each block in

---

28 Unstressed [o] is pronounced only in unassimilated loans in Russian, and even in most loans being dispensed with.

29 For variety’s sake, other forms for the frame sentence, in equal numbers and randomized in the list were “The word X sounds bad”, “The word X sounds good”, and “The word X is easy to memorize” (zapominaetsja legko).
a formal style, described as “like a newscaster might”. After a brief pause, the speaker then repeated the same block of sentences, instructed this time to read at a faster pace. Blocks were kept small enough that the speaker was able to keep a more or less constant speech rate throughout each one. Recordings were digitized at 22.050 KHz and analyzed using the Praat 4.0.2 speech analysis software (Copyright@1992-2001 by Paul Boersma and David Weenink). Measures of duration for target vowels were taken from linked spectrographic and waveform displays. Formant measurements were taken using LPC autocorrelation analysis from the approximate midpoint of the vowel in question where formant values were as close to steady-state as was achieved in each case. Discarding tokens containing second-pretonic syllables from which the vowel had actually been deleted (or effaced beyond capacity for formant measurement), I was left with measurements from 92 first pretonic vowels and 60 second pretonic. In some cases a single token contained both. Results were as follows.

Durations did vary considerably as a function of speech rate. Mean durations over both tempos were 83.5 ms. (s =13.8) for first pretonics and 40.8 ms. (s = 9.7) for second pretonics. Mean durations for first pretonics at the faster rate were 75.3 ms. as compared to 90.4 ms. at the slower rate. For second pretonics mean durations were 34.9 ms. at the faster rate and 44.8 at the slower. Mean F1 for first pretonic vowels was 552.8 Hz (s =

97
37.1). For second pretonic vowels mean F1 was 462 Hz (s = 45.7). Neither durational nor spectral differences between underlying /a/ or /o/ were found to be significant.

The most important result in the present context, however, is this: There was a strong correlation between duration and F1 for second pretonic vowels (correlation coefficient, r, (60) = .602, p < .0001). This can be seen in the scatterplot shown below:

(10) Russian second pretonic /a/, /o/: F1 vs. Duration

---

---

98

---

\[^{30}\] All statistics computed in StatView 5.0.1 (SAS Institute Copyright 1992-98) unless otherwise noted.

\[^{31}\] Scatterplots created with Microsoft Excel 98 (Copyright 1995-1998 Microsoft Corporation).
Equally interesting, however, are the results of the tests for vowels in the first pretonic syllable. In Russian, Degree 1 reduction along the F1 dimension is not a function of duration, in contradistinction to the UVR in Standard Italian unstressed syllables. Here, the correlation results were $r_i(92) = .179, p > .05$. This is illustrated in (11):

(11) Russian first pretonic /a/, /o/: F1 vs. Duration

In second pretonic syllables, then, vowel height is largely a function of phonetically-assigned duration, varying in accord with such factors as speech rate and degree of final lengthening. There is little reason, in light of this, to imagine that the
phonological specification of this vowel is any different from that of any other [a] (← /a/, /o/). The realization of /a/, /o/ in non-first-pretonic syllables is as close to pretonic [a] as durational restrictions allow it to be (identical in many cases under the more extreme degrees of lengthening found in absolute word-initial position, hiatus with another low vowel, and phrase-final position), and likewise exactly as far from [a] as durational restrictions in other cases force it to be.

First pretonic /a, o/ may simply have its own F1 target specification, narrowly enough defined so as not to be influenced by durational pressures. Another interpretation of the results above, though, is simply that the durations of first pretonic vowels never actually got low enough in my experiment to make their influence felt consistently on F1. Under this interpretation, then, all unstressed syllables share a single F1 target. First pretonic syllables can now be seen as just one position among several (absolute initial, phrase-final, hiatus) in which sufficient duration is present that pressure to raise [a] to [ø] is not significant. [a] in syllables in which reduction to schwa does take place share the same F1 target as [a] in the first pretonic syllable. This target window is

---

32 Whether that specification is independent of the specification of stressed /a/, as the literature suggests, or is in some way just reflecting conditioned variation (window resizing in the terms discussed in Chapter 1) of the specification for stressed /a/ is a matter to be determined through further empirical inquiry.
wide enough, however, that variation can occur when durational pressures are great enough.\footnote{The assumption of a single target, as discussed in chapter 1, is consistent with the interpretation that reduction to schwa is not phonological, all unstressed instances of /a/ having the same featural representation in the phonology. It is not, however, a necessary consequence of that representation.}

The scatter plot below in (12) superimposes the results shown above in (10) and (11), providing dramatic confirmation of the view the above interpretation. Viewing the results for Degree 1 and Degree 2 reduction contexts together, we can see that F1 values remain relatively stable, uninfluenced by duration down to approximately 60 ms., at which point durational impoverishment begins to take its toll on the speaker’s ability to realize a sufficiently open vocal tract to produce the F1 value otherwise specified for the vowel in question.
Again, phonologically, the two “degrees” of reduction of /a, o/ in Russian turn out in fact not to be representationally distinct at all. Rather, they are the result of a single set of processes: phonological neutralization of unstressed /a/-/o/ to a single featural specification (whatever we assume that to be), which then receives gradiently differing phonetic implementations as a function of the amount of duration assigned by the phonetics to the syllable in question.

If, on the other hand, the two levels of reduction of unstressed /a/ and /o/ were actually to be considered phonologically distinct categories with different constellations
of features (or rather, in this case, the features of [a] vs. complete underspecification), we would expect two completely independent sets of spectral realizations, one corresponding to an [a], and the other functioning something like schwa in English, truly phonetically targetless with formant values interpolated from the action of coarticulation with surrounding segments superimposed on an otherwise neutral vocal tract (see Browman and Goldstein 1992 for discussion of the need for a target articulation for schwa). In such a situation, the rise of F1 values toward those characteristic of [a]/[e] as permitted by adequate duration is entirely unexpected and inexplicable. That they should actually reach those values when given the opportunity is stranger still.

While there is ample evidence demonstrating the difficulty of articulating a low vowel under a certain durational threshold, which explains a raising trend in shorter syllables, there is no reason whatsoever to expect a mechanical link between lowering of schwa toward [a] and small increases of duration. In this experiment the highest recorded duration for a second pretonic /a, o/ was 61 ms. and yet nonetheless the vowels in the upper durational range have F1 values more characteristic of the range found for first pretonic /a,o/\(^4\). Had the experiment included phrase-final or absolute word-initial /a, o/.

\(^4\) While some languages do avoid stressing inherently shorter vowels such as schwa, fail to lengthen them under certain circumstances otherwise requiring it, and perhaps even enhance their sonority in positions of strength, in such cases the relevant degree of durational enhancement is far greater than the paltry amount

103
in which the spectral shift is so conspicuous that it is noted in impressionistic
descriptions, the asymptoting effect for F1 of non-first-pretonics as duration increased
would have been all the more dramatic.

It is clear from the above, then, that Russian has one phonological reduction
process (Degree 1), and one phonetic reduction process (Degree 2). Furthermore, both of
these apply to all unstressed syllables equally. The merger of /a/ and /o/ takes place
categorically whenever the vowels in question appear outside the stressed syllable.

Gradient raising of unstressed /a/-/o/ toward [ə] occurs in all unstressed syllables as a
function of the duration allotted to the vowels in question by the phonetics. Where this
duration is above a certain threshold (apparently somewhere just over 60 ms), no
reduction is seen.

The only catch here is that the phonetics must be able to assign different durations
to different unstressed syllables on a language-specific basis, since while absolute initial
vowels, phrase-final vowels, and vowels in hiatus before identical vowels receive
additional duration for reasons not specific to the phonology of Russian, significant

| in question here. I know of no reason a true schwa should respond so dramatically to so marginal an increase of duration as that found in this experiment. | 104 |

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
lengthening of vowels in first pretonic syllables is a property of certain Slavic languages in particular.  

Language-specific apportionment of duration by the phonetics is nothing controversial, however. In Russian, as in other languages, we already need to have a phonetics capable of assigning radically different durations (at least for speakers such as this one, and those described by Matusevich above) to equally moraic vowels stressed and unstressed (first pretonic). We also need it to assign different durations to monomoraic phrase-final vowels than it does to monomoraic phrase-internal word-final vowels. We need it to alter durations of monomoraic vowels according to the complexity or presence of onsets and codas in their syllables, to lengthen and shorten monomoraic vowels according to the voicing specifications (and often other sonority-scale differences as well - nasal, liquid, glide) of following consonants, according to the aspiration of preceding consonants, to the number of syllables in the word, to the position of the word in the utterance, to the position within the word of the monomoraic vowel regardless of placement of stress, etc., etc. It is difficult in light of this to see what should prevent the phonetics from assigning different durations to phonologically identical vowels as a function of their position relative to the stressed syllable, as is the case in East Slavic.

---

35 A trace of an earlier prosodic system with bimoraic pitch accents

105
2.6. Phonetic Determinism in UVR systems

I have argued in this chapter so far that the typological characteristics of systems of unstressed vowel reduction are best accounted for using a phonologization-based approach. This approach sees regularities in phonological patterns of UVR as a result of the fact that most UVR systems arise from the phonologization of phonetic patterns of vowel realization approximation and overlap cause by durational pressures on the articulation of vowels in unstressed syllables. I have also illustrated in case studies instances of both the phonetic and phonological ends of the phonologization process, including cases in which the two are operative in a single language (Brazilian Portuguese, Russian) or at opposite ends of a dialect continuum (Bulgarian). The patterns start off as functions of the apportionment of duration in unstressed syllables, but once phonologized, the neutralizations take place regardless of the phonetic characteristics of the syllables in which they occur. In this sense, then, phonologized patterns of UVR are to a large extent synchronically arbitrary. This phonetic arbitrariness is underscored by the existence of patterns of phonological UVR not resulting from phonetically-conditioned mergers in unstressed syllables to begin with.

Section 2.3.2 above raised the possibility that the East Slavic neutralization of /a/ and /o/ in unstressed syllables is actually the result of a failure of earlier short and long /a/
to split into qualitatively distinct vowels in those positions. If this is ultimately born out, a story of articulatory approximation followed by misperception and phonologization of the merger is obviously entirely incorrect. A clear case of such a set of developments is found in some northern Russian dialects where etymological /o/ under specific accentual circumstances (See Timberlake 1983a. and 1993 for details) splits into /o/ and what is traditionally symbolized /ø/ in the Slavic literature (and in which the contrast between /a/ and /o/ is realized in unstressed syllables). This, in combination with a subsequent merger of */u/ with original /o/, produces a situation in which /ø/ (higher and often diphthongized) contrasts with /o/ only under stress. This is so not because unstressed syllables were not durationally strong enough to license the contrast (in these dialects there is apparently little reduction of unstressed vowels to begin with - hence the persistence of the /a/ - /o/ contrast), but because /ø/ arose only in syllables which bore particular types of accent.

The resulting licensing asymmetry is clearly not “phonetically determined” in a synchronic sense, though it is phonetically explicable as an instance of phonologization. Indeed, Timberlake (1993) suggests a phonetic explanation for the split of original */ø/ in North Russian dialects involving the pitch curves and attendant durational patterns produced by the earlier pitch accent system, rephonologized upon the breakdown of that system in the relevant dialects. Approaching the typology of UVR systems from a...
phonetically-motivated phonologization model, we can account equally well for the existence of these synchronically similar licensing asymmetries by simply identifying more than one phonetic source for a given synchronic pattern (both durational reduction of unstressed vowels and durational and articulatory enhancement of accented vowels\textsuperscript{36} can result in differing sets of contrasts available in stressed and unstressed syllables).

Indeed, in many cases synchronic reduction patterns will not be phonetically motivated even in diachrony, but rather, will result from the layering of several not-necessarily-related sound changes, each phonetically motivated in its own right, but in aggregate creating a pattern which is not wholly phonetically explicable. Such patterns will never be found in gradient versions such as those discussed above for phonetically-motivated UVR patterns, since they are not in fact the result of the phonologization of a phonetically-conditioned gradient shift from one articulation to another. Such a case is found in Seediq, an Austronesian language of Taiwan, belonging to the Atayalic family.

2.6.1. Vowel Reduction in Seediq

Holmer (1996) presents a description of the Paran dialect of Seediq (also called Sedik in the literature - Asai 1953, Li 1977, Li 1980). Seediq is described as having five

\textsuperscript{36} See de Jong (1995) and Cho (2001: 77-80) for discussion of the nature of this enhancement under accent.
underlying vowels /i, e, a, o, u/, all of which are contrastive in stressed syllables. In
pretonic syllables, however, only one vowel is realized: /u/37. In unstressed syllables /u/
has the allophonic realizations [u], [u], and [a], according to Li (1977). Li analyzes the
pretonic /u/ as in fact epenthetic. Unstressed vowels reduced to /u/ in this position are
frequently deleted, their realization largely predictable from sonority restrictions on
consonant clusters. Li therefore analyzes forms liked ‘mupubulebin’, ‘will pull’, as
having underlying clusters before the tonic syllable: /mpblebin/. Be this as it may, the
reduced vowels of the posttonic syllable cannot be analyzed as such. In these, the
following neutralizations take place.

(13) Posttonic vowel realizations in Seediq

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Realization</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i/</td>
<td>[i]</td>
</tr>
<tr>
<td>/a/</td>
<td>[a]/[ə]</td>
</tr>
<tr>
<td>/e, o, u/</td>
<td>[u]</td>
</tr>
</tbody>
</table>

Alternations caused by these neutralizations can be observed under affixation, where
stress shifts onto the former posttonic final syllable, as shown in (14), where for clarity’s
sake Holmer’s intermediate stages are included as a derivation.

37 Stress in Seediq is fixed penultimate.
38 /a/ and /i/ are said by to retain their stressed qualities “to a certain extent” (Holmer 1996: 26). I assume
this means that they are somewhat centralized.
(14) Vowel reduction in Seediq

/hanəd/ 'cook'
- > h-m-anəd -> hmanəts -> [humənuts], (Actor Focus)
infixation d -> ts epenthesis and UVR

- > hanəd-an -> hnedan -> [hunədan], (Locative Focus)
suffixation a -> ø epenthesis

/heyeg/ 'stand'
- > m-heyeg -> mheyew -> mheyuw39 -> mheyu -> [meheyu] (AF)
prefix g->w UVR w -> ø epenthesis, translaryngeal harmony

- > h-n-eyeg -> hneyeg-an -> hneyegan -> [hunuyégan] (LF, Pret.)
infix suffix e -> ø epenthesis

There are lexical exceptions to the pattern, but it is otherwise said to be quite general. The problem, of course, is that in an unstressed syllable which licenses the faithful realization of [i], it is difficult to imagine what would make [u] an optimal realization for unstressed /e/. /e/ -> [i] involves the change of [-high] to [+high], involved in any case in /e/ -> [u] as well. But /e/ -> [u] involves changes to [back] and [round] as well (or just [back], depending on one's view of underspecification).

39 Final -Vw simplified to -V is independently attested and is not the source of [-u] here. Other dialects retain /w/.
The answer is that [u] is not in fact an optimal realization for /e/ under vowel reduction under any circumstances (which is to say, in comparison with [i] or [æ]. [u] is perfectly fine from the point of view of dispersion). It is, however, the realization that a layering of sound changes in Seediq ultimately provides. There was no gradual drift of unstressed /e/ toward [u], finally resulting in reanalysis (an implausible tendency at best). The authors describing these facts offer no explanation of the rise of this pattern, and not all the facts are clear to me still concerning precisely what did happen, but the crucial pieces of the puzzle appear to be these: Stressed /e/ in Seediq comes from *Proto-Austronesian /a/ (Asai 1953: 13). Asai likewise attests that when deprived of stress, /e/ is realized as [i, ə, u]. He also claims that /i/ has the same realizations in unstressed syllables, but gives examples of this only in pretonic syllables. By contrast, there are plenty of instances of posttonic [i] throughout the grammar (e.g. [mari] ‘buy’, [muheə] ‘bearing fruits’), but apparently none of [e]. It is plausible to assume, in fact, given available information, that [e] never arose outside stressed syllables in the first place in Seediq. In any case, the second crucial piece of information is that in the Paran dialect at the time of Asai’s writing (1953), *PAN /u/, realized as such even in other Atayalic dialects, is being realized as [u] under stress in Seediq. About this transcription, too, Asai is quite specific, describing it that a high, back, non-round vowel. At least some
instances of */u/ > Seediq rounded [u] are attested in Asai, particularly in final syllables: [kücü̯h] < *kuku 'claw, and others. It is not clear to me whether this holds of all and/or only final syllables, as examples in Asai are few. What is clear, however, is that earlier /u/ is being realized regularly as [u] in stressed syllables in Asai’s description, but in Li’s and Holmer’s later work, also based on direct work with speakers, all these stressed reflexes of */u/ are transcribed as [u], with no indication that they are pronounced any other way (Though unstressed /u/ is said to vary from [u] to [u] to [u], note that all three variants imply rounding).

Asai reports far more variation in the realization of unstressed vowels than do the later researchers, perhaps indicating relatively recent transformation of the processes from gradient to categorical, which they now seem to be. Whether or not this is the case, what must have happened in Seediq is as follows: Stressed /e/ alternates with unstressed [i, ə, u], which must be contextual variants of one another. Unstressed [u] must have been realized in this same range, such that the two vowels become phonologically neutralized in unstressed syllables. Then at some point before the stage described by Li and Holmer, the pronunciation of [u] became rounded across the board, including the

---

It is tempting to see the final syllable [u] realization as crucial to the UVR pattern, but without a better understanding of the distribution of this reflex this might be premature.
unstressed allophones of /e/. This is reflected in the current situation, where no mention
of a high back unrounded vowel is made in either account, and /e/ alternates with /u/ in
unstressed syllable. The unstressed inventory created by these processes is, it must be
noted, perfectly well dispersed (it is even tempting to attribute the rounding of [ɯ] to the
enhancement of the contrast between high vowels, increasing the goodness of the
dispersion). The problem is the relationship between /e/ and /u/, curious from both
phonetic and phonological perspectives. While Seediq UVR is typologically odd, the fact
that the phonologization approach to typological regularities locates the source of their
phonetic naturalness in diachronic development means that the phonological grammar is
in no way prevented from implementing a pattern such a pattern, should it somehow
come to arise.

2.7. Summary

I have argued in this chapter there is no reason to believe that categorical patterns
of vowel reduction have any more than a single formal implementation synchronically.
The formal system, be it based on Licensing constraints such as those employed by
Crosswhite for some UVR processes, or some other variant of Positional Faithfulness or
Positional Markedness should be chosen according to its ability to implement the attested
range of phonological patterns, rather than according to its ability to justify the existence

113
of that pattern from a functionalist perspective. If it is true that we must accept more than one distinct formal mechanism for the synchronic implementation of UVR patterns, it should be possible to demonstrate that those two patterns actually behave differently from a synchronic point of view (rather than merely being attributable in a non-specific way to two different sets of general functional principles). I have identified two (not unexpected) patterns of this sort in the crosslinguistic patterning of UVR systems: those which are categorical or phonological in nature, and those which are gradient or phonetic. The former are the consequence of a true neutralization, a merger of phonological entities, while are the latter are cases of phonetic approximation or overlap of realizations conditioned by durational pressure on articulation. I have also shown that far more UVR patterns than were previously thought in fact belong to that gradient, phonetic type, including that of Standard Bulgarian and Degree 2 of reduction in Brazilian Portuguese and Russian. Furthermore, I have presented the phonologization model as the link between the two categories, the force which transforms the latter into the former. That the categorical patterns should frequently be explicable with reference to phonetic tendencies or perceptual limitations is because they are more often than not the direct result in a diachronic sense of those actual tendencies. Some patterns, however, such as Seediq.

---

41 Principles which may indeed be active in the phonetics.
clearly have their origins elsewhere, but it is in no way obvious that this unnatural conception ultimately causes them to behave synchronically any differently than patterns with clean phonetic pedigrees. Unless it can be shown otherwise, therefore, there is no reason to believe that their synchronic phonological implementation should be in any way formally distinct from those others either. The phonologization approach to typological regularity, based as it is on the diachronic development of real phonetic patterns, makes predictions about which types of UVR systems should be attested and which types which should not, and those predictions are largely correct. That UVR patterns should center overwhelmingly on neutralizations of vowel height contrasts, with few if any instances of systems based on the neutralization of front/back, round or ATR contrasts in unstressed syllables is predicted in the phonologization model, which identifies as the primary source of UVR patterns durational pressures on the articulatory implementation of height contrasts. No such pressure, on the other hand, is known to exist on front/back, rounding, or ATR contrasts, meaning that UVR involving these should be less likely to arise. That a contrast of vowel nasalization should be neutralized outside stressed syllables, as it is in Copala Trique, for example, is also predictable, since it is well-known that in many languages (as noted above for Brazilian Portuguese, where maintaining the contrast nonetheless seems more important to the categorical phonology than reducing the durations of all unstressed syllables) nasalized vowels are significantly longer than their
oral counterparts. Other theories of vowel reduction, such as that advanced in Crosswhite (2001) fail to predict these typological regularities, in fact even predicting the reverse in the case of contrast-enhancing UVR. The phonologization model, by contrast, both accounts for typology of UVR systems, and in doing so creates the possibility for a much simpler treatment of UVR in the phonology, since the latter is no longer compelled to duplicate the work of phonologization in generating the phonetic naturalness of typological patterns.
Chapter 3. Final Syllables

This chapter looks at the range of positional neutralization effects found involving the final syllables of a prosodic domain. Final syllables present an interesting test case for theories of positional neutralization for a number of reasons. First and foremost, they commonly share with stressed syllables the phonetic cue taken by phonologists such as Steriade to be among the most crucial in licensing vowel contrasts: duration. The phonetic durations of domain-final syllables are known to be enhanced in many languages by a process known as final lengthening. For this reason, we might legitimately expect to find the same types of licensing asymmetries in final syllables as we found in stressed syllables (Chapter 2).

Secondly, final syllables are also known to bear a particular psycholinguistic significance during acquisition, a fact which we might expect to make the existence of licensing asymmetries here all the more likely (Kehoe and Stoel-Gammon (1997). Indeed, final syllables have been claimed to be strong licensors for vowel quality contrasts, but not unambiguously so.

Whatever phonological strength effects may be found in final syllables, the right edge of word and phrase alike are also famously host to segmental weakening effects such as devoicing, reduction and deletion. This latter class of effects, which has no counterpart in stressed syllables, introduces a contradiction into the status of final
position as a licenser of contrasts, and demands an explanation from any theory seeking
to account for the typology of positional neutralization effects found in this position. In
this chapter I offer an explanation of these contradictory phonological patterns which
attributes their emergence to the effects of a diverse constellation of phonetic properties
associated with domain-final position. These effects, while perfectly compatible with one
another from a production standpoint, have consequences for the perception of contrasts
which frequently prove mutually contradictory. Whether the phonetic patterns tending to
enhance perceptual robustness or those tending to diminish it are ultimately reinterpreted
as intentional and phonologized is determined by the relative magnitudes of the language-
specific instantiations of each phonetic tendency.

While the phonological effects of this uneasy coexistence in final position among
perceptually antagonistic phonetic trends are comprehensible according to the
phonologization approach to typological patterning, final syllables present a problem for
Universalist Pure Prominence approaches to PN. The equally robust attestation of final
syllables as strong and weak licensers of vowel contrasts precludes the specification in
UG of final position as either inherently Strong or inherently Weak. While the licensing
capabilities of final syllables in a given language will be motivated from a diachronic-
phonetic standpoint, synchronically phonological Strength or Weakness for this position
appears to be a parameter which must be set arbitrarily on a language-specific basis.
Insofar, too, as the strong or weak licensing capacity of final position must be divined from empirical evidence during the acquisition process, it is reasonable to question the extent to which it is necessary or realistic to assume restriction in UG of the status vis-a-vis potential neutralization patterns of any other specific position either.

Section 3.1 reviews arguments for the phonetic and psycholinguistic prominence of final position, and section 3.2 catalogues and analyzes the phonological strength effects which result from this prominence. Section 3.3 and 3.4 compare the strength effects found in final position to those discussed in Chapter 2 for stressed syllables. The problem of reconciling these effects with tendencies to phonetic and phonological weakness in final syllables is taken up in 3.6, and Section 3.7 is devoted to the licensing of vowel quantity in final position and its comparison with similar effects in stressed syllables.

3.1. Final syllable prominence

3.1.1 Phonetic prominence

Domain-final syllables have been shown by many researchers to be the locus of a certain degree of vowel lengthening (Oller 1973, Klatt 1975, 1976, Beckman and Edwards 1987, Wightman et al. 1992, Keating, Wright and Zhang 1999, inter alia on English). Lengthening effects of any kind are of immediate relevance to the present study.
insofar as we have seen how positional durational asymmetries can give rise to neutralization patterns in other positions, such as stressed and unstressed syllables. Languages in which final lengthening has been identified experimentally include American Sign Language (Brentari 1995), Beijing Chinese (Zhang 2001), Brazilian Portuguese (Major 1985, Simões 1991), Bulgarian (Savitska and Bojadzhiev 1988), Creek (Johnson and Martin 2001), Czech (Dankovičová 1997), Dutch (Cambier-Langeveld 1997, 1999), English (see above), French (Delattre 1966, Fletcher 1991), German (Kohler 1983), Hausa (Newman and Van Heuven 1981), Israeli Hebrew (Berkovits 1984, 1993a, 1993b, 1994.), Italian (Farnetani and Kori 1990), Japanese (Ueyama 1999), Jicarilla Apache (Tuttle 2000), Jordanian Arabic (de Jong and Zawaydeh 1999), Korean (Cho 2000), Luganda (Zhang 2001), Russian (Zlatoustova 1981: 13-17), Spanish (Delattre 1966), and Swedish (Nord 1974, 1986). It has been described or implied in numerous descriptive grammars as well, as shall be seen below. It is difficult to assess the universality of the phenomenon due to the tendency for experimental studies to focus on European languages (and primarily English). Most grammar writers, to compound the problem, for fairly obvious reasons make no reference, positive or negative, to the phenomenon positive. Nonetheless, I know of no study experimental or impressionistic making the explicit claim that a language lacks any form of final lengthening. The existence of phrase-final lengthening in such pre- or non-linguistic
behaviors as infant babbling, musical performance, and bird song and insect chirps lead
Johnson and Martin (2001) (among others) to speculate that final lengthening is in fact a
property of motor performance in general.

However wide-spread, it is clear that domain-final lengthening is nonetheless
subject to significant language-specific variation. Johnson and Martin (2001) cite
Delattre’s 1966 study of syllable durations in four languages, in which French, English,
Spanish and German are shown to differ in the magnitude of final lengthening observed
in comparable domains, and in Johnson and Martin’s study, Creek is seen to differ from
all four of these in degree of lengthening. To this we may add the comparative study of
accentual and final lengthening in English and Dutch of Cambier-Langeveld (1999),
which shows that while English words with phrasal accent in phrase-final position are
longer than unaccented phrase-final words, in Dutch items in phrase-final position
undergo the same amount of lengthening whether accented or not. Unfortunately, most
studies of domain-final lengthening to date have focused only on a single language, such
that there is a pressing need for additional comparative studies to determine the extent of
crosslinguistic variation in the implementation of this possibly otherwise universal
phenomenon.

A number of studies of final lengthening have been devoted to determining the
extent of its operation in prosodic domains of different levels. These are especially
relevant to the present study for their help in understanding how the phonologization of final syllable effects proceed from the level of the phrase to the level of the word. It is difficult in some respects to compare the results of these studies directly. For an experiment to show conclusively how two languages differ in their implementation of final lengthening at a given prosodic level, extreme care would have to be taken not only to replicate general aspects of experimental design and presentation, but in particular to make certain that the domains correspond to comparable levels of the Prosodic Hierarchy in both languages, that target phrases and words were of the same length, that syllabic and segmental environment in the vicinity of the boundaries were identical, and so forth. Studies of this type, preferably of languages otherwise with dissimilar prosodic systems, are sorely needed for further progress in understanding the extent of possible variation in the implementation of final lengthening. In addition, while many studies or descriptions report only utterance-final lengthening, it is often not clear whether other phrasal domains were tested or observed at all. Be this as it may, we can draw a number of important generalizations from the studies on this question: A number of studies (e.g. Beckman and Edwards 1990, Cho 2001, de Jong and Zawaydeh 1999, Ueyama 1999) have shown the magnitude of pre-boundary lengthening effects to be greater, the higher the level of the prosodic domain in which the relevant syllables are final. Thus, de Jong and Zawaydeh find for Jordanian Arabic the least lengthening before boundaries of non-
utterance final phrases with no F0 contour marking the end of the phrase, more lengthening in non-utterance final phrases with boundaries marked by a clear F0 contour, and the most lengthening of all utterance-finally. Ueyama finds more pre-boundary lengthening in Japanese Intonational Phrases than in lower-level Accentual Phrases. While Beckman and Edwards 1990 report a certain degree of (phrase-internal) word-final lengthening in English, this lengthening is clearly less robust both in degree and consistency of application than is the pre-boundary lengthening they detect at the ends of higher-level phrases. The implicational statement following from this observation seems valid for all languages in which final lengthening has been measured or described: the occurrence of lengthening before a lower-level domain boundary implies the occurrence of lengthening at all higher levels. As will be shown below, this pattern is mirrored in the crosslinguistic distribution of positional neutralization effects involving final position as well: if a domain-final effect applies before a phrase-internal word-boundary, it will apply when that word boundary coincides with higher-level phrase boundaries as well. The reverse is not the case.

Additional phonetic characteristics of final lengthening, in particular issues concerning its articulatory implementation and domain of application, will be discussed later on in this chapter in connection with the phonological patterns to which they are relevant. For now it is sufficient to characterize domain-final lengthening as a phonetic
effect which is potentially universal in its cross-linguistic distribution, subject however to
language-specific variation in its implementation, and most robust in higher-level
prosodic constituents. To the extent that final lengthening is universal, we might expect
final strength effects to be equally widespread.

3.1.2. Psycholinguistic prominence

In addition to their well-documented phonetic enhancement, final syllables are
also known to exhibit special salience in a psycholinguistic sense. Kehoe and Stoel-
Gammon’s 1997 study of prosodic acquisition demonstrates this well. In this study of the
truncation patterns of children acquiring English, the authors show that in truncated
productions of polysyllabic words, children are most likely to retain internal stressed
syllables and final syllables, stressed and unstressed. They characterize their results as
follows (Kehoe and Stoel-Gammon 1997: 120-121):

In WS words, children produced the stressed syllable; in S'S, 'SWS, and SW'S words, children
produced both stressed syllables; in WSW, SWW, and WSWW words, children produced the
stressed syllable and the unstressed syllable in word-final position; and in 'SWSW, SW'SW, and
S'SWW words, children preserved both stressed syllables and the unstressed syllable in word-final
position. Children preserved a medial rather than a final unstressed syllable in their truncations
only infrequently.

Where 'S refers to a syllable with primary stress, S to a syllable with secondary stress, and W to an
unstressed syllable.
Kehoe and Stoel-Gammon suggest that the increased perceptual salience of final syllables due to final lengthening is a possible source for this pattern, but that perceptual difficulty with non-prominent material cannot ultimately tell the full story. Firstly, it has been shown that children do in fact accurately perceive the material they otherwise omit from truncated productions. Segmental material from deleted syllables (e.g. their onset C’s) indeed often appears in truncated forms, reinforcing this point. Furthermore, in children’s earliest truncations, it is the rightmost stressed syllable in a word which is most often the one to surface. When the adult form has final stress, the final syllable along surfaces, and when stress is penultimate, children’s most frequent early truncations preserve both the stressed and the final unstressed syllables (p. 123). All this suggests that there is some additional importance attached to the right-edge adjacent material in the word which goes beyond simple relative ease of perception of the segmental material located there. Kehoe and Stoel-Gammon argue that children’s phonological representations are relatively intact (compared to those of adults), but that production constraints give rise to the omission of all but a few privileged portions of the word. This enhanced psycholinguistic salience accorded to prosodic heads and edge-adjacent material can be modeled using the parametrized faithfulness constraints now familiar from a variety of OT approaches to positional neutralization. Kehoe and Stoel-Gammon are not specific, however, as to the precise nature of the psycholinguistic salience of word-final syllables. Still, the
preferential retention of final syllables in children's truncations does suggest the possibility that the privileged psycholinguistic status of these syllables could lead to analogous phonological effects in adult grammars.

3.2. Phonological strength effects in final position

The foregoing makes the prediction that phonological strength effects should be attested in final position, and this is indeed the case. The following sections provide an account of those effects. I begin with a brief discussion of previous work on tone in final position. I then move on to segmental licensing asymmetries, beginning with two case studies illustrating the complexity of the issue from the standpoint of the relative roles of phonetics and phonology in determining the shape of PN patterns. I then present a catalogue of attested final strength patterns, from which I extract a typological generalization concerning the role of syllable structure in determining the distribution of final strength effects. This generalization allows us to clearly disambiguate the relative roles of phonetic and psycholinguistic prominence in producing final-position PN patterns.
3.2.1. Tone in final position

One type of phonological strength effect that has been thoroughly documented in final position is the tendency toward preferential licensing of contour tones on final syllables. Zhang (2001), building on Gordon’s (1999) crosslinguistic study of the phonetic properties associated with tone-bearing segments, provides evidence for the strength of prosodic-final syllables with respect to contour tones. Zhang cites the behavior of contour tones in Kikuyu, shown in (15), as an example of this pattern.

(15) Contour tones in Kikuyu: phrase-final vs. non-final

a. kāriōkī moēyā -> kāriōkī moēyā ‘good Kiriuki’
   \L H L \L H L

b. kāriōkī -> kāriōkī ‘Kiriuki’
   \L H \L H

In the non-prosodic-final context in (a), the floating high tone alone occupies the final syllable of the noun, while the low tone is realized only on the first two syllables. In the prosodic-final context in (b), by contrast, the high associates to the final syllable, forming a contour with the preceding low. Zhang explains this pattern as a consequence of the additional phonetic duration associated with domain-final syllables. Contour tones, for obvious reasons requiring more duration for their realization than other tones, are shown
to be restricted in many languages to syllables characterized by a longer sonorous-rhyme duration. Among these are syllables with long vowels, syllables with short vowels closed by sonorants, stressed syllables, domain-final syllables, and those located in shorter words.

3.2.2. Vocalic strength effects in final position

Zhang cites Steriade (1994) as claiming that final syllables can sometimes behave as strong positions for other contrasts as well. Steriade lists four such cases, in which either [round] and/or [back] (Hausa, Timugon Murut, and Eastern Cheremis) or [lax] are preferentially licensed in final syllables. Zhang takes this to mean that final syllables may be strong positions specifically for vowel contrasts which are difficult to perceive. Such contrasts profit from realization in positions associated with additional duration, since, in Steriade's words, "extra duration means extra exposure to a dubious vowel quality and thus a better chance to identify it correctly" (Steriade 1994: 20). Of Steriade's examples of languages with final-syllable strength, the most straightforward is Hausa, which is exemplified in the following section.

Zhang argues on the basis of this evidence for a "direct" approach to the licensing of tone, his formalization of Steriade's Licensing-by-Cue theory. Simply put, what all the positions preferentially licensing contour tones have in common is additional phonetic duration of the sonorous portions of their rhymes. If the phonology can refer directly to this phonetic duration, it can characterize contour-tone licensing in a unified and explanatory fashion that would be impossible with traditional phonological representations (i.e. through reference to moraicity).
3.2.2.1. Hausa

Hausa has a vowel inventory consisting of five vowels, long and short, /i, i:, e, e:, a, a:, o, o:, u, u:/ This complete inventory, however, is reliably distinguished only in final position. While long vowels are faithfully realized in all positions, the situation is markedly different for the short vowels. Steriade claims that the feature combination [round]-[back] among short high vowels is distinctive in Hausa only phrase-finally (Steriade 1994: 13). As noted above, Hausa has been shown to have robust (though non-neutralizing) phrase-final lengthening. In phrase-final position, long and short vowels are realized with identical quality. Short vowels, however, are “cut off by a glottal closure”, while a long vowel “has longer duration and just dies away” (Camochan 1988). Camochan (1951) notes additionally that the long vowels are often realized “with breathy release”, which I will interpret as partial devoicing.

Non-phrase-final short /i/ and /u/ are clearly subject to a great deal of variation in their realizations, determined by both consonantal and vocalic environment (Schuh and Yalwa 1999). Whether or not this variation is ultimately neutralizing as per Steriade 1994

---

44 Zhang adds to Steriade’s list final syllables acting as strong positions for nasalization and fine degrees of vowel height, but cites no such examples. The claim is ultimately correct, as will be seen below.

45 I leave aside here what Newman and Van Heuven show to be a morphologically-restricted third, intermediate class of vowels: those which are long non-prepausally, but prepausally are realized with the glottal closure characteristic of short vowels, and a duration in between that of ordinary long and short prepausal vowels.
is another matter. The possibly conservative description of Carnochan (1988) makes no mention of such a neutralization, transcribing these vowels as somewhat centralized [i] and [u]. Steriade cites Schuh (1971) for the neutralization data. Schuh and Yalwa (1999), however, say the following (p. 90):

Short /i/ may thus range across [i - i - i], and short /u/ may range across [u - u - u]. In normal conversational speech, medial short high vowels are probably frequently neutralized to a high, centralized vowel, with rounding or lack of rounding determined by environment.

Newman (2000: 398-399), noting the variation and in some cases overlap in the realizations of these vowels, nonetheless argues that they are still phonologically distinct, if perhaps on the way to an ultimate merger. There is a serious difference, in other words, between segments whose realizations are frequently overlapping to some extent and those which are actually phonologically neutralized. What we seem to be dealing with in Hausa here is a gradient process akin, for example, to the gradient neutralization of back mid and high vowels in the unstressed syllables of Standard Bulgarian discussed in the previous chapter. The merger would thus be ill-characterized as the neutralization of the categorical phonological features [back] and [round].

Hausa short mid vowels behave differently. While these are distinguished consistently in word-final position, there are no underlying short mid-vowels in medial position in non-ideophonic native vocabulary. Medial short mid-vowels could in principle be created through the process of closed syllable shortening effective in Hausa. This outcome is in fact described by Carnochan 1988, who gives derived closed-syllable forms with the lowered and centralized short mid vowels shown in (16):
(16) Hausa medial short mid vowels à la Carnochan 1988

<table>
<thead>
<tr>
<th>Word</th>
<th>Pronunciation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>kare</td>
<td>[kɛɾeː]</td>
<td>‘dog’</td>
</tr>
<tr>
<td>mace</td>
<td>[mɑʃeː]</td>
<td>‘woman’</td>
</tr>
<tr>
<td>goro</td>
<td>[goɾoʔ]</td>
<td>‘kolanut’</td>
</tr>
</tbody>
</table>

- karen Audu  | [kɛɾeŋʔauduʔ] | ‘Audu’s dog’     |
- macen Audu  | [mɑʃeŋʔauduʔ] | ‘Audu’s woman’   |
- goron Audu  | [goɾoŋʔauduʔ] | ‘Audu’s kolanut’ |

Both Schuh and Yalwa 1988 and Newman 2000, however, are in agreement that at least for most speakers in closed syllables the contrast between short /e/, /o/ and /a/ is neutralized. The resulting vowel is transcribed [ə], but in fact varies greatly in realization depending on its environment. This is shown in (17), with forms taken from both sources:

(17) Hausa medial short vowels

<table>
<thead>
<tr>
<th>Word</th>
<th>Pronunciation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>zoːbeː</td>
<td>[zoːbeː]</td>
<td>‘ring’</td>
</tr>
<tr>
<td>reːʃeː</td>
<td>[reːʃeː]</td>
<td>‘branch’</td>
</tr>
<tr>
<td>toːmaː</td>
<td>[toːmaː]</td>
<td>‘dig up’</td>
</tr>
<tr>
<td>saːboː</td>
<td>[saːboː]</td>
<td>‘new’</td>
</tr>
</tbody>
</table>

- zabbaː      | [zabbaː]     | ‘rings’          |
- rəsaːː      | [rəsaːː]     | ‘branches’       |
- tantɔːnaː   | [tantɔːnaː]  | ‘dig up, pluract.’ |
- saːbɔn sɔxʒaː | [saːbɔn sɔxʒaː] | ‘new soldier’ |

The merger of the non-high medial short vowels as schwa thus seems to be a better candidate for description as a categorical neutralization process\(^{46}\). Both merger patterns

---

\(^{46}\) Newman notes additional details: recent foreign loans with short /e/ and /o/ in closed syllables (such as [fɛnʃl] ‘pencil’ and [bɒs] ‘boss’) constitute lexical exceptions to this process, tending not to merge them with /a/. Also while the word-internal closed-syllable merger is categorical, often even reflected in orthography, non-prepausal word-final short /e/ and /o/ are said only in fast speech to have a tendency to centralize. This suggests that the word-final version of the merger is actually gradient, another instance
clearly have as their ultimate source the durational asymmetry between short vowels in phrase-final and non-phrase-final position. As with the mixed vowel reduction systems discussed in Chapter 2, however, in the gradient cases (medial short high vowels and potentially phrase-medial word-final short non-high vowels as well), the likelihood or degree of phonetic overlap will be determined by the interaction of the duration of the vowel with the coarticulatory influence of surrounding segments, while in the categorical case merger will occur or fail to occur based solely on the position of the vowel within the word or phrase, actual durational characteristics of any particular instantiation of the vowel in that position notwithstanding.

In Hausa, then, the domain-final syllable is actually the sole position licensing the full array of contrasts available in the language. All other positions allow the realization of only a subset of those contrasts, making this essentially the final-syllable analogue of stress-based vowel reduction. Yet, interestingly, such cases are far from the norm. In fact, uncontroversial instances in which this is the case irrespective of the placement of stress, and in which the phonetic source of the pattern is clearly the additional duration supplied by final lengthening are few and far between. Steriade’s inclusion of Pasiego Spanish in her list of such systems is illustrative in this connection.

such as those of Uyghur and Nawuri discussed below in which a word-internal categorical change operates gradiently in word-final position.
3.2.2.2. Pasiego Spanish

Pasiego Spanish (Penny 1969, McCarthy 1984, Flemming 1993, Sanders 1994, Dyck 1995, *inter alia*), as Steriade notes, does indeed allow a contrast between tense and lax vowels only in final syllables. More specifically, [tense] and [lax] are contrastive only for high back vowels in absolute word-final position\(^47\). Whenever the lax -\(U\) suffix appears, however, all preceding vowels in the word are realized as lax as well, a process usually analyzed as a right-to-left laxing harmony\(^48\).

Penny (1969: 148-149) gives the vowel inventory below in (18) for the tonic and non-final atonic syllables of Pasiego Spanish. He describes the lax variants of the vowels in the following way: The high front -\(I\) - is centralized and lowered, as is -\(U\). -\(A\) - is “fronted and slightly raised, almost [\(e\)]”, and -\(O\) - is similar to the front rounded vowels of French, but “without any great protrusion of the lips”. In final unstressed syllables, only the [\(e, a, u, U\)] appear. Examples of the spreading of the lax feature are given in [19].

\(^{47}\)Still more specifically, the only underlyingly lax vowel in the language is in the lax /-\(U\)/ suffix of masculine singular count nouns. The suffix for masculine mass nouns is a tense /-\(U\)/, raised from an earlier /\(o\)/ to contrast with the now-lax original /\(U\)/ of the count-noun suffix. I follow here the transcription of lax vowels with capital letters used in McCarthy (1984). Penny (1969) marks lax vowels with an acute accent (e.g. -\(u\) vs. -\(u\)), disturbingly reminiscent of tone or stress. Questions concerning the actual phonetic realizations of the lax vowels would make use of IPA symbols somewhat arbitrary.

133

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
(18) Pasiego Spanish vowel inventory

\[
\begin{array}{c|c|c}
    & a & b \\
    \hline
    \text{i} & \text{I} & \text{I} \\
    \text{u} & \text{U} & \text{U} \\
    \text{e} & \text{O} & \text{O} \\
    \text{o} & \text{A} & \text{A} \\
\end{array}
\]


\textbf{a. plurals} \hspace{1cm} \textbf{masc. sg. count}
\textit{tense} \hspace{1cm} \textit{lax}

\begin{align*}
\text{abi} \text{\textacute{\textalpha}nus} & \quad \text{'hazels'} & \text{AvI\textacute{\textalpha}nU} & \quad \text{'hazel'} \\
\text{sold\textacute{\textalpha}s} & \quad \text{'soldiers'} & \text{sOld\textacute{\textalpha}s} & \quad \text{'soldier'} \\
\text{pu\textacute{\textalpha}kus} & \quad \text{'young chickens'} & \text{pU\textacute{\textalpha}kU} & \quad \text{'young chicken'} \\
\text{kant\textacute{\textalpha}rus} & \quad \text{'5 gal. jugs'} & \text{kAnt\textacute{\textalpha}rU} & \quad \text{'5 gal. jug'} \\
\end{align*}

\textbf{b. sg. mass} \hspace{1cm} \textbf{sg. count}
\begin{align*}
\text{m\textacute{\textalpha}lu} & \quad \text{'evil'} & \text{m\textacute{\textalpha}lU} \\
\text{li}\text{mpju} & \quad \text{'clean'} & \text{li}\text{mpjU} \\
\text{s\textacute{\textalpha}\textacute{\textalpha}ju} & \quad \text{'dirty'} & \text{s\textacute{\textalpha}\textacute{\textalpha}jU} \\
\end{align*}

\textbf{c. augmentative} \hspace{1cm} \textbf{count}
\begin{align*}
\text{bu\textacute{\textalpha}k\textacute{\textalpha}n} & \quad \text{'donkey'} & \text{bU\textacute{\textalpha}k\textacute{\textalpha}kU} & \quad \text{'donkey'} \ (\text{pej.}) \\
\text{a\textacute{\textalpha}y\textacute{\textalpha}n} & \quad \text{'big arroyo'} & \text{A\textacute{\textalpha}y\textacute{\textalpha}yU} & \quad \text{'arroyo'} \\
\text{kant\textacute{\textalpha}n} & \quad \text{'rock that has been rounded and worn by water'} & \text{k\textacute{\textalpha}ntU} \\
\end{align*}

\footnote{In fact, the domain of this harmony is actually the clitic group, which Penny calls the “syntagma”.}

134

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Again, Steriade attributes the strong licensing capacity of the final syllable to the extra duration found in that position, duration which (Zhang adds) assists in the perception of difficult contrasts. But is this really the case in Pasiego Spanish? A number of facts cast suspicion on this hypothesis. First among these is the description Penny (1969: 149) gives of the realization of final syllables in Pasiego. According to him, in both closed and open final syllables, “the vowel sounds of pasiego speech are very relaxed (although not as much as in Portuguese, for example) and are almost entirely devoiced when a voiceless consonant or [rr] precedes”. In final syllables, even the tense /u/ is realized lower (at “similar” aperture as /u/) and laxer, the difference between /U/ and /u/ now being only that the former is “centralized” while the latter is a “clear velar”. /e/ in final position is raised (“almost an open ‘i’”), which suggests why there is no contrast between front mid and high vowels in these syllables. Recall also from note 6 above that the masculine mass noun suffix /-u/ comes from an earlier -o. All this suggests vowel weakening in the final syllable, perhaps if anything less duration than in other unstressed syllables.

To this we can add facts from neighboring Tudanca Spanish (Hualde 1989, Flemming 1993, Dyck 1995). In Tudanca Spanish word-final high vowels are phonetically laxed⁴⁹. This laxness spreads left to the stressed vowel, laxing any

⁴⁹ There is no phonemic contrast between tense and lax vowels in Tudanca.
intervening vowels as well in the process. Additionally, in word-final unstressed syllables
the contrast between underlying /i/ and /e/ is neutralized, both vowels being realized as
lax [i] or [ə]\(^{30}\). Again, this neutralization suggests a position susceptible to vowel
reduction, i.e. actually a rather poor place to perceive fine contrasts in vowel quality. A
third piece of evidence comes from Sanders (1994), who presents an extensive study of a
similar (but unrelated) laxing harmony in the dialects of Eastern Andalusia. Sanders
provides data from an experimental study of the phonetics of vowels in this dialect,
showing that word-final (phrase-internal) vowels in this dialect (both tense and lax) are
shorter than the vowels of stressed syllables, but longer than the vowels of pretonic
unstressed syllables. For example, mean duration from three speakers for tonic tense /a/
was 86.98 ms, for word-initial tense /a/ 60.28 ms, and for word-final tense /a/ 76.85 ms.
Values for lax /A/ (not Sanders’s notation) were 90.18 ms, 58.24 ms, and 73.50 ms
respectively. Ratios for other vowels were similar (Sanders 1994: 168). Sanders had
hypothesized that his lax vowels might be phonetically longer than his tense vowels. The
reason he chose to measure final vowels in non-pre-pausal position was, in his words, “to
prevent vowel reduction due to suprasegmental processes” (1994: 159). Specifically, “In

\(^{30}\) This neutralization does create surface contrasts between laxed vowels preceding underlying [i] and tense
vowels preceding underlying [e], as in the pair /abri/ [Ábra] ‘open’ and /abra/ [Ábra] (Flemming 1993: 11).
eastern Andalusian the phrase-final position is subject to considerable reduction" (1994: 118).

From the above it is clear that in Pasiego and other laxing harmony languages of Spain we are not dealing with the type of system in which certain features occur contrastively only in positions where they are perceptually most robust (such as in some of the vowel reduction systems discussed in Chapter 2, or the strong-to-weak harmony systems dealt with below in Chapter 4). In fact, the very opposite appears to be the case. In Pasiego Spanish tense and lax contrast underlyingly only in a position, whose phonetic characteristics would actually hinder accurate perception of fine differences in vowel quality (i.e. a position characterized by shorter durations, laxing or centralization of vowels, in some systems even raising of mid vowels leading to merger with highs, and increased susceptibility to devoicing). In these systems the direction of spreading then is not from strong position to weak position, but from weak (phonetically non-prominent) to strong. This characterization places laxing harmony in a class with other weak-to-strong assimilation processes such as those traditionally labeled umlauts or metaphonies. The notion of umlaut or metaphony being motivated primarily by the perceptual weakness of

---

51 In fact, the very northwestern Spanish dialects in question also exhibit vowel height harmonies. While the systems are complex (see Flemming 1993 or Dyck 1995 for details), in Pasiego for example, the /-U/ suffix conditioning laxing harmony also raises tonic mid vowels to lax high vowels in a manner reminiscent of the metaphony systems of Italian dialects. This in itself is perhaps the strongest evidence of the type of process involved here.
the trigger (as opposed to its phonetic salience, as Steriade suggests for Pasiegos) is central to Ohala's theory of sound change (as in e.g. Ohala 1993, discussed below). Its isolation as the feature differentiating harmonies proceeding from non-prominent to prominent positions (perception-based) from those from operating from prominent to non-prominent (articulation-based) is the central thesis of McCormick (1982), and is explored more recently in Majors (1998: 160-167) and Walker (2001).

From a diachronic point of view such patterns have a relatively straightforward explanation. Specifically, they are examples of hypocorrective change, in the Ohalian parlance (Ohala passim, but especially e.g. 1993). The term hypocorrective change refers to instances in which the listener hears, for example, the coarticulatory effect on a vowel of a neighboring segment, but rather than correctly attributing it to its actual source and disregarding it, or taking it as a remote cue for the identity of the source segment, the speaker instead takes the extended feature to be a property of the vowel in question, and constructs a new underlying representation for the form as result. A simple example would be the development of a contrastively nasal vowel from an earlier vowel-nasal sequence. In Ohala's schema, the speaker, with a UR of /VN/, produces the coarticulated sequence [vN]. The listener then fails to attribute the nasalization he or she hears to the following nasal consonant, hearing instead [v], and believing this to be the speaker's intended pronunciation, constructs a new representation for the form in question: / v/. In
the metaphony case, then, a sequence /éCi#/ would perhaps include some raising of the
tonic mid vowel as a result of vowel-to-vowel coarticulation. A correct interpretation of
the speaker’s intentions would attribute the raising effect to the following high vowel and
parse the tonic vowel as mid regardless. The final high vowel, however, occurs in a
position in which accurate perception is impeded by the phonetic weakening effects
discussed above. It is possible, then, that a listener could misperceive the /éCi/ sequence
such that he or she failed to attribute the coarticulatory raising to a following high vowel,
reconstructing the speaker’s intended production instead to be an actual stressed high
vowel in the penultimate syllable.

Another, synchronically-oriented approach to weak-to-strong harmony patterns is
that of Walker (2001). In treating, *inter alia*, the metaphony patterns found in Italian

52 This scenario is not without complications. It is most convincing, to my mind, in instances where the
conditioning environment for the change is actually lost, a result of the perceptual failure preventing correct
attribution of the coarticulatory effect. In the metaphony case (and others), however, the conditioning
environment is not lost. Ohala (1993: 246-247) suggests that in such cases the listener might perceive both
segments accurately, but lack the experience to attribute the coarticulatory effect to its actual source (e.g.,
the listener is a child, still acquiring the phonological patterns of the language). One problem here is that it
makes clear predictions about differences between child-initiated and adult-initiated sound changes, which,
to the extent that these are distinguishable, are wanting in empirical substantiation. Another problem is that
the acquisition-based explanation makes irrelevant the fact that the trigger for such changes is in a
perceptually-non-optimal environment (or is actually weakened itself articulatorily), but surely this is not
accidental. The seeds of a resolution to this problem may lie in another aspect of this type of change. In
Ohala’s conception, the listener’s new UR for a given item will often give rise to lexical doublets
(assuming the listener does at least occasionally parse this item correctly). It is possible to imagine in such
instances a process of leveling across competing representations leading to forms containing both the
misparsed target and the original trigger of the change. One is reminded in this connection of the exemplar-
based models of the lexicon defended in, e.g., Pierrehumbert 2001). These problems await resolution.

139
dialects, Walker notes the differences between these patterns and, for example, stressed-syllable triggered harmonies (also present in Romance dialects). She characterizes the two patterns as based on different phonetic motivations and deserving of different synchronic treatments. Walker too sees the metaphony patterns as resulting from the perceptual weakness of the triggering segment (inherent or positionally acquired), and attributes the feature spreading not to misperception, but to the purposeful extension of perceptually-weak features to phonetically-prominent positions such as the stressed syllable. Both the realization of the weak(ened) features in a perceptually-more-hospitable environment and the mere fact that the extended feature now occurs over a span of longer duration would contribute to increased perceptual robustness. Here Walker is invoking the principle central to the thesis of Kaun's treatment of rounding harmony (which Kaun attributes in part to Suomi 1983): bad vowels spread (perceptually difficult contrasts are more easily perceived if they are realized over longer spans within the word). To which it seems necessary to add, in the case of metaphonies involving major distinctions in vowel height, that in fact even good vowels will spread if they wind up in bad enough neighborhoods. The notion behind Walker's story then is essentially a functionalist one, in which the spreading of vowel features takes place in order to preserve vowel contrasts which are otherwise in danger of collapse due to features of their environments. Walker notes in this connection a generalization made by Dyck.
(1995) about metaphony in Spanish and Italian dialects: metaphony does not occur (high vowels do not raise preceding vowels) from positions in which there is no contrast between high and mid vowels. If metaphony were not at least in part concerned with the preservation of contrasts in the triggering syllable, such a generalization would go unexplained. It also seems relevant in this regard that in the languages in question the triggering vowels are generally the sole exponents of one or another morphological category, meaning that neutralization of contrasts here would result in the loss of grammatical information. Recall that in the case of the only vowel triggering laxing harmony in Pasiego, neutralization of the tense-lax opposition would result in the collapse of the mass vs. count opposition for masculine nouns as well.

---

53 For this generalization to hold, of course, systems in which subsequent neutralizing reductions (of all Finals to schwa, of -o- to -u-, etc.) have introduced opacity into the metaphony pattern must be analyzed as having a high-mid contrast at least underlyingly.

54 It is possible too that in some ways Ohala’s view and Walker’s are not so mutually exclusive as they otherwise seem. In the Ohalian scenario, for example, the original coarticulatory effect is generally taken as a given. It is something properly “unintended”, which the listener for whatever reason erroneously parses as “intentional”. While Ohala is clearly referring to the realm of the phonemic here, it is by now well known that coarticulation strategies are both planned and language-specific (which is to say, as “intentional” as anything else in phonology). It is also commonly accepted that segments may be cued by features which are not temporally coextensive with them (one thinks of vowel duration as a cue for the voicing distinction in following obstruents in English). It is perhaps not far-fetched to imagine that cues for a positionally-challenged contrast might become phonetically maximized, such that the coarticulatory effect involved in the hypocorrection is in fact particularly robust in those very instances in which the hypocorrections take place. The input to the change would thus be inherently skewed toward the occurrence of that same, perhaps suggesting as well a way of understanding why the change takes place despite the retention of the conditioning environment. All this is fertile ground for future experimental investigation.
It is clear from the above that in systems such as Pasiego Spanish, the privileged licensing of the features in question is not due to the phonetic characteristics of that position in the way that, e.g., retroflexion contrasts are preferentially licensed in postvocalic position due to the salience of cues for this distinction in that positions (i.e. VC transitions [Steriade 1999]). In many cases contrasts for which final position is phonologically strong are not "licensed" in any sense by the phonetic characteristics of that position. In some cases, such as Pasiego Spanish, those characteristics may even be detrimental to the perception of the contrast (a fact which one way or another provides the ultimate motivation for the harmony patterns which may develop there). Further examples of cases of final syllable phonological strength owing explicitly to final syllable phonetic weakness are discussed in the section 3.7 below. More often than not this turns out to be the case in systems where final position alone is the sole licenser of the full range of contrasts in the inventory. In this sense Hausa, described above, is the exception. None of the other putative instances of (non-tonal) final syllable strength has this characteristic. Rather, where final strength is observed at all, it is most often in the form of which I will call Final Resistance, introduced in the next section.

5 Those conditions may in fact be the ultimate diachronic source of the contrast (as in the Eastern Andalusian laxing harmony systems in which the lax final vowels of the plural suffix develop from loss of an earlier word-final /-s/, presumably with the intrinsic periods of breathiness on the preceding vowels becoming extrinsic thereby (Sanders 1994).
3.2.3. Final syllable resistance to assimilation and reduction

By far the most common vocalic strength effect that can observed in final syllables is some degree of resistance to reduction or assimilation processes which would otherwise target vowels of similar prosodic status (unstressed, short, non-initial, etc.). In such cases the final syllable is not the strongest licensor in the word (as it is in Hausa), but rather shares this distinction with some other prosodically-prominent syllable. In many, perhaps most of these systems, the strength effect is observed only at the phrase or utterance level, with phrase-internal word-final syllables submitting to reduction (often described as gradient or optional), in contrast with the obligatory categorical reduction or assimilation of other non-prominent vowels. Very often the strength effect itself bears some hallmark of gradient application: it is described as “optional”, rate- or style-dependent, or even only partial. The following is a list of the languages in which it has been possible to identify relatively clear cases of final-syllable resistance. As this particular pattern of final strength seems not to be widely remarked upon in the

56 A major problem for the researcher wishing to determine whether, say, a given vowel reduction language exhibits this effect lies simply in assessing the possible significance of a source’s silence on the matter. In the course of this survey it became clear that for the subtler effects even in generously-described languages native to societies with long-standing local traditions of phonetic experimentation (e.g. Russian), only the most conscientious descriptions were likely to contain a direct assertion on this topic, to say nothing of secondary sources from the international phonological literature.
literature, a few words on each system are in order as well. The catalogue below in (20) is illustrative, but by no means exhaustive. It is followed by subsections discussing common characteristics of Final Resistance patterns and the ramifications of these for the places of phonetic and psycholinguistic factors in an account of final strength effects.

(20) Reduction or assimilation-resistant final syllables


*Maximal inventory:* [i, e, a, o, u, i]

*Reduction processes:* Basic facts (See chapter 2 for some additional details)
- In first pretonic syllables, /a/, /o/ -> [e]
- In first pretonic syllables following C[^1], /a/, /e/, /i/ -> [i]
- Other unstressed syllables: /a/, /o/ -> [a]
- Other syllables following C[^1], /a/, /e/, /i/ -> [i]

*Resistance:* In phrase-final open syllables, /a/, /o/ -> / [æ], as in first pretonic syllables (also associated with more duration than other unstressed syllables). Following palatalized consonants, /i/, /e/ -> /i/ but /a/ -> approx. [æ]

*Conditions:* Phrase-final only. Phrase-internal final vowels are often extremely weak, prone to devoicing or deletion. Reduction to [æ] is gradient in all cases (see Chapter 2). Phrase-final /a/ and /o/ often described as somewhere between the first and second degrees of reduction. Rate-dependent, greater reduction likely in fast speech or less "conscientious" style.

b. **Belorussian** (Slavic) (Czekman and Smulkowa 1988: 223-234)

*Maximal inventory:* [i, e, a, o, u]

---

57 Flemming 1993 notes Yakan and Maltese, though see below on some problems here. Kirchner 1998 gives an analysis of Nawuri, but in a somewhat different context.

144
Reduction processes: Basic facts (See chapter 2 for additional discussion)
In first pretonic syllables, /a/, /o/ /e/ -> [a]
In other pretonic syllables, /a/, /o/, /e/ -> [e]

Resistance: In final open syllables, /a/, /o/, /e/ -> /[a]/. As in first pretonic syllables.
Conditions: Resistance does not reverse categorical neutralizations, only realization of height of resulting low vowel. Not clear from this source whether resistance occurs in all word-final syllables or only in phrase-final position.

c. Ukrainian dialects (Slavic) (Shevelov 1979: 218-240)

Maximal inventory: [i, ɪ, e, a, ɔ, u]
Reduction processes: /i/, /e/ -> [ɪ]
(/ui, /o/ -> [ʊ])

The dialect geography of vowel reduction patterns in Ukrainian is complex, and cannot be treated at length here. The processes most commonly involved are the raising of /e/ and /o/ toward /i/ and /u/ to one extent or another and in various combinations. The pattern in which the distinction between /e/ and /i/ is categorically neutralized in unstressed syllables is the most common, being characteristic of all Ukrainian dialects save the eastern and central Polissian and the Lermkian (northern periphery adj. to Belorussian and far western periphery in Slovakia respectively). Standard Ukrainian has only this reduction (Shevelov 1979, Pugh and Press 1999). Disregarding here less robust instantiations of the change, across-the-board neutralization of /o/ and /u/ is much less general, obtaining primarily in a North-South strip from Bukovyna to Pidljašia (on the western periphery, but excluding some far western dialects (Shevelov 1979: 519).

58 Some descriptions of Belorussian do not note the raising of the low vowel in the reduced system of oppositions outside the first pretonic syllable, and also claim that /e/ outside the first pretonic is realized as [e], curiously reversing its merger with /a/ and /o/ in a phonetically-less-prominent syllable than the one in which the merger takes place. These descriptions are apparently based on the spelling norms of the Belorussian literary language, and do not reflect phonetic reality (Czekman and Smulkowa 1988: 229-230).

59 The term used appears to be the Polish translation of Auslaut.

60 Depending on the dialect, or even the individual speaker in some areas, both vowels are realized as [e], as [ɪ], or as something varying between the two, in some dialects assimilating to some degree to the height of the stressed vowel (Shevelov 1979: 520).
Resistance: In most dialects reduction does not apply to word-final vowels (sources make no mention of phrase vs. word distinction. In the dialects of the Berestja region (northwestern), by contrast, reduction is apparently resisted in word-final open and closed syllables.

Conditions: Shevelov describes reduction as categorical in the dialects where it occurs regularly. Other sources speak of a gradient-sounding process, e.g. the neutralizing vowels having only a “tendency to approximate one another” (Pugh and Press 1999), or of reduction being optional (Janiak 1995).


Maximal (oral vowel) inventory: /i, e, a, ɔ, o, u/
Reduction processes:
- Pretonic syllables - /e, ɛ/ -> [e], /o, ɔ/ -> [o]
- Posttonic syllables - /i, ɛ/ -> [i], /u, ɔ/ -> [u]

Resistance: For some speakers the second degree of reduction (posttonic high-mid neutralization) does not take place in prepausal word-final vowels (Major 1985: 266-267).

Conditions: The pretonic mid-vowel neutralization is categorical (non-rate dependent, not reversible in careful speech. The posttonic reduction is gradient (rate dependent, reversible in careful speech). Major 1985 suggests that for those speakers whose lengthened prepausal final vowels reduce nonetheless, posttonic reduction too has become “lexical”. In casual speech, pretonic high and mid vowels are optionally subject to the same (gradient) reduction as posttonic vowels (Major 1985: 266)


Inventory: Full: /i, y, e, ə, a, o, u/ Reduced: /a/

Reduction: Stress falls on the rightmost (underived) full vowel of the word. Words containing only reduced vowels take the default initial stress. Only the reduced vowel
occurs in non-final posttonic syllables.

Harmony: Posttonic vowels harmonize with the stressed vowel in backness and roundness. Final (non-low) vowels do so in a categorical, neutralizing fashion. Non-final reduced vowels have variants with allophonic fronting and rounding in agreement with the stressed vowel.

Final strength effects: No short (reduced) vowels surface word-finally (Flemming 1993). Results in alternations, whereby root-final or suffix mid vowels surface as reduced when appearing non-finally. I class this with resistance to reduction effects in that all posttonic vowels save those in word-final open syllables are realized as schwa. The disallowing of short word-final vowels, on the other hand, results in the neutralization of potential full vs. reduced contrasts.

ex. tələze 'moon' -> tələzəste 'in the moon' (inessive)
   ludʃə 'reader' -> ludʃən 'reader's' (genitive)
   pəɾəʃə 'in the house' -> pəɾəʃəzo 'in his house'

Conditions: Described as categorical, affecting all word-final vowels. Varies greatly across dialects.

f. Uyghur (Turkic) (Hahn 1991)

Inventory: /i, y, e, æ, ə, i, a, o, u/  

Reduction: Low, short vowels in non-initial open syllables are raised to [i]

Resistance: Applies only to non-prepausal word-final vowels (Hahn 1991: 52).

Conditions: Applies categorically within words but is rate-dependent across word boundaries.

62 No contrasts, however, are preferentially preserved in this position, as final unreduced mid-vowels harmonize with the earlier stressed vowel. In light of this, Eastern Mari does not seem to exhibit strong licensing of any kind word-finally, pace Steriade 1994.

63 /u/ and /u/, analyzed as independent phonemes, in fact share the same surface realizations, differing only in the harmony pattern they condition (front or back).
g. **Central Eastern Catalan (Romance)** *(Recasens 1991)*

*Inventory:* /i, e, e, a, ə, o, u/

*Reduction:* Unstressed syllables - /e, e, a/ -> [ə]  
/u, o ə/ -> [u]  
/i/ -> [i]

*Resistance:* In a number of dialects in this area, e.g. Oden (transitional between Northeastern and Central eastern), word-final vowels are resistant to the elimination of oppositions associated with unstressed syllables *(Recasens 1991: 111)*

h. **English (Germanic)** *(Hammond 1997)*

*Inventory:* /i, ɪ, e, ɛ, æ, a, ə, ɔ, o, u, w/

*Reduction:* Unstressed word-internal preconsonantal syllables contain only [ə]*

Unstressed word-internal prevocalic syllables contain [i, u, o, ə]

*Final strength effects:* Word-final open syllables show [i, u, o, ə]

i. **Yakan (Austronesian: Philippines, SW Mindanao, Basilan Island)** *(Behrens 1975, Flemming 1993)*

*Inventory:* /i, e, a, o, u/

*Reduction:* Primary stress is penultimate, with secondaries on alternating syllables preceding. In unstressed syllables /a/ -> [e].

*Resistance:* In word-final open and closed syllables reduction does not take place (nor does it in pretonic syllables when the following syllable is also [a]).

*Hammond’s analysis admits only one reduced vowel.*

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
j. Maltese (Semitic) (Puech 1978)

Harmony: In Standard Maltese and closely related dialects, there is a left-to-right rounding harmony which targets all short non-low vowels triggered by [o]. In Siggieha and Gozitan dialect, harmony affects vowels of all heights, affecting specifications for [back].

Resistance: While in some dialects suffix vowels only harmonize if they are in a final closed syllable (with no harmony taking place if additional syllables intervening between the stem and the word-final syllable, even if intervening syllables contain otherwise-eligible vowels), in no dialect do vowels in final open syllables ever harmonize.65

k. Nawuri (Kwa) (Casali 1995)

Inventory: /i, I, e, a, ə, o, u, u/

Reduction: Underlying short, front vowels centralize: /i, ɪ, e, ə/ -> [i, ɪ, ə, ɻ].

Resistance: Does not apply to absolute word-initial vowels (presumably due to increased duration, cf. Russian vowel reduction). Does not apply to vowels in phrase-final position.

65 A possible explanation for this not involving final lengthening directly is that while synchronically all surface final vowels are short (phonologically), underlyingly they are all long. Maltese has undergone a common sound change (to be discussed below in 3.6 and 3.7), whereby long final vowels shorten and short final vowels are deleted. The original length of the remaining final vowels could have prevented their harmonization. Flemming (1993: 21) does not discuss harmony, but notes that final syllables closed and open are never subject to unstressed vowel deletion in Maltese, and attributes this resistance to final lengthening. This may ultimately be so, though I see two reasons for skepticism. First is the tendency for final closed syllables (indeed stronger than for internal closed syllables in some dialects), but not final open syllables, to undergo harmony, suggesting if anything decreased duration in final closed syllables. Another problem is in the nature of the deletion process itself. Syncope, at least in many Arabic dialects is similar to French schwa deletion in that it is restricted in that it only targets vowels in open syllables. While final open syllable resistance is therefore possibly due to final phonetic prominence, final closed syllable resistance is more likely a consequence of syllable structure.
Casali notes that in word-internal syllables, centralization is obligatory, even in very deliberate speech. To (phrase-medial) word-final vowels centralization applies “postlexically”. Kirchner (1998) uses Nawuri centralization as an argument for the need to treat even universally non-contrastive phonetic properties such as the additional duration of final vowels as phonological. His reasons are as follows: There is a lexical process of left-to-right rounding harmony targeting high non-front vowels which applies from root-to-prefix and holds as a static generalization within roots as well. This rounding harmony does apply to prefixes which undergo centralization (i.e. become non-front), but (of course) does not apply to absolute word-initial prefix vowels (which fail to undergo centralization. Kirchner argues from this that since the output of centralization conditions phonological rounding harmony, centralization (which is sensitive to subphonemic durational variations) cannot be “relegated” to the phonetic component, and hence subphonemic durational features must be available to the lexical phonology. This, however, cannot be right for the following reason: Kirchner ignores Casali’s claim that centralization applies to word-final vowels only postlexically, while its internal application is lexical and obligatory/non-rate-dependent. The lexical application, then, is clearly not conditioned by subphonemic variations in duration, while the postlexical application seems to be. This, then, is another instance of the categorical application of a process word-internally, with gradient application word-finally. Additional support for
this view comes from the application of the clearly-lexical rounding harmony. Casali notes (p. 653, note 4) that in addition to its lexical application, rounding harmony also applies postlexically across a word-boundary to (postlexically) centralized phrase-internal word-final vowels. This postlexical application, however, is "optional", and produces final vowels of only "an intermediate degree of rounding" (viz. applies gradiently). Clearly then both centralization and rounding harmony apply both categorically in the lexical phonology and gradiently in the postlexical phonology. Further proof of the dual nature of these processes comes from the absolute initial vowels. The exceptionality of these is clearly due historically to increased duration. Synchronically, however, absolute initial vowels never centralize even postlexically (p. 655), and hence are never subject to rounding harmony lexical or postlexical (which is to say, their exceptionality, previously a function of their phonetic duration, has since been phonologized, such that it is now derived from their structural description alone. Final syllables retain the earlier state, in which resistance to alterations is still directly contingent on physical duration).

1. **Shimakonde (Bantu, Mozambique)** (Liphola 1999)

   **Inventory:** /i, e, a, o, u/

   **Reduction:** Optional reduction of pretonic mid vowels /e/ and /o/ to [a]. Stress is fixed penultimate.
Resistance: Final vowels never reduce.  

m. Bonggi (Austronesian, NE Borneo, Banggi and Balambangan Islands) (Boutin 1993, Kroeger 1992)  

Inventory: /i, ə, a, ɔ, u/ underlying in native roots (Kroeger 1992: 291)  

Reduction: Pretonic - Non-high vowels reduce to /a/  
Posttonic (final syllable) - /a/ -> [a]  

Resistance: Does not apply to word-final vowels  

n. Timugon Murut (Austronesian, NE Borneo) (Prentice 1972, Kroeger 1992)  

Inventory: /i, a, o, u/  

Reduction: /o/ is not contrastive in pretonic syllables. All occurrences are derivable synchronically from left-to-right spreading of [round].

---

66 A number of the systems noted here (Shimakonde, Yakan, Murut) have fixed penultimate stress and final syllable vowels which avoid reduction, Yakan and Murut even in closed syllables (an oddity as will be shown below). It is not outside the realm of possibility that the failure of these final syllable vowels to reduce is due not to duration accorded them by final lengthening, but rather due to their presence in the head foot of the word (see Alderete 1995 for discussion of faithfulness effects in prosodic heads). Zoll (1997) presents an instance of this effect in non-final position in Guugu Yimidhirr, in which a contrast between long and short vowels is permitted only in the head foot of the word (the initial and peninitial syllables). In the cases at hand, the lack of any non-final posttonic syllables makes it impossible to determine whether resistance to reduction is a property of the final alone, or posttonic syllables more generally.

67 All instances of final syllable [ə] have [ə] in the preceding stressed syllable, indicating their derivation by rounding spread (independently attested in other contexts).

68 Neighboring related languages (e.g. Dusunic Kimaragang, Kroeger 1992) merge /a/ and /ə/ in pretonic syllables as /a/. This situation is most likely the immediate precursor to the distribution in Timugon Murut, which seems to have restored full vowel quality to pretonic non-highs, with [a] and [o] predictable from the identity of the stressed syllable vowel.
Resistance: /o/ does occur contrastively in the stressed (penultimate) syllable and final syllables (closed or open) (see note 51 on Shimakonde). Furthermore, [o] only surfaces in stressed syllables if the following syllable also contains [o], meaning that the distribution of [o] is actually least restricted in the (posttonic) final. See below for a cautionary tale, however, on the dangers of extrapolating claims of special final prominence from distributional evidence alone.

3.2.3.1. The phonetic roots of Final Resistance

This pattern of final syllable resistance to vowel reduction or assimilation processes is very much expected given what is known about the phonetics of domain-final syllables crosslinguistically. If these processes target vowels which are for whatever reason durationally-deficient, final lengthening would naturally serve to exempt the vowels of final syllables from the effects of the duration-sensitive processes. Qualitative vowel reduction as in Russian or Brazilian Portuguese targets unstressed vowels and occurs primarily in languages with strongly duration-cued stress (see Chapter 2). In Nawuri, centralization exempts specifically long vowels, absolute-initial vowels, and phrase-final vowels, as noted, all three positions characteristically associated with increased duration. In Uyghur, raising applies only to non-initial low vowels in unstressed syllables. This may seem odd at first, particularly since Uyghur, like most Turkic languages, has fixed final (not initial) stress. It makes more sense in light of phonetic facts documented in other Turkic languages: Turkic final stress is not strongly correlated with increased vowel duration (see Konrot 1981 for experimental evidence
In Turkish (Barnes 2001a, and Chapter 4 below) and the closely-related Turkmen (Mollaev 1980), it has been shown that vowels in word-initial syllables, regardless of the placement of stress, are realized with longer phonetic durations than the vowels of comparable word-internal syllables. Additionally, in Anatolian Turkish, contrary to near-universal expectations, all things being equal vowels in closed syllables are longer than vowels in comparable open syllables (Lahiri and Hankamer 1988, Jannedy 1995, Kopkali-Yavuz 2000, Barnes 2001b). The shorter duration characteristic of open syllables in Turkish has the result of conditioning frequent reduction of /a/ to [ə] in ordinary speech. If these two durational regularities are found also in Uyghur, then raising can be seen to fail specifically in initial syllables, closed syllables, syllables with (underlying) long vowels, and phrase-final open syllables; these are all positions in which vowels would have characteristic additional phonetic duration.

Phonetic studies of final lengthening add empirical weight to this hypothesis. In addition to the temporal augmentation of vowels in domain-final syllables, several researchers have identified a process of (spatial) articulatory strengthening as well.

---

69 Stress in Uyghur is generally final in native vocabulary, and is described by Hahn as relying primarily on a pitch distinction, with durational differences between stressed and unstressed syllables being “less pronounced ... than in most European languages” (Hahn 1991: 27).

70 Or at least were found at the time of the development of the raising process.

71 Long vowels created through compensatory lengthening still undergo raising where applicable (Hahn 1991: 56), highlighting the independence from actual phonetic duration of the categorical word-internal raising process.
Edwards et al. (1991) in a study of the gestural implementation of final lengthening in English concluded that while phrase-final stressed /a/ was longer and slower than its non-final counterpart (indicating a local change in gestural stiffness), a phrase-final reduced vowel [ə] was not only longer, but also enhanced by greater gestural displacement (in their measure of jaw height).

Vayra and Fowler (1992) find a strikingly similar pattern in their investigation of supralaryngeal articulatory declination in Tuscan Italian. While for stressed vowels (/a/ and /i/) measures of F0, amplitude and jaw opening decreased monotonically from beginning to end in trisyllables uttered in isolation, for unstressed vowels (again /a/ and /i/) the same declination was observed for F0 and amplitude, but measures of F1 and jaw opening showed minima in the second syllable, while unstressed vowels in final syllables showed higher measures, suggesting supralaryngeal strengthening of final unstressed vowels of the same type observed by Edwards, Beckman and Fletcher.

Cho (2001), in a detailed study of gestural strengthening in accented and boundary-adjacent syllables in English found articulatory strengthening reminiscent of that found in pitch accented syllables for preboundary vowels as well. Specifically, Cho

\[\text{\footnotesize\textsuperscript{72}}\text{ A non-metaphony dialect in which there can be occasional apocope of phrase-internal word-final vowels following sonorants, but no such process phrase-finally. This is in contrast with certain southern dialects, e.g. Pugliese, which reflect metaphony and in which all posttonic vowels are reduced to schwa and word-final vowels are deleted phrase-finally. (Giannelli 1997: 298, Loporcaro 1997: 340-341). See below for discussion of final vowel weakening processes.}\]
noted in domain-final vowels what he refers to as sonority enhancement, measured as a lower tongue position and greater lip opening gesture for both phrase-final [o] and [a]. He did not, however, record the lower jaw position found by Edwards et al. and Vayra and Fowler, and which he himself notes for vowels under phrasal accent. Note, however, that Cho measured only vowels carrying at least lexical stress, and then varied their prosodic position with the phrase. Since the lower jaw position in final vowels recorded by both Vayra and Fowler and Edwards et al. was found only for lexically unstressed vowels, it is unsurprising that Cho saw no such effect. In addition to this, Cho also demonstrated increased resistance to vowel-to-vowel coarticulation for both accented and phrase-final vowels, another result that provides insight into the phonetic origins of the phonological final strength patterns described above. The fact that so many of the phonetic studies of final lengthening have detected the phenomenon in a robust and consistent manifestation only before higher phrase boundaries likewise makes clear the reason why so many of the Final Resistance patterns do not operate in word-final syllables internal to the phrase.

Though studies of the articulatory properties of final syllable vowels have so far dealt with only a very few languages, the frequency of occurrence of Final Resistance

73 While Cho finds sonority expansion for domain-final vowels, he found no evidence of “featural enhancement” of those same. Under accent, in agreement to some extent with the finding of de Jong (1995), Cho found that [i] was realized fronter (but not necessarily higher) and that [a] was realized lower (but not necessarily backer).
effects (in conjunction with the understandable reticence of most grammar writers on this matter whether the effect is present in the language or not) might suggest that the articulatory characteristics documented for English and Tuscan Italian final lengthening might in fact be properties of the phenomenon in general, and hence that most vowel reduction systems should show at the very least a low-level phonetic version of Final Resistance. There exist at least two studies of the spectral characteristics of final syllable vowels in languages with unstressed vowel reduction, however, which show this not to be the case: Nord’s studies of vowel reduction and duration in Swedish (1975,1986), and Johnson and Martin’s (2001) study of vowel reduction in Muskogee Creek.

Nord’s studies showed that despite clear application of final lengthening (with final-syllable unstressed vowels approximating initial-syllable stressed vowels), all vowels tested (/e/ and /a/ in 1975, /e/, /a/, /i/ and /a/ in 1986) still underwent reduction. It is, however, worth noting several things about these studies: Firstly, Nord measures only vowels in closed final syllables, which, as will be shown in the following section, in many languages at least undergo less lengthening than vowels in final open syllables (immediately adjacent to the word- or phrase-boundary). Also, while the final-syllable unstressed vowels do reduce in comparison with their stressed counterparts, it is clear from Nord’s data that for vowels other than /a/, there is a statistically significant difference between formant measurements for final-syllable unstressed vowels and
unstressed vowels in initial syllables. Strikingly, for /a/, /i/ and /e/, the final syllable unstressed vowel has a higher F1. Nord seems to take this to be the “coarticulation with the rest position” he claims is responsible for the movement toward a schwa-like vowel particularly characteristic of final unstressed /e/ in Swedish. But the higher F1 for /a/ as well suggests that this is not necessarily the case. In fact, it is more reminiscent of the final sonority-expansion documented by Cho (2001) for English. If this is true, then final lengthening does inhibit reduction somewhat in unstressed final closed syllables in Swedish74. A comparable study of open syllables would most likely clarify this matter.

Johnson and Martin (2001) show more conclusively that in Muskogee Creek final vowels are generally more centralized than non-final, though final lengthening is operative. This suggests that at least in Creek final lengthening is carried out without concomitant articulatory strengthening. Johnson and Martin wish to attribute this to a “final fade” in supralaryngeal articulations similar to that found in F0 and amplitude at the ends of phrases. They hypothesize that, along with lengthening, decreased gestural stiffness could also cause articulatory reduction, citing Beckman et al. (1992), Vayra and Fowler (1992), as well work by Krakow (1993) and Vassiere (1986) showing declination of velum height.

74 Recall that even in Russian final lengthening does not always reverse reduction of /a/ completely.
As described above, however, the results of Edwards, Beckman, and Fletcher’s later study, and in fact the results of this very study of Vayra and Fowler show just the opposite of final fade when it comes to unstressed vowels, any declination being found only for the stressed counterparts thereof. Also, later work by Vassiere (1988) demonstrates stronger velic closure in phrase-final syllables than in phrase-medial, again arguing against a general phenomenon of final supralaryngeal fade. To this can be added the work of Keating, Wright and Zhang (1999), who found increased linguopalatal contact for phrase-final consonants in English, and of course Cho (2001), again demonstrating the link between phrase-final lengthening and articulatory strengthening, the later in fact specifically for lexically stressed syllables.

In sum, at least for unstressed syllables there is very little (if any) direct articulatory evidence for supralaryngeal “final fade”, while at the same time there is an abundance of evidence for the opposite effect: phrase-final strengthening of unstressed syllables, and in some studies stressed syllables as well. While this casts doubt on Johnson and Martin’s more general interpretation of their results, it still leaves the facts of Creek unexplained. Were this all simply a matter of phrase-final vowels lengthening while still reducing to the same degree as phrase-internal vowels, it would be possible to account for the application of reduction in the presence of lengthening through phonologization, as the generalization of spectral targets from the phrase-internal variants...
of word-final vowels to all tokens of those vowels regardless of phrasal position or
duration. As noted above, Major (1985) argues precisely this to have taken place for
those speakers of Brazilian Portuguese who have generalized reduction of the mid vowels
to [i] and [u] even in phrase-final lengthening contexts75. Even this, however, fails to
resolve the entire problem: a number of languages show reduction of vowels only or
especially in phrase-final position. Here it is clear that no such generalization has taken
place. The phrase-final syllable is simply weaker phonetically than the non-phrase-final.
Two examples of languages apparently displaying this pattern of phrase-final reduction
have already been cited76. I will return to this problem below in discussion of final
weakening effects. For now it is enough to note that if a process of supralaryngeal final
fade exists at all, it is far from universal, given the abundant attestation of the very
opposite pattern in both phonetic studies and patterns of resistance to reduction and
assimilation phenomena which should otherwise have targeted the resistant final vowel.
As for Final Resistance, the foregoing shows clear phonetic motivations for the existence
of such a pattern. Studies like that of Johnson and Martin, though, provoke questions

75 In Bulgarian, for example, phrase-internal word-final unstressed vowels are generally extremely short,
moreso even than other unstressed vowels. In phrase-final position, however, those vowels undergo
lengthening. This could then be another such case. I have not located any claim in the literature
documenting Final Resistance in Bulgarian, nor have I noticed it impressionistically in my work with the
language. Of course, neither of these facts necessarily means that it does not exist.
76 Eastern Andalusian Spanish, where Sanders (1994: 118) claims that phrase-final vowels undergo an
especially strong degree of reduction, and Pugliese dialects of Italian, where posttonic vowels reduce to
schwa but are retained, while phrase-final schwa is deleted (Loporcaro 1997: 340-341).
about the universality of the implementational aspects of final lengthening which give rise to it. The topic of reduction in final position is taken up again in section 3.6 below.

3.2.3.2. Syllable structure effects on Final Resistance

In discussion of the phonetics and phonology of final lengthening effects I have thus far alluded to, but not addressed directly the role of syllable structure in the implementation of lengthening. Looking at the typology of Final Strength effects in the phonology of vowel systems, a clear pattern emerges. In the overwhelming majority of cases in which final syllables behave as strong licensors in one way or another, Final Strength occurs only in final open syllables (which is to say, affects only absolute word-final vowels). There are very few uncontroversial cases of Final Strength effects in both open and closed final syllables, and none to my knowledge operating solely in final closed syllables, with final open syllables behaving as weak. Of the first type listed above are Hausa, Russian and Belorussian, nearly all dialects of Ukrainian where reduction exists at all, Brazilian Portuguese, Eastern Mari, Uyghur, Catalan, English, Maltese, Nawuri, Shimakonde, and Bonggi. Showing strength of both closed and open syllables were only Yakan, Murut (to the extent that this is an instance of final strength at all), and apparently the Berestja area dialect of Ukrainian (Shevelov 1979). To this can be added the Polish dialect of the city of Niemirowa nad Bugiem, in which pretonic syllables
contain only [i, a, u], while (fixed penultimate) stressed syllables and final syllables (open
and closed) contain [i, e, a, o, u] (Janiak 1995). Note however that three of four of these
examples have fixed penultimate stress, making it ambiguous whether we have resistance
to reduction of final syllables specifically due to final lengthening, or rather failure to
reduce of posttonic syllables in general (all of which happen to be final). This
cooccurrence of fixed penultimate stress and syllable-structure independent Final
Resistance may well not be accidental. In contrast to the Polish spoken in the city of
Niemirowa nad Bugiem, the local Ukrainian dialect has mobile stress and Final
Resistance to vowel reduction in open syllables only. Niemirowa is located in the
Berestja dialect region to which Shevelov attributes resistance to vowel reduction in all
final-syllables. At least this dialect, then, is an exception. It is tempting to ascribe any
Final Resistance of the type Shevelov describes in Berestja Ukrainian to the influence of
the Polish of the area which apparently also shows Final Resistance of this type, and has
penultimate stress. In the absence of further information, however, this is purely
speculative.

With respect to the syllable-structure generalization about Final Resistance
patterns, it is necessary to note that certain of the languages listed above conform to the
generalization only vacuously, insofar as they simply lack final closed syllables of any
description. Still though, the pattern is clear. It is worth noting finally that this syllable-
structure asymmetry could well be at least in part responsible for the fact that the final lengthening study of Edwards et al. (1991) found strengthening of the vowels of final unstressed syllables only. As it happened (and they in fact note), all the unstressed final syllables they measured were open, while all the stressed finals in the study were closed.

3.2.3.3. Final Strength and psycholinguistic prominence

Recall now the hypothesis sketched in 3.1.2 that Final Strength effects could be due not to phonetic realization, but to the psycholinguistic prominence of final syllables in general, a prominence which might mandate faithful realization of vowels in those syllables over that of vowels in less prominent positions (just as it did in the study of Kehoe and Stoel-Gammon on children's truncations in English). The asymmetry between closed and open syllables observed above for Final Strength effects, however, is deeply problematic for the psycholinguistic-prominence-only theory. Were it psycholinguistic status that gives rise to strength effects, we would naturally expect these effects to be distributed evenly among syllable types (as is the case in the acquisition study), since all final syllables would be equally prominent in this respect. The fact that this is not case should suggest that the proximal cause of Final Strength effects is to be found elsewhere.

From a phonetic point of view, on the other hand, the open syllable/closed syllable asymmetry is readily explicable. A number of studies, including de Jong and

163

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Zawaydeh (1999) on Jordanian Arabic, Edwards, Beckman and Fletcher 1991, Byrd 2000, and Cho 2001 remark on the propensity for vowels in phrase-final open syllables to undergo more lengthening than vowels in phrase-final closed syllables. This effect is generally attributed to the fact that degree of lengthening of a given segment seems to decrease as distance from the boundary increases. When the final syllable is closed, it is the final consonant that undergoes stronger lengthening, with the vowel lengthening somewhat less than it would have were it in absolute domain-final position. Byrd (2000) draws a connection between this effect and the results of numerous studies (see Chapter 4), including her own, in which domain-initial consonant articulations are strengthened. Specifically, within a task-dynamic model of speech production, Byrd posits a non-tract variable she calls a π-gesture, which straddles the boundary between phrases and causes a local decrease in gestural stiffness, with the effect of lengthening, and conceivably strengthening (but see Cho 2001 for discussion) the segments to either side of that boundary. The effect would be strongest on segments closest to the boundary, which would account for why domain-final vowels and domain-initial consonants are strengthened significantly, while vowels in domain-final closed syllables and vowels in domain-initial syllables with onsets undergo lengthening to a much lesser degree (or in the case of the latter, in fact, at least in English, do not do so at all. Again, see chapter 4). All this is not to say, of course, that Final Strength effects in both closed and open
syllables could never arise in a system. Indeed, the fact that vowels in final closed syllables do lengthen, if perhaps somewhat less, suggests that the phonologization of non-reductions owing to the phonetic length in both syllable types should nonetheless be possible. From the phonetic point of view, we would merely expect such systems to be relatively less common, which is clearly the case. Ultimately, the durational asymmetry itself may be found only in certain languages, with other languages implementing final lengthening differently. One study by Berkovits (1993), in fact, of Israeli Hebrew, suggests precisely this. In this study Berkovits demonstrates that vowels in Hebrew undergo lengthening of equal magnitude in both closed and open final syllables. In both final and non-final contexts, however, the vowels in the open syllables are longer than the vowels in closed syllables by a mean 11 ms.

While the phonetic features associated with final lengthening are clearly better suited to account for the patterning of Final Strength effects than an appeal to the psycholinguistic status of final syllables, I should emphasize that this does not necessarily mean that psycholinguistic status plays no role at all here. It is conceivable, for example, that psycholinguistic status is in some way responsible for the fact that final syllables are phonetically prominent in the first place (though of course given that final lengthening appears to be primarily a phrasal rather than a lexical process, this seems unlikely). Be this as it may, I claim here that the specific phonetic characteristics of segments realized
in final position which give rise to crosslinguistic regularities in the forms and distributions of final strength effects. It is therefore these phonetic characteristics to which we must turn for an explanation of those regularities.

3.3. Why Final Strength is different from Stressed Syllable Strength

At this point it is possible to return to the two general approaches to Positional Neutralization sketched and contrasted in Chapters 1 and 2. For stressed syllables a phonetically-motivated approach like Licensing-by-Cue, despite the serious failings it was shown to have in Chapter 2, does at least make the prediction that vowel reduction should occur most readily in languages in which duration is a primary cue for stress, a prediction which a purely structural approach obviously fails to make. Here too the two approaches differ in the degree of success they enjoy in accounting for the typological regularities observed in the patterns and distributions of phonological strength effects. The purely structural account of PN, which we may call the Monolithic Strength approach, again does not predict the existence of regular restrictions on the nature of strength effects in final position, nor does it predict the differences observed between the strength effects found in stressed and final syllables.77 In an approach making no explicit

---

77 Put more strongly, without further elaboration of some kind (e.g. of the type proposed by Smith 2002), it actually predicts the non-existence of those restrictions and differences.
connection to phonetic factors, one in which the Strength or Weakness of a given position
is merely the stipulated ability of the grammar (in OT, Con) to formulate Faithfulness or
Markedness constraints parametrized to that position. Strength effects are simply the
interaction of those constraints with their general equivalents. The free reranking of these
constraints in turn explicitly predicts that a given (generable) neutralization pattern will
be just as likely to surface in one Strong position as in any other. The typological facts
tell us otherwise, making the abstract grammatical approach alone a poor predictor of
crosslinguistic typology.

An approach incorporating phonetic information into the grammar, on the other
hand, such as Steriade’s Licensing-by-Cue or other direct phonetics approaches such as
that of Flemming (2001) do far better in this respect. In the sections above I have
demonstrated the striking isomorphism between the distribution of phonetic duration and
the occurrence and patterning of strength effects in final position. Now we can take up
the differences between the strength effects in stressed and final syllables and examine
the phonetic origins thereof.

As noted, while Strength effects in stressed syllables (generally but not always the
result of unstressed vowel reduction) are both common and robust in languages with
duration-cued stress, Final Strength effects seem both less common and when found often
not so unequivocal as their stressed-syllable equivalents. Specifically, Stressed-syllable
Strength effects commonly make the stressed syllable the sole licenser of the entire vowel inventory in a given word. In the examples of Final Strength given above, on the other hand, the final syllable is rarely the unique strongest licenser in the word. The strength effects found there are in fact often best characterized as exceptions to a more general pattern of neutralization (what I have termed Final Resistance above). Of course, this distinction may not hold in all cases, and is not meant as a formal or definitional statement 78. Indeed, it is often difficult to identify criteria by which to designate one strong position “primary” and another “exceptional” in this way. In Uyghur, for example, is low-vowel raising primarily an instance of initial-syllable strength, closed-syllable strength, or even bimoraic vowel strength (despite the phonetic manifestation of the vowel length distinction in Uyghur being somewhat ambivalent)?

There are a number of phonetically-based reasons why stressed syllables and final syllables manifest different types of strength effects, and why the final-syllable equivalent of unstressed vowel reduction seems not to exist. The first comes from Cho’s characterization of the gestural correlates of English final strengthening. While Cho does find sonority expansion (essentially, slight lowering) of vowels in final position, he does not note any of the featural enhancement effects that he describes for accented syllables, 78.

78 In a standard Positional Faithfulness (Beckman 1998) account, in fact, the two would be formally indistinguishable.
and which were reported by de Jong in his 1995 study characterizing accent as local hyperarticulation (affecting all dimensions of vowel articulation, such that [i] would be raised rather than lowered, while [a] would be lowered rather than raised). This would account for the resistance of final syllables to processes decreasing vowel height, but would not necessarily make all vowels perceptually more distinct (in fact, see below for the suggestion that this sonority expansion may in some cases contribute to the development of the phonological weak licensing capacity of final syllables as well), and certainly would not make them as able licensors as stressed syllables. High vowels in these syllables, for example, would be realized lower than their peripheral stressed counterparts. The resistance to coarticulation Cho describes for final vowels is clearly relevant here as well in the development of resistance to assimilatory processes as in, for example, Maltese.

The second reason why final syllables are less robust and consistent strong licensors than stressed syllables has to do with the level at which the relevant phonetic strengthening effects take place. In vowel reduction systems, there is a substantial durational asymmetry between lexically stressed and unstressed vowels, even if the stressed syllable does not ultimately carry a nuclear pitch accent. This asymmetry is thus present in all potential tokens of a given lexical form. Final lengthening, on the other hand, the phonetic source of Final Strength effects, has been shown in many if not all
languages to be strongest and most consistent before higher-level phrase boundaries (such of those of IP or Utterance). At the word level, by contrast, final lengthening is often realized inconsistently and to a much weaker degree than in phrase-final syllables. In Italian, in fact, Farnetani and Kori 1990 claim that consistent, robust final lengthening takes place only utterance finally, with the phrase-final effect being somewhat ephemeral and less consistent (a pattern which they attribute to the “syllable-timed” rhythm of the language). This means ultimately that relatively few tokens of a given lexical form will be pronounced with substantial final lengthening. While the consistency of the durational asymmetry between lexically stressed and unstressed vowels then makes this pattern an extremely likely candidate for eventual phonologization (as categorical vowel reduction), the opposite pattern for final lengthening makes phonologization less likely, with any effect to development often remaining gradient and phrase-level indefinitely (as in Uyghur).

While the degree of any positional durational asymmetry is thus at the word level greater in the stressed/unstressed pair than in the final/non-final, the nature of the asymmetry plays a role as well. In the final lengthening case, while the duration of the domain-final syllable increases in some cases quite significantly, this does not mean that the domain-internal syllables in the word are made in any way durationally deficient. On the contrary, they are as long (or short) as they would be in any other token of the same
word at the same rate of speech, such that their non-final status alone does not cause them to be realized with a duration which would in any way compromise them articulatorily or perceptually. There is even evidence from a number of studies (e.g. Berkovits 1993, Cambier-Langeveld 1999) that phrase-final lengthening often increases to some extent at least the durations of segments in the word prior to the final syllables as well, though such lengthening decreases in magnitude as distance from the boundary increases. Nonetheless, this failure of the process to single out the final syllable as the unique locus of additional durational prominence would serve to decrease the likelihood that the effects of such an asymmetry would ultimately become categorical through phonologization.

The asymmetry in the stressed/unstressed case, on the other hand, is quite different. In most systems of phonological vowel reduction, it is not merely that the stressed syllable is longer than the unstressed syllables, but rather that the unstressed syllables (or some subset thereof) are actually durationally contracted, or realized with durations sufficiently short regardless to cause articulatory undershoot (see Chapter 2 - truncation of the relevant gestures due to contracted intergestural phasing such that they fall short of their targets), or to cause certain contrasts to become difficult to perceive with accuracy. Thus, while the additional duration of final syllables may allow them to resist processes targeting non-prominent vowels, by itself it is unlikely to lead to a
situation in which the final syllable alone licenses the system's full vowel inventory, while all non-final syllables are reduced to some subset thereof. Indeed, even Hausa (§3.2.2.1) is not unambiguously such a system, with the phrase-internal neutralizations targeting specifically the short vowels, which by all accounts are by nature short and variable enough to overlap significantly in casual speech if not subjected to some additional lengthening. We might then expect other such systems to arise in languages with non-duration-cued accentual systems and a vowel quantity distinction such as that found frequently in Indo-Iranian languages like Kurdish or Punjabi, in which the short vowels are strongly centralized and variable\textsuperscript{79}.

3.4. Categorical and gradient effects in Final Strength systems

The preceding sections have presented arguments for viewing the typology of final strength effects as a consequence of the peculiarities of the implementation of final lengthening. The phonetics of final lengthening were shown to account for the fact that final strength effects are often found at the level of the phrase rather than the word, that final syllables are rarely if ever the sole strong licensers of vowel contrasts in the word domain, and that Final Resistance in the vast majority of cases targets only the vowels of

\textsuperscript{79} Such as found in the Volga region Turkic and Uralic languages and Indo-Iranian languages like Kurdish.
final open syllables, rather than final syllables of any shape. A Pure Prominence approach to PN effects in final position predicts none of this.

In this section I present a sketch of how the phonetic details treated above are transformed into phonological patterns through the medium of phonologization. In many ways the process is similar to that described for stressed syllables in Chapter 2, though here additional details (including interaction with the phonologization of UVR) make this process worth additional attention.

Just as with the UVR, final strength effects start off gradient, and become categorical through the process of phonologization. This is even clearer, in fact, in the case of final syllables, where the phonetic roots of the effect are most robustly observed in the gradient application of phrase-final lengthening. A Final Resistance pattern might thus emerge in the following way (taking unstressed vowel reduction to be the operative neutralization pattern. See chapter 2 for background).

The first stage in the development of a gradient vowel reduction process yields a system like that of Bulgarian: unstressed vowels are realized with short durations. The degree to which reduction takes place is a function of vowel duration. Assuming a robust-enough implementation of phrase-final lengthening, gradient phrase-final resistance à la Russian or Brazilian Portuguese is present from the outset of the development of vowel
reduction, emerging automatically as a consequence of the duration-dependent reduction of vowels.

The second stage is marked by the phonologization of part or all of the word-internal reduction pattern. In Uyghur, for example, word-internal vowel raising is already categorical. It occurs in that position whenever its structural description is met, regardless of the physical duration of the segments involved (as evidenced by the raising of new long vowels from compensatory lengthening). In word-final position, however, the process is still gradient and rate-dependent, taking place when the vowel is shorter (phrase internally or in fast speech), and meeting resistance when the vowel is longer (phrase-final position or in careful speech). In such a system, then, the internal categorical reduction pattern is an innovation, while the final-syllable gradient pattern is a relic from a previous system. It is appropriate at this point to wonder why it should be that the reduction effect is phonologized only word-internally, leaving a gradient effect in the final syllable, while the reverse seems not to occur. As noted above, phonologization of an effect is most likely where that effect is robust and consistent. This is the case for internal unstressed syllables in the immediate predecessor to a system like Uyghur. Within a given speech style or tempo (to the extent that such an entity exists

---

80 Referring of course only to the Degree Two gradient reduction patterns of these latter, the Degree One patterns being categorical in both.
monolithically over time in connected speech), word-internal unstressed vowels will be subject to relatively little durational perturbation, comparatively free as they are from the effects of processes like boundary-adjacent strengthening. Word-final vowels, on the other hand, will undergo substantial durational variation depending on the position of the word within the phrase, meaning essentially on the degree of application of the final lengthening effect. The resulting inconsistency of application of the gradient vowel reduction process to vowels in word-final position will hinder the process of phonologization of reduction in that position, giving it a crosslinguistic tendency to remain gradient there long after reduction is phonologized in other positions within the word. Note that in Uyghur or Nawuri, the fact that the final syllable undergoes the reduction or assimilation process only gradiently means that the failure of the categorical version of that process to apply must be specifically mentioned in the phonology.

This explicit mention of certain structural positions as apparent arbitrary exceptions to a process is something that proponents of a direct approach to phonetics such as Kirchner argue is undesirable, and a primary reason for allowing the phonology to refer directly to the phonetic features that unify all exceptional positions as a natural class of sorts. In Uyghur then, if we could just say that raising applies to all vowels under X ms of duration (formalized as e.g. Kirchner's [+ partially long]), we would avoid having to arbitrarily stipulate that long vowels, initial vowels, closed syllable vowels and
phrase-final vowels all fail to undergo the process, since all these could be marked [+partially long]. This approach does simplify the description of the process, and brings its original phonetic motivation directly into its description, but in so doing it oversimplifies the specific characteristics of the implementation of the process, and thus obscures its true nature. As noted above, in Uyghur all underlying long vowels fail to undergo the raising process, despite the fact that they durational distinction between these and underlying short vowels has been largely effaced with time. Furthermore, the process *does* apply to new long vowels created by compensatory lengthening, despite the fact that they must clearly be over whatever subphonemic durational threshold we take [+partially long] to encode. In other words, the process has been phonologized. It applies categorically word-internally, without regard for matters of phonetic duration. The exceptional final syllables, on the other hand, do not share this property. They are said to optionally undergo raising non-prepausally, which I take to mean that raising applies to them only in such case as they actually are under some minimum durational threshold, which is to say, it applies gradiently. We are dealing, in other words, with two distinct processes, one of which is responsive to the realization of phonetic duration, the other of which is not. That the latter includes a list of arbitrary-seeming exceptions to its application (in whatever form) is not a shortcoming. Rather, it is a *virtue*, in that
synchronously those exceptions really are arbitrary. The relevant structures are exceptional regardless of their precise phonetic realizations.

It should be noted at this point that, as with vowel reduction, phonologization of the process here refers only to the introduction of a new, independent target realization for the vowel in the position of reduction. Prior to phonologization, the vowel would have shared the same target window as its unreduced counterpart, with the caveat that that window would be modeled as sufficiently wide to allow substantial deviation under the influence of durational pressures (in Keating 1996’s terms, meaning essentially that the need to comply with the dictates of durational pressures and restrictions on articulatory effort are weighted more heavily than the need to attain an optimal target height). The introduction of an independent target means that the process is no longer contingent on duration or other factors altering the realization of the original target. Reduction now takes place regardless of phonetic circumstances, assuming that the structural description of its application is met (e.g. the vowel is in a phonologically unstressed syllable). The process thus far may correspond to Hyman’s stage 2 of phonologization, in which the phonetic pattern is extrinsic, but subphonemic. Whether or not the addition of the independent target for the unstressed vowel has also resulted in the categorical neutralization of phonological contrasts (i.e. the proximity of the realizations of two phonologically distinct vowels in unstressed syllables has led to the substitution of a
single target for both, and with this the merger of the two phonological categories into one in this position) is another matter altogether.

In Stage 3, the status of the final syllable is set through phonologization. Resistance to phonologization notwithstanding, it is still possible for the final syllable to follow its word internal counterparts in becoming phonologized. This stage in the developmental path of final strength is in some ways the most difficult to find incontrovertible examples of for (at least) two reasons: Without experimental evidence (of which there is a limited supply on this question for languages which are neither Romance nor Germanic), it not possible to be certain that the status of the final syllable is truly phonologized, despite assertions or implications found in descriptive studies. In this respect the assertion that a process is not phonologized, or remains gradient, is far more indicative and likely to be correct, variation perhaps being more readily perceptible to impressionistic evaluation than the lack thereof. The second problem is that even if a process is shown to be phonologized in the present, though we can suppose that it must have passed through some gradient stage or stages to arrive at that point, there is rarely clear evidence that this was actually so.

Nonetheless, one possibility is still for the pattern of Final Resistance to become phonologized. This amounts to the generalization of the least reduced variant of the final-syllable vowel (that which appeared in phrase-final lengthening contexts), meaning in
effect the expiration of the vestigial gradient reduction pattern altogether, even in its last
refuge, the domain-final syllable. In the window model, gradient duration-dependent
reduction is to be modeled as a wide window for the realization of the unstressed vowels,
expressing the looser specification of the height target for the vowel, such that additional
factors such as decreased duration might alter the vowel’s realization significantly. The
phonologization of that reduction meant the introduction of a new, independent target
window for the unstressed vowel, replacing the widened one of the stressed vowel. The
phonologization of final resistance, then, must be seen as a re-narrowing of the target
window for the vowel in question in those unstressed syllables which, like the final, have
not already received independent targets through the phonologization of the reduction
itself. The vowel there is now realized as a largely unreduced instantiation of its
phonological specification regardless of the actual duration of the final syllable.

Assuming an earlier stage of the English system in which final /i/, /o/ and /u/ were
left unscathed by the categorical phonology (did not become phonological schwa as did
their word-internal preconsonantal counterparts), but were subject to gradient qualitative
reduction in the phonetics, the modern situation in which /i/, /o/ and /u/ remain relatively
peripheral in final open syllables is the result of the suspension of that gradient reduction
pattern, which at a still-earlier point must have been operating generally throughout the
It is difficult to tell without experimental evidence or a direct positive statement in a description whether a final resistance effect is truly phonologized in this way at the phrase-level (or the word- and phrase-levels, depending). Eastern Mari, however, appears to present a relatively case for being just such a system.

The other option for phonologization of the reduction status of the final vowel or syllable is the generalization of the variants in the gradient system with the greatest amount of reduction (the phrase-intemal or casual speech realizations). This has the effect of eliminating the Final Strength effect entirely. This is the situation for those speakers of Brazilian Portuguese whom Major (1985) describes as having lexicalized the reduction of word-final mid vowels such that it applies even under conditions of final lengthening. For such speakers presumably all posttonic vowels are categorically reduced in this fashion, with finals simply ceasing to be an exception.

The sections above provide a phonetic explanation for the observed differences between phonological strength effects found in stressed and final syllables, and provide a

---

81 This account is based on the synchronic description of Hammond 1999, in which unstressed word-intemal preconsonantal syllables are said to license only the reduced vowel [a], while unstressed open final syllables permit [i, u, o, a]. There is the possibility, however, that this account, and by extension the diachronic scenario sketched above are irretrievably flawed in the following sense: There is a no small circularity in Hammond’s account in the reckoning of unstressed vowel distributions. These are determined by the fact that the only [o], [u] or [i] found in word-intemal preconsonantal positions are in syllables said to bear secondary stress (e.g. [jʌɪ’tʌnd], [ˌmoʊˈtʌkt], etc.). To the best of my knowledge, however, the fact that these syllables bear lexical secondary stress is determined largely if not solely on the basis of the failure of the vowels in these syllables to undergo vowel reduction. Impressionistically at least, these vowels actually seem to undergo a fair bit of gradient reduction nonetheless in ordinary speech, suggesting that the situation in English is not yet fully phonologized in the first place.

180
unified phonologization scenario accounting for the development of a range of systems of Final Strength effects, relating them one to another in the process. But we have yet to discuss what is perhaps the most crucial factor in accounting for the ambiguous status of final position within the typology of Positional Neutralization. This is the alarming (in terms of the discussion so far) propensity for phonetic and phonological weakening to occur in those very same syllables, whose propensity for strengthening we have just reviewed in some detail.

3.5. On assumptions concerning licensing capacity and phonetic prominence

All the instances of phonological final resistance seen in the catalogue in (20) above (§3.2.3) are cases in which more than one position within the word fails to undergo some reduction or assimilation process, and in each case the exceptional positions have also shared the characteristic of heavy relatively longer phonetic durations than the reducing or assimilating positions. The case of Brazilian Portuguese, for speakers for whom both pre- and posttonic reduction is phonologized, even has three distinct levels of licensing capacity (stressed, pretonic and posttonic), which correspond directly to three characteristic degrees of phonetic length. The case of the Pasiego Spanish tense/lax contrast was presented above as evidence of the danger present in assuming automatically that the ability to license additional contrasts in a system must be correlated with an
increased amount of phonetic duration characteristically realized in the relevant position.

Another example, that of Timugon Murut, illustrates a similar danger in the assumption, even when duration is involved in determining licensing capacity in a system, that the relationship between these two is a simple direct proportion.

As outlined in 3.2.3-n. above, Timugon Murut, an Austronesian language of Sabah in Borneo (Prentice 1971, Kroeger 1992), has four contrastive vowel qualities: /i, a, o, u/. These all only contrast, however, in the (fixed penultimate) stressed syllable and the posttonic final. Steriade notes of Murut, however, that “the contrast of rounding among non-high vowels occurs also - in a more limited fashion - in the stressed syllable, which occurs in penult position. Thus final [o] contrasts with [a] in words of any shape, whereas penult [o] contrasts with [a] only in words in which the final is an [o] (Steriade 1994: 13). This is illustrated in (21)

(21) Distribution of non-high vowels in Timugon Murut

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
</table>
a. | tanom | ‘plant’  |
    | baloy | ‘house’   |
    | limog | ‘dew’     |
    | iloŋ  | ‘look!’    |

182
As for the non-high vowels in pretonic position, here there is no contrast between /a/ and /o/. Their realizations can be summarized as follows: where the tonic and posttonic are /o/, all contiguous pretonic non-highs vowels are realized as [o]. Where the tonic and posttonic are not [o] (or the pretonic non-high vowel is separated from these by a high vowel), pretonic high vowels are realized as /a/. Kroeger (1992) follows Prentice in seeing restrictions on the realization of [o], and also a right-to-left harmony pattern triggered by the combined presence of [o] in the tonic and posttonic syllable, a rounding harmony with a bisyllabic trigger rather like that treated by Walker (to appear) for Classical Manchu and Tungusic. As seen in (a) above, [round] does not spread from the posttonic to the tonic. The harmony pattern is shown in (22), where in (a) and (b), addition of a suffix without [o] causes preceding high vowels to become [a], and in (c) the leftward spread of [round] from the stressed syllable and posttonic.

(22) Rounding harmony with a bisyllabic trigger in Timugon Murut

a. orop + an -> arapan 'perch' (Referent Focus)
ongoy + an -> angayan 'go' (RF)
in + abot + an -> inabatan 'belt' (RF, past tense)
Since the final syllable is the most unrestricted licenser in the word in Murut, it is tempting to imagine that this perhaps owes to some particularly outstanding phonetic prominence, a prominence which allows it to excel even the stressed syllable in licensing capacity. This, however, would in all likelihood be incorrect.

I have no instrumental data on Timugon Murut from which to determine the precise prosodic characteristics of the final syllable, but a diachronic approach to the problem makes clear both the source of the current distribution of vowels in the language and the prominence relationships which must have led to them. Two preliminary facts are necessary to understanding the situation: First, the vowel transcribed ‘o’ in Murut (and many related neighboring languages) is generally not realized as IPA [o]. In Timugon Murut this vowel is [o] only before [w] (Prentice 1971: 19). The base symbol used by Prentice in phonetic transcription, [ə], is also somewhat misleading, as the vowel is realized with this quality only before velar consonants in closed syllables. The elsewhere
realization of ‘o’ in this language is apparently something more like [ø], what Prentice describes as a “voiced lower-mid central half-rounded vocoid” (1971: 19). Realization of this vowel also varies significantly among the Murutic, Dusunic, and Paitanic languages of the area (Boutin and Pekkanen 1993). In Kimaragang (Dusunic, Kroeger 1992), it is realized as [ɔ] under stress, while other Dusunic languages have it as “a back unrounded or only slightly rounded vowel, roughly [ɤ], with considerable tensing of the tongue back” (Kroeger 1992: 280). Kroeger even refers to it as “the neutral vowel” in Kimaragang. None of this is unexpected, though, in light of a second consideration: the historical source of /o/ in most cases in the languages in question is *PAN /a/ (Robert Blust, p.c.).

While the system of realizations of the pretonic vowels differ significantly among some of the languages in question, Timugon Murut shares with Dusunic Kimaragang its distribution of [o] in the penultimate and final syllables (Kroeber 1992). This is the older part of the system, the pretonic harmony being secondary. It is the development of this part of the system, additionally, which gives the lie to the assumption that the final is in some way prosodically stronger than the tonic, or that rounding has some need to be anchored to the word edge (as Walker argues for Oroqen). Facts of the emergence of the distribution of [o] in these positions are the following (*PAN and *Proto-Malayo-Polynesian reconstructions from Blust 1999, 2000, Dusunic Tatana’ forms from

(23) *PAN /a/ in the Murutic and Dusunic final syllables

- *PMP tanəm > TM tanom ‘plant’
- *PAN qaNəb > Tatana’ aŋob ‘door’
- *PAN Sipəs > Tatana’ lipos ‘cockroach’
- *PMP nipən > Tatana’ dipon ‘tooth’

Given that the vowel in question is derived from the overshort *PAN /a/ along with what we know about the realizations of this vowel in the modern languages, this should occasion no surprise, and certainly does not require the assumption of special prosodic prominence in the final. The opposite, in fact, may be more plausible, in light of the fact that in the stressed penult, *PAN /a/ becomes TM [a]. This development, central to the formation of the licensing asymmetry described above, should actually be understood as a strengthening of schwa under stress, making its realization lower and longer than in the weaker final syllable. Several examples of changes of this sort, in Cho’s (2001) phonetic sonority enhancement and the Positional Augmentation of Smith (2002), have already been exemplified in the previous chapter, and more come in the discussion of final lowering in 3.6. The heart of the strong licensing capacity of final, then, originates in the fact that it is prosodically weaker than the tonic, not stronger. The
dependent licensing of [o] in the tonic is the result of an exception to this pattern, whereby *PAN /ə/ > TM [ʊ] if the penult also contains [o]. While such "support" phenomena commonly account for the failure of certain vowels to reduce (see discussion of harmony and reduction in chapter 4), it is less usual that a vowel should resist *strengthening due to the presence of an adjacent vowel. Nonetheless, it must be concluded that the presence of adjacent [o] in the penult is enhancing the perception of stressed [o], preventing it from being realized low and unround. Indeed, the combination of the fact that the final syllable is comparatively weaker prosodically than the tonic, and the fact that whatever rounding was present on the reflex of schwa cannot have been very great (given its realizations in the daughter languages) makes it tempting to see the emergence of stressed [o] as an umlaut- or metaphony-like change. In such a scenario, the already-weak rounding, being realized in a prosodically weak position, is heard instead on the stressed non-high vowel, and reinterpretation occurs. See the discussion of the motivations for metaphony above in the section on Pasiego Spanish for various takes on this phenomenon. If this is the case for Murutic/Dusunic, though, then the entire limited licensing of [o] in the stressed syllable is the result of the prosodic weakness of the final, and cannot be considered a final strength effect. Examples follow in (24):
(24) *PAN /a/ in Murutic/Dusunic stressed syllables

a.  *PAN talu > talu ‘three’
    *PAN Sapat > Kadazan/Dusun apat ‘four’

b.  *PMP anam > onom ‘six’
    *PMP damdam > Tatana’ rondom ‘dark’
    *PAN gem ꞉ > gomgoŋ ‘fist’
    *PMP depa > lopo ‘fathom’

The realization of */a/ as [a] in the stressed syllable can be seen above in (21) (e.g. tanom - to plant). Adding to the asymmetry, however, */a/ is also realized as TM [o] in final syllables of a certain shape. Since the most regularly attested of these seems to be word-final /a/ (and since a > [o] here makes little sense as a final strength effect), this appears to be the result of a final reduction process such as those discussed in 3.6.5.2, a hypothesis strengthened by Kroeger’s (1993: 38) observation that in Kimaragang, there are few vowel-final words, and what final non-high vowels ([a], [o]) there exist are realized with a “slight breathiness or aspiration”. The regularity of the change in some of these environments (e.g. /_y#) is still uncertain, and /_h# could simply be the result of an earlier loss of final [h], leaving /a/ final:

82 From Blust (1988), meaning ‘grasp in the fist’.
(25)  *PAN /a/ > [o] in Timugon Murut posttonic syllables

a. /_# *lima > limo ‘five’
   *duSa > duo ‘two’
   *kita > kito ‘see’
   *tuba > tuo ‘fish poison’
   *mata > mato ‘eye’

b. /_j# *PMP m-atay ‘to die’ > patoy ‘die’
   *PMP sakay-an ‘vehicle, ride in’ > sakoy ‘mount, ride’

c. /_h# *qumah > umo ‘cultivated field’

As for the pretonic syllables, the right-to-left harmony exhibited in Timugon Murut makes most sense in light of facts from other related languages, such as Kimaragang, in which Kroeger notes that pretonic /a/ and /o/ both “tend to reduce to schwa in normal speech (1993: 38). In Timugon Murut, then, this reduction must also have been present. Where the stressed vowel was [o], preceding [a] would have assimilated in rounding to the tonic vowel, while in the presence of other tonic vowels, it remained [a]. At some point in its later development, these reduced pretonics must then have been lengthened, replacing the reduced pretonic non-high vowels with the unreduced variants found there today. Where the schwa was rounded, the result was [o], and where it was not, the result was [a]. In this account, rounding simply spreads from a round tonic vowel to preceding reduced vowels, an unremarkable scenario from the point of view of vowel-to-vowel coarticulation. The appearance of a “bisyllabic trigger” is purely accidental. Only a stressed [o] is necessary, but this stressed [o] happens only to
be found as the product of the metaphonic assimilation from final [o], as described above.

Note that with respect to the behavior of pretonic syllables, the licensing of [o] in finals 
deeply depends on some additional phonetic prominence. Despite the weakness of this position relative to the stressed syllable described above, the final must nonetheless have some additional duration resulting from final lengthening such that it resists reduction the reduction to schwa which takes place in pretonic syllables. There is, therefore, a final strength effect in Timugon Murut, but only relative to these reduced pretonics. The remainder of the distributional restrictions on [o] follow entirely from other factors. The following illustrates the development of some of the alternations discussed above:

(26) Rounding and Unrounding in Timugon Murut

*tanəm > tánəm
stressed /a/ > [a], final schwa > [o]

/tanom - in/ > tanámin ‘plant (Referent Focus)
stressed */a/ > [a], pretonic non-high > *[ə] > [a]

/tanom -on/ > tonómon ‘plant (Object Focus)
stressed */a/ > [o]/Co(C)#, pretonic non-high > *[ə] > [o]

3.6. Final Weakening

The weakening, often leading to loss, of word-final segmental material is well-known to historical linguists and synchronic phonologists alike, far more so no doubt than
the final strengthening trends discussed above. One sees repeated without qualification in study after study, in fact, the claim that final position is “weak”, usually in the sense that the speaker deigns to spend less “articulatory effort” in this position. The eminent Russian phonetician Shcherba, writing in 1912, refers to “a general flaccidity\(^3\) of the articulatory organs... coming on toward the end of the word”. Gauthiot’s 1913 monograph, *Le Fin de Mot en Indo-Europeenne*, is an encyclopedic treatment of the phonological history of final syllables in Indo-European; a narrative of what fell off where and when. More recently Hock (1999), in an article on phonological processes in phrase-final position concludes that final position is a monolithically “weakening” environment, deriving a number of distinct phenomena found there from this general tendency (i.e. final devoicing, stress retraction, vowel loss)\(^4\). This conclusion is seriously at odds with the findings catalogued above.

For consonants, previous work suggests a solution to this contradiction. Certainly the study of Keating, Wright and Zhang (1999) shows a marked strengthening effect for consonants in phrase-final position in English, such that their linguopalatal contact

---

\(^3\) *v’alost*, also perhaps ‘withering’, ‘limpness’ or ‘languour’.

\(^4\) An important aspect of Hock’s conclusions, though, is on the need to distinguish phrase-final (in his terms “Finality”) effects from word-final (“Right Edge”), in that in phrase-final position there is often phonetic motivation for weakening (e.g. for final devoicing, see below), while in word-final but phrase-internal position that motivation may be lacking. This means that weakening processes common to both environments almost certainly begin in the larger domain and are generalized to the smaller. This parallels what we have seen for final lengthening above.
becomes in some cases equal to that of comparable word-initial consonants in the same phrasal position, a pattern standing in direct opposition to the idea of phrase-final articulatory weakness. On the other hand, Steriade (1997) analyzes the tendency for a variety of consonantal contrasts to be neutralized in word-final position (and elsewhere) as stemming from a lack of robust cues to consonant identity in that environment. Specifically, lack of CV transitions and release burst obscures cues to consonant place and laryngeal specification. We may add to this the sharp amplitude drop extremely common if not universal in most phrase- or utterance-final syllables, and along with that the tendency to phrase-final voicelessness cited by Hock, and what emerges is a picture of the extreme perceptual weakness of consonants in final position, despite whatever articulatory strength they may exhibit simultaneously. Regardless of degree of linguopalatal contact or strengthening of any other supralaryngeal articulations, the perceptual difficulties associated with final consonants would suffice to engender such changes as neutralization of laryngeal features (voicing, aspiration, etc.), neutralization of place features, or even ultimately debuccalization and loss. In many instances, then, perceptual factors could ultimately win out in determining the fate of the final consonant even if the type of articulatory strengthening documented by Keating, Wright and Zhang were operative.

---

85 Results are for the consonants /t, d, n, l/. The effect was weaker for /l/ than for the other consonants.
Turning now to vowels, I will propose a similar explanation of the contradictory phonological trends encountered in final position. Before turning to the relevant vowel-quality neutralization patterns, I will review a number of other phonetic tendencies commonly found in domain-final position. These are relevant here in that, while they are natural concomitants of the phonetic final lengthening described above, their effects on the perceptual robustness of vowel contrasts realized in final syllables are the opposite of those of final lengthening and strengthening. To wit, these phonetic patterns have the effect ultimately of obscuring the realizations of vowel contrasts in final position. This phonetic anti-prominence, in turn, can not only counteract the prominence enhancing effects of final lengthening, but can even result in the phonologization of neutralizations taking place specifically in domain-final syllables. It is thus these phonetic factors which are responsible for the phonological ambiguity of final syllables with respect to patterns of positional neutralization. Because final syllables have received little attention in the context of PN in previous literature, and because even the phonetic literature is rather one-dimensional, focused primarily on phrase-final lengthening, the following sections have as an ancillary purpose the documentation in as much detail as possible of the crosslinguistic range of these patterns. Each section will therefore be quite extensive.
3.6.1. Devoicing

The first such effect to be discussed, touched upon already above with reference to Hock’s Weakening theory, is the tendency to phrase-final voicelessness. The significance of this effect for consonants is commonly known, so I shall limit the discussion here to the consequences of this effect for the realization of vowels. Domain-final devoicing of vowels is extremely common crosslinguistically. It ranges from the radically low-level, sporadic effect rarely even mentioned in descriptions of languages where one finds it such as, for example, Russian, to the heavily morphologized devoicing patterns in the “utterance final” forms of Oneida analyzed by Michelson (1999). It often figures in descriptions as the addition of aspiration, or the epenthesis of final /h/ (as in Sapir and Hoijer’s 1967 description of Navaho). It is also a common phonetic source of final vowel deletion, as can be seen in the Austronesian languages discussed in Blevins (1997) and Blevins and Garrett (1999).

Final vowel devoicing is generally assumed to have its phonetic roots in the steep drop in subglottal pressure associated in most languages with the ends of phrases or utterances along with final lowering of F0 (Dauer 1980, Gordon 1998). This decreased subglottal pressure can result (particularly with a close supralaryngeal constriction downstream) in the elimination of the pressure drop across the glottis necessary for voicing to be maintained. In this original form we may conceive of it as a form of passive
devoicing (as discussed in e.g. Ladefoged and Maddieson 1996: 49), in which devoicing occurs as a consequence of some other phonetic process, rather than as a result of the planned implementation of, e.g. glottal abduction.

Gordon (1998) presents a crosslinguistic survey of vowel devoicing in both internal and final positions, in which he makes a number of generalizations of relevance to this discussion. The first is that in nearly all cases the presence in a language of word-internal vowel devoicing implies the presence of final vowel devoicing, though the opposite is not true. Gordon cites four apparent exceptions to this generalization, noting that in three of them (Turkish, Azerbaijani, and Montreal French) stress is fixed on the final syllable. We will return to the topic of the devoicing of stressed vowels later in this section. A second generalization which emerges from Gordon's survey and is mirrored in my own is that the devoicing of word-final vowels in a language implies the devoicing of phrase-final vowels, while the reverse is not true: there are many languages in which only phrase-final vowels are subject to devoicing (and deletion, most likely a later consequence). This pattern falls out directly from the phonetic facts discussed above concerning the regular extreme lowering of subglottal pressure in phrase- (but not word-)

---

86 Stressed vowels rarely devoice crosslinguistically, presumably due to the various language-specific combinations of increases in F0, amplitude, or duration associated with stressed vowels. It should be noted that Turkish does not actually constitute an exception in this regard. Word-final stressed high vowels in Turkish are in fact frequently realized voiceless, particularly when the preceding consonant is a (phonetically) aspirated obstruent (Orhan Orgun, p.c.). In Turkish, stress is cued by higher F0. Amplitude and duration increases are not correlates of final stress (Konrot 1981).
final syllables. From a diachronic point of view, this almost certainly means (as it does for word-final consonant devoicing), that the pattern gets its start in phrase-final position, and is extended only subsequently to phrase-internal word-final vowels. In this respect devoicing mirrors the final strength effects discussed above in following what is the usual view of the development of word-edge phonology in traditional historical linguistics.

From this diachronic scenario an important prediction concerning both the strengthening and devoicing patterns emerges. Gordon notes in his description of final devoicing that just as with final strengthening above, in a great number of cases the effect is optional or gradient, with degree of devoicing occurring along a continuum from fully voiced to fully devoiced. Any single instance of devoicing would then be a function of other phonetic factors, such as emphasis, preceding consonant, speech rate, and the like. One the other hand, many instances of vowel devoicing are clearly phonological, due either to surface contrast with voiced vowels or interaction with other processes in the system which are believed to be phonological (Gordon 1998: 96). The prediction then is this: Where final vowel devoicing has been generalized from the position in which it is phonetically-motivated (phrase-final) to a position in which it is not (or perhaps less) phonetically-motivated (word-final), it should no longer retain the gradient character it may have had in the beginning, which is to say, it must be phonologized. Here the term "phonologization" must be understood specifically in the sense of the Hyman (1976),

196
where it refers to the transition of a given phonetic feature from *intrinsic* to *extrinsic* implementation in a given environment, but does not necessarily imply phonemicization. Interestingly, in the case of voiceless vowels, as Gordon notes, the phonemicization of the contrast is certainly logically possible, though there seem to be no completely uncontroversial instances of its occurrence. Again, in its intrinsic state, a phonetic feature is realized automatically, as a function of the presence of another feature, and in this sense need not be considered intentional (or commanded, as Hyman puts it). In the latter case the same feature is implemented in such a way or to such a degree that it can no longer be considered an automatic consequence of something else in the speech signal. Rather, it must be realized intentionally, a planned characteristic of the target articulation.

At this point it is necessary to underscore a significant difference in the operation of the phonologization process as it is conceived here in the window model. In all previous discussions of the realization of vowel height, with the exception of schwa, unreduced, gradiently reduced and categorically reduced vowels have all had some target window specified for the realization of their height features. As proposed by Keating

---

87 In Turkana, one of the strongest cases, for example, voiced and voiceless vowels are certainly contrasted on the surface in final syllables. Dimmendaal and Breedveld (1986) ultimately argue, however, that this contrast is a derived effect stemming from underlying lexical tone specifications. The analysis is somewhat abstract, perhaps, but clear enough.

88 Hyman’s example is the relatively minor FO perturbations in the onsets of vowels caused by the voicing of the preceding consonant where tone is not contrastive. Such automatic features can be used as cues for the phonological distinction of which they are a direct function, but this is a distinct issue.
(1996), the narrowing or widening of target windows allows us to model various degrees of phonetic underspecification, rather than conceiving of it as the all or nothing presence or absence of a specified target for a given dimension of a segment's phonetic realization. The discussion of vowels herein has thus availed itself of this capacity, assuming in all cases that vowels, save a putatively-targetless schwa, bear some spectral target specification or another, whether it be one with a wide window, indicating greater variation in response to the pressures of duration, or a narrower one, indicating relative stability despite such pressures. In discussion of the transition from passive (contingent) devoicing to active (planned) devoicing, the point of phonologization will certainly entail the addition of a target (i.e. a glottal abduction gesture) where none was previously found. This is not, however, to suggest that the gradient devoicing in the pre-phonologization state was entirely automatic, subject to no control whatsoever on the part of the speaker. There is obviously a tremendous amount of variation in just how much passive final vowel devoicing a given language will tolerate, which variation must be accounted for through the narrowing or widening of existing target specifications, such as those controlling subglottal pressure and the durations of the laryngeal states associated with final vowels. Devoicing in this system may not be specified, but voicing obviously must be.
The prediction that for final vowel devoicing to generalize beyond the phrase-final domain, it must first (or simultaneously) have been phonologized is more difficult to verify, since the vast majority of the languages involved have not been subjected to the kind of phonetic analysis necessary to determine such a thing (e.g. laryngoscopy, airflow measurement, etc.). The majority of descriptive sources for their part mention presence or absence of devoicing in a single sentence often without qualifications as to degree or domain of application. In addition, the situation is complicated somewhat by the variety of phonetic circumstances which can lead to vowel devoicing. For instance, while word-level final vowel devoicing can be considered phonetically unnatural or unmotivated in that this position is not generally associated with the degree of subglottal pressure decrease found finally in larger domains, other phonetic properties specific to word-final position might actually enhance the likelihood or degree of final vowel devoicing. It is well-known, for example, that shorter vowels are more likely to devoice than longer vowels (presumably because the duration and robustness of the necessary vocal-fold adduction gesture is sufficient to prevent longer vowels being overlapped by neighboring voiceless segments or falling victim to complete devoicing as a result of decreased subglottal pressure). Word-final vowels, though perhaps produced without the phrase-final drop in subglottal pressure, are also produced without phrase-final lengthening. Without this lengthening and a following phrase boundary, in many languages word-final
vowels can actually be among the shortest possible in the system, making them susceptible to elision in hiatus contexts or interconsonantal devoicing of the same sort discussed by Gordon for word-internal vowels. In Major's study of Brazilian Portuguese timing patterns (1981), di- and trisyllabic nonsense forms of the segmental shape [lala] or [lalala] situated in phrase medial contexts were realized (for a single speaker) with the following durational patterns:

(27) Mean vowel durations in Brazilian Portuguese di- and trisyllables (Major 1981)

<table>
<thead>
<tr>
<th>Token</th>
<th>Vowel Durations in ms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>v1</td>
</tr>
<tr>
<td>a. /lala/</td>
<td>151</td>
</tr>
<tr>
<td>/lála/</td>
<td>217</td>
</tr>
<tr>
<td>b. /lalala/</td>
<td>138</td>
</tr>
<tr>
<td>/lála/</td>
<td>208</td>
</tr>
</tbody>
</table>

The forms in (a.) demonstrate that posttonic vowels are shorter than pretonic in disyllables, while the forms in (b.) show both this asymmetry, and that the vowel of the second posttonic (word-final open syllable) is substantially shorter even than the vowel of first posttonic. In Bulgarian, similarly, all posttonic syllables are characterized by durational reduction of up to %70 in comparison with the corresponding stressed vowels (Savitska and Bojadzhiev 1988: 52), with unstressed final vowels (especially, it seems, [u] < /o/, /u/) often barely perceptible if articulated at all, even in forms such as the
complementizer /kato/ (frequently realized in colloquial speech and even probably lexicalized dialectally as proclitic [kat]), which due to syntactic restrictions are excluded from ever appearing in phrase-final contexts in the first place. It would not, therefore, be surprisingly if these vowels were seen to undergo a gradient devoicing process completely independent of any concerns regarding domain-final subglottal pressure, in which, for example, unstressed word-final high vowels following a voiceless consonant phrase-medially before a voiceless-obstruent-initial word were to devoice completely, while variations of these parameters would induce lesser degrees of devoicing. The existence of gradient devoicing patterns phrase-internally is thus not in and of itself evidence that the phonologization scenario and prediction given above are incorrect.

That said, I know of no clear instances of counterexamples to the prediction. Certain devoicing patterns not mentioned by Gordon, such as that of the Cushitic language Dasenech (Sasse 1976), would be crucial test cases in this regard. In Dasenech non-pre-pausal final vowels are said often to be voiceless or even to delete in fast speech, while in pre-pausal position the same vowels are retained in all cases, though realized with final "aspiration". If the internal process is not gradient, this could be an instance of a case analogous to some of the "final clipping" systems described below, in which short final vowels devoice completely and long vowels are devoiced partially, potentially phonologizing devoicing of a more or less set vocalic span, regardless of the total
duration of the vowel in question, a mora’s worth in Hayward’s terms (1998). If the process is gradient, however, as it may be, though the conditions for its occurrence are not elaborated in the description, we may perhaps suspect the circumstances described above, though without additional data is impossible to know.

Returning now to Gordon’s observation concerning stress and final vowel devoicing, findings from my survey support the claim that in general stress impedes vowel devoicing (Turkish notwithstanding), and that no language devoices stressed vowels but not unstressed vowels. In fact, there are cases in which final vowels which would otherwise devoice are realized fully voiced when they receive some sort of secondary stress or emphasis: Blevins and Garrett (1998) discuss the cases of Rotuman and Kwara’ae, in which secondary stresses placed on final vowels in marked speech styles save them from devoicing or deletion. Additionally, Sapir notes for Southern Paiute a type of “rhetorical emphasis” sometimes placed on words which allows final vowels to avoid devoicing, to be realized instead “lengthened and generally followed by glottal stop” (see below for discussion of domain-final glottalization).

This does not mean that stressed vowels cannot be realized with devoicing. Gordon mentions one such language, Papago. To this we may add as an example Goajiro (Arawakan) as described by Mansen 1967, in which unstressed final short vowels are devoiced completely, but stressed final short vowels are lengthened and only partially
devoiced (as in the clipping patterns mentioned above). At least one case, however, stands out as potentially problematic: In Afar, as described by Bliese (1976), utterance-final stressed open syllables are realized with “final aspiration”, while utterance-internally they are not:

(28) Utterance-final stressed vowel devoicing in Afar (Bliese 1976)

/aba: 'na-h/ 'they do' cf. internal /aba: 'na/
/li 'yo-h/ 'I have'

Bliese does not mention here the fate of utterance-final unstressed vowels (though he does note some optional preboundary glottalization, Bliese 1976: 164), which allows several interpretations: if this is not an oversight, it is conceivable that final unstressed vowels are never in fact devoiced at all, which would be difficult to explain from the point of view of the generalizations and phonetic explanations detailed above.

Another possibility is that devoicing of unstressed final vowels is partial, as it is with stressed vowels, only due to the lower subglottal pressure we can assume is present in unstressed finals, the devoicing does not have the turbulence to be considered “aspiration”, and in this way perhaps escaped description. But if this is the case, there is an obvious problem, viz. why is utterance-final devoicing realized more robustly (or even at all) on a syllable with high enough subglottal pressure to give the impression of
frication when it applies, given that the putative phonetic root of phrase-final devoicing is actually dramatically lowered subglottal pressure? And why at the same time are the final syllables in this language that might actually realize this lower subglottal pressure (unstressed finals) being realized with devoicing of a far less salient degree?

The answer must lie in phonologization. Specifically, if devoicing has ceased to be an automatic consequence of its phonetic surroundings (i.e. has been phonologized), such that the target pronunciation of utterance-final syllables includes partial devoicing of the final vowel (i.e. is phonetically specified to include devoicing), then all utterance-final vowels, regardless of the degree of subglottal pressure associated with any given token, will be realized with some amount of devoicing. Furthermore, and perhaps most importantly, the implementation of that planned devoicing would now potentially vary according to its phonetic environment, meaning that in syllables with increased subglottal pressure it would be realized with greater turbulence and hence be perceptually more salient. Again, the distinction is essentially that drawn by Ladefoged and Maddieson (1996: 49) between passive devoicing and active devoicing. In the former case, that involved in gradient final devoicing, subglottal pressure drops at the end of the phrase to the point where voicing can no long be maintained. The vocal folds thus cease to vibrate, but not because they have been actively configured to do so. In the case of active devoicing, on the other hand, what is by hypothesis occurring in the phonologized cases
is a purposeful devoicing of all or part of the relevant vowel, most likely through some
degree of intentional glottal abduction. With an active abduction gesture, devoicing is no
longer a function of low subglottal pressure. An increase in subglottal pressure, such as
that associated with a stressed vowel in some languages, would now serve actually to
enhance the perceptual robustness of the devoicing, rather than to impede its
implementation, as in the unphonologized cases. A similar account is presented below of
the phonologization process in analogous cases of final glottalization.

Further evidence that final devoicing is phonologized in this way in Afar comes
from what Bliese describes as “h-epenthesis” (Bliese 1976: 160). As Bliese describes it, a
stressed word-final vowel may be followed by and [h] phrase-internally when the
following word is vowel-initial. He characterizes the process as an optional means of
keeping the two words separate to avoid the confusion which the application of
assimilation rules in external sandhi would cause. In the case of certain morphological
contexts, however, h-epenthesis is obligatory. All this suggests a process now reasonably
distant from its original phonetic roots, subject to, among other things, morphological
conditioning.

Before moving on, it is necessary to address a final observation made by Gordon.
Gordon notes that the lowered subglottal pressure that causes final devoicing is “in direct
competition with another common crosslinguistic property of final position: final
lengthening” (p. 101). This is clearly true with respect to the potential phonologization patterns that might result from these phonetic properties, insofar as completely devoiced final vowels are unlikely candidates for the phonologization of final resistance to qualitative vowel reduction, and vice versa. But there is another sense in which the claim is problematic. Gordon notes (correctly) that despite the evidence that the increased duration associated with final long vowels or final stressed vowels impedes final devoicing, duration contributed by final lengthening generally does not rescue final vowels from devoicing (though the existence of cases such as Dasenech above probably count as exceptions to this otherwise valid generalization). Gordon attributes this pattern to the fact that final lengthening, as opposed to stressed vowel lengthening, does not involve any increase in the magnitude of associated gestures, but rather is just a consequence of local gestural slowing (citing Beckman et al. 1992). If anything, in fact, Gordon notes that since devoiced vowels are actually typically shorter than their crosslinguistic voiced counterparts, the subglottal pressure drop “not only inhibits final lengthening, it also appears to induce final shortening”.

Given the evidence of the numerous phonetic studies and phonological patterns cited above, the claim that final unstressed vowels (and perhaps stressed as well) are lengthened without any increase in gestural magnitude just cannot be accurate, at the very least in a large number of cases. Though the phonetic studies deal only with a few
languages, were this claim true more generally, the Final Resistance patterns discussed above would by all rights not exist. It is my contention here, however, that the two phonetic trends, (final lengthening and a phrase-final drop in subglottal pressure) are not in fact contradictory in synchronous phonetic implementation in any way. If the defining characteristics of final lengthening are a reduction in gestural stiffness and an increase in the magnitude of certain supralaryngeal gestures (e.g. Cho’s sonority enhancement), then it is not clear what possible detriment to the integrity of this pattern’s implementation could be caused by simultaneous low subglottal pressure (except over time through reinterpretation). That the perceptual consequences of this cooccurrence could result in diachronic implementational changes owing to the lowered perceptual salience of the supralaryngeal strengthening in this environment is clear enough, but it is not what is at issue here. The two phenomena can and regularly do coexist within a system, as exemplified by the phrase-final vowels of Russian, exemplified below in (29).

---

89 Where examples of implementationally contradictory phonetic properties would be something like the relationship between nasalization and oral frication, or voicing and turbulent frication downstream, where the realization of the former in each case actually obliterates the phonetic conditions necessary for the implementation of the latter (i.e. a substantial pressure drop across the oral constriction) (Ohala and Ohala 1993).
In this phrase-final token of the word /granáta/ 'grenade', it is clear that final lengthening occurs\textsuperscript{91}. The posttonic vowel, ordinarily among the shortest in a given word, is 121 ms long (measured conservatively). The stressed vowel has a duration of 119 ms and the pretonic 106 ms. The (relatively late-occurring) F1 maximum for the final vowel

\textsuperscript{90} Or potentially 'pomegranate', genitive singular.

\textsuperscript{91} This spectrogram and all the attendant analysis was made using the Praat 4.0.2 speech analysis software (Copyright@1992-2001 by Paul Boersma and David Weenink). Formant measures were taken using LPC autocorrelation analysis.
is 568 Hz, with an F2 of 1239 Hz. A short study of formant values for unstressed /a/ in 63 open first pretonic syllables (in non-nasal contexts followed by voiced or voiceless obstruents) from the speaker producing the spectrogram above (a Leningrad native) yielded a mean F1 of 555 Hz and a mean F2 of 1216 Hz, placing the values for the final vowel of this token well within the speaker's range for Degree 1 reduction (see Chapter 2). The unreduced stressed /a/ in this token, as expected, has a higher F1: 608 Hz, with an F2 of 1308 (a bit high, probably due to the surrounding coronals). The final is then clearly resistant to reduction in the sense outlined above. It also has a low amplitude, dropping off drastically toward the end. The irregular wide spacing of the glottal pulses together with the variations in their amplitudes indicates (audible) creaky phonation92, and the very final portion of the vowel appears to devoice slightly before the pause. All this suggests the low subglottal pressure we expect to find in phrase-final syllables (see below for the less-than-obvious reason this should be the case for creaky voice).

As noted, the coexistence of low subglottal pressure (with concomitant devoicing) and final lengthening (with concomitant supralaryngeal gestural strengthening), while not inherently problematic from a production point of view, is nonetheless potentially vulnerable to perceptually-driven reanalysis, due to the conflicting acoustic results of these patterns (essentially the obscuring of strengthening by low amplitude, non-modal

92 Itself actually potentially contributing some raising of F1 in this example (Gordon and Ladefoged 2001).
phonation and devoicing). This vulnerability can ultimately lead to phonologization of a situation with very different articulatory characteristics. Leaving aside the Russian pattern which has only minor devoicing toward the end of the lengthened vowel, suppose we have a language in which the laryngeal specifications and subglottal pressure in phrase-final position are such that they frequently result in the devoicing of a significant portion of an otherwise lengthened short vowel (i.e., the supralaryngeal gestures of which are slowed and enhanced in some respects, and which perhaps is realized as fully or almost fully voiced in some percentage of its instantiations). Again, there is no particular reason that the implementation of final lengthening should not coexist with the severe drop in subglottal pressure or even with the devoicing pattern. Of course, if devoicing becomes frequent enough and lengthy enough, there will be little acoustic evidence left for the lengthening or strengthening gestures of phrase-final vowels (or for accurate identification of any kind for that matter). This situation could easily lead to reanalysis by the listener, with phonologization of the devoicing, such that the target articulation of a phrase-final vowel should now contain the active devoicing plan discussed above. At this point voicelessness is no longer a function of subglottal pressure, and as a targeted articulation becomes subject to the phonetically-induced variation common to the realization of segments of its type. Which is to say, in circumstances of raised subglottal pressure, the voicelessness, rather than being impeded, would actually be accompanied
by more turbulent airflow through the glottis. Gordon notes that voiceless vowels are generally shorter than comparable voiced vowels, and indeed an active glottal abduction gesture in the absence of lowered subglottal pressure would certainly increase volume-velocity of airflow through the glottis, perhaps with the result of a substantially briefer phrase-final vowel.

But is this really an instance of "phrase-final shortening", in the sense of cancellation or even reversal of the otherwise-common phonetic tendency toward phrase-final lengthening? The answer here must be no. Phrase-final lengthening was implemented as normal up until the moment of the phonologization of the final vowel's voicelessness, at which point the vowels durational properties would be changed completely to correspond to the phonetic characteristics of the new target articulation. So what happens to the final lengthening in the process? Certainly at this point phrase-final vowels are now being realized without such lengthening (insofar as they are short and voiceless by design). But this does not exclude the possibility that there is still final lengthening in the system: While the conclusive data relevant to the example here would come from, e.g., Turkana, an extremely interesting finding in an experiment on final lengthening in Dutch by Cambier-Langeveld suggests what the answer may be. Cambier-Langeveld (1997) found that in Dutch, words with final syllables containing full vowels undergo phrase-final lengthening implemented primarily on the rhyme of the word-final
syllable. Previous syllables are not significantly effected. On the other hand, when a word-final open syllable has a nucleus with the reduced vowel schwa, final lengthening reaches further back into the word, lengthening the preceding syllable as well. Words with a final closed syllable containing schwa also saw some lengthening of the penult, though not as much. Cambier-Langeveld's analysis is that the overshort vowel schwa may not be capable of lengthening to the degree that other final vowels are. If this is so, and the domain of application of final lengthening (the period of activation of the π-gesture in Byrd's terms) has a relatively fixed duration, the failure of the schwa to occupy that entire period would result in the extension of the lengthening past the final syllable deeper into the word. In Dutch then, final reduced vowels may undergo little final lengthening themselves, but this is a completely different claim than one which asserts that final lengthening does not operate in words with final schwa, or that the presence of schwa in the final syllable induces a process of final shortening. If final lengthening as a phenomenon is as general and universal as the evidence (linguistic and non-linguistic) suggests that it is, the very same pattern may be manifested even in languages with complete devoicing of phrase-final vowels. If these devoiced vowels are short, it means that final lengthening does not significantly affect phrase-final devoiced vowels. It does not mean that the language has no phrase-final lengthening, or that any lengthening it has differs significantly in kind (rather than in domain of application) from that found in other
languages. Lowered subglottal pressure in this case has resulted in the phonologization of phrase-final vowel devoicing, with concomitant (probably predictable) changes to the segmental domain of final lengthening in the relevant phrases. It has not, by contrast, resulted in the effacement (let alone reversal) of the phrase-final lengthening pattern as such. This argument is important for analogous situations in the following two sections.

Final vowel devoicing can thus be phonologized directly as one of a variety of patterns of final devoicing, aspiration, clipping or shortening of domain-final material. As I will argue below, it can also contribute to phonologization in a more subtle way, by contributing to the perceptual obscuring of vowel quality contrasts which would otherwise be robust in a lengthened final syllable.

3.6.2. Glottalization

As seen above, along with the devoicing caused by the characteristic phrase-final subglottal pressure drop (and the equally-characteristic and related phrase-final lowering of F0), another common salient feature of final syllables crosslinguistically is glottalization or laryngealization. I will use the term glottalization throughout (despite the contradictions this entails described below) to highlight the connection between the allophonic creaky phonation characteristic of phrase-final syllables in many languages (e.g., English, as treated in Henton and Bladon (1988), or in Russian as manifest in the

213
spectrogram above) and phonological epenthesis of final /ʔ/, often ultimately in heavily grammaticalized circumstances, as Hyman (1988) illustrates for Dagbaani. I will assume, along with Hyman (1988), that at least in many cases the epenthetic final glottal stop so common in the languages of the world is ultimately the phonologization of allophonic phrase-final creak. That there is an intimate enough connection both articulatorily and perceptually for such a scenario to be plausible is clear from the discussion of glottal stops in Ladefoged and Maddieson (1996: 73-77) as well as from the analysis of compensatory lengthening triggered by loss of glottal stop in Kavitskaya (2001). Specifically, while glottal stop is definitionally at least produced by a complete adduction of the vocal folds resulting in the momentary cessation of glottal airflow, it is in fact often realized allophonically in many (if not most) languages whose inventories contain such a segment without full closure of the glottis. According to Ladefoged and Maddieson (1996: 75), “In place of a true stop, a very compressed form of creaky voice or some less extreme form of stiff phonation may be superimposed on the vocalic stream”.

I will depart from this aspect of Hyman’s analysis, however, in one respect: Hyman claims that though final allophonic creak is surely the ultimate diachronic source of final glottal stop insertion in many cases, this may not be historically true of all cases. While contradicting this claim outright would be rash, I would nonetheless cast doubt on
the case for a second source proposed by Hyman to account for comparative evidence from Akan dialects and from the Guang subgroup of Volta-Comoe. The pattern is as follows. For certain cognate forms, one language or set of languages shows a vowel-final disyllable, while another language or set of languages shows a monosyllable ending in a glottalized sonorant. Thus, in data from Snider (1986):

(30)  Final vowel loss in Guang

<table>
<thead>
<tr>
<th>Gonja</th>
<th>Chumburung</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ka-wul?</td>
<td>wuri</td>
<td>'skin'</td>
</tr>
<tr>
<td>ku-ful?</td>
<td>ki-furi</td>
<td>'moon'</td>
</tr>
<tr>
<td>e-nin?</td>
<td>o-pari</td>
<td>'man'</td>
</tr>
</tbody>
</table>

Hyman suggests then that rather than having the “phonologization of pause” found in some languages, in this instance we are dealing with the reduction of a final segment, with glottalization of the consonant representing a trace of the lost final vowel. It is certainly common enough that the loss of a final vowel leaves traces on preceding segments (umlauts of various sorts on vowels, distinctive palatalization or affrication of preceding consonants, etc.). It seems unlikely, however, that a vowel would leave of all things glottal constriction on a preceding consonant unless it were itself glottalized. We might hypothesize then that the direct ancestor to the Gonja forms above was one with a short, low amplitude, creaky-voiced final vowel, whose laryngealization may either have affected the sonorant itself in implementation, or at least have obscured the already
lower-salience acoustic modulation of the transition from the sonorant to the short vowel\textsuperscript{93}, such that the sequence could be reinterpreted as merely a laryngealized sonorant\textsuperscript{94}. This would mean Hyman is correct to see this correspondence as the product of reduction of a final segment, with the sole caveat that the pre-deletion phonetic realization of the final segment in question was quite probably the result of the phonologization of prepausal creakiness. This may seem a pedantic distinction, but it is an important one as well if the question we are posing is an etiological one\textsuperscript{95}.

Before looking more closely at the crosslinguistic typology of final glottalization, it is necessary to address the phonetic issues surrounding the development of creaky voice in phrase-final syllables. Most sources treating creaky phonation characterize its production and acoustic and aerodynamic consequences similarly: Gordon and Ladefoged (2001: 386) describe it as owing to “vocal folds tightly adducted but open enough along a portion of the length to allow for voicing”. Ladefoged and Maddieson (1996: 53) speak of “a great deal of tension in the intrinsic laryngeal musculature”, and note that for a given amount of subglottal pressure, the tight adduction of the vocal folds yields substantially lower rates of airflow through the glottis than would be the case for

\textsuperscript{93} As compared with, e.g., that of the transition from obstruent to vowel.

\textsuperscript{94} The additional duration supplied by the misparsed final vowel could actually help in this regard, making a (probably shorter) onset sonorant sound more like a (probably longer) coda sonorant.

\textsuperscript{95} In treating these facts, of course, one might wonder about the role of the ATR specification of the lost vowel in this connection. The cognate dialect forms retaining the vowel, however, show both vowel series yielding the same result upon deletion.
modal voicing (1996: 50). From this we can extrapolate that for a given rate of airflow through the glottis, creaky phonation should require greater subglottal pressure than modal voice. The facts concerning pitch are ambiguous (and tellingly so, as will be shown): Creaky phonation is most commonly associated with a low F0 relative to modal voice, but in many languages is in fact correlated with high F0 tonal contours (Gordon and Ladefoged 2001: 400). This latter is odd, particularly in light of Silverman’s (1997: 149-152) discussion of the interaction of creaky voice and pitch. Silverman notes that in many ways high F0 and creaky phonation are actually far more compatible than low F0 and creaky phonation, given the articulatory properties attributed to both (1997: 150-151):

First, low pitch involves decreased vocal fold tension, while creakiness is enhanced by increasing vocal fold tension. Also, creakiness is enhanced by increasing subglottal pressure, while lowness is enhanced by reducing subglottal pressure. Finally, low tones involve pronounced larynx lowering, while creakiness may be accompanied by larynx raising.

Gordon and Ladefoged mention allophonic phrase-final and phrase-initial glottalization as cues to prosodic constituency, but make no mention of any potential contradiction in these two phonetic environments. Comparing the two, the higher pitch and subglottal pressure associated with the beginnings of phrases is well-suited to the tightly-adducted-vocal-folds implementation of creak. In initial position, Gordon and Ladefoged also note that glottalization is greater on accented syllables than on unaccented (Pierrehumbert 1995, Dilley et al. 1996).
As concerns our phrase-final syllables, however, the relevant phonetic properties seen so far have been the exact opposite of this: all things being equal phrase-final position is associated with lowered pitch, lowered subglottal pressure, and the syllables that most concern us are unstressed. Compare the two Russian phrase-final words in (31) and (32), one with final stress, and one with medial. The unstressed final vowel (bottom) is clearly creaky throughout, while the stressed vowel (top) is modal for at least 120 ms, becoming increasingly breathy thereafter, with no creak to be found.

(31) Russian final-stressed [kæbələ] ‘kabbalah’

![Waveform and Spectrogram of Russian final-stressed [kæbələ] ‘kabbalah’]
In other words, the creak appears on the vowel by all accounts least likely to host it, and fails to do so on its far more appropriate stressed counterpart. It is also not at all clear in light of the foregoing how to explain the properties of the final vowel in (29), [grenâte] above, in which amplitude declines gradually to the end of the vowel, first with creakiness, apparently easing into voicelessness at the end (with no evidence of an actual glottal occlusion in the area).

Silverman (1997) recognizes these contradictions in the context of the appearance of creakiness on low tones in Chong and Mandarin. As a solution, he proposes that the
creakiness occurring in the low tone environment actually has a different source and mode of production than the creaky voice described articulatorily in most studies. Specifically, he hypothesizes that low-tone creakiness is in fact a consequence of the lowering of subglottal pressure (and hence airflow across the glottis) needed for the F0 drop, such that “the vocal folds may more readily seal the subglottal chamber, and thus are only intermittently blown apart by eventual subglottal pressure increases. This slow and irregular glottal pulse may give rise to creakiness” (1997: 151). He hypothesizes further that this may be the case with phrase-final glottalization as well.

Silverman’s hypothesis is largely confirmed by the work of Slifka, reported in Hanson, Stevens, Kuo, Chen and Slifka (2001). Slifka shows that in periods of phrase-final “glottalization”, subglottal pressure is in fact gradually falling, while airflow is increasing, and thus that the vocal folds are in fact being gradually abducted during the production of this vowel. She concludes that the irregular phonation in question cannot possibly be glottalization (as described above), but rather that the irregular glottal pulsing arises from a combination of falling subglottal pressure, vocal fold slackening and vocal fold abduction. She continues to suggest that glottalization in the environment of low tone has the same origin, resulting from the reduction in subglottal pressure and vocal fold slackening necessary to produce the desired pitch level.
This explanation is in perfect accord with the observed facts of phrase-final allophonic glottalization (a term I will continue to employ despite its demonstrated inaccuracy) in Russian. It is likewise immediately clear how the smooth transition from creak to breathiness and voicelessness can occur in the course of the final vowel discussed above. Both final devoicing and final glottalization thus have essentially the same phonetic source (subglottal pressure drop and gradual laryngeal-gesture “fade” to use the term suggested inaccurately above for all final gestures). They can also both be characterized as resulting from phrase-final phonetic weakening and prominence-reduction, though of course, in their phonologized states, there may be nothing weak about them, from an articulatory point of view.

The crosslinguistic patterning of final glottalization is parallel to that described above for final devoicing. Gordon’s generalization that word-final devoicing implies phrase-final devoicing seems to be continued here: I was unable to locate any clear instance of a language that glottalizes word-final vowels only when they are phrase-internal, which is fairly unsurprising given the above. And as above, it is primarily unstressed vowels that are subject to phrase-final glottalization. Again, however, there are some crucial counterexamples.

Hyman quotes Timothy Vance writing that in Tokyo Japanese, which inserts glottal stop after short vowels before pause, “...the glottal stop after a short vowel is more
salient when a speaker is excited and emphatic” (Hyman 1988: 114). From what we now know of phrase-final glottalization, however, “excitement” or emphasis should if anything prevent glottalization of the final short vowel. Likewise, recall from the discussion of devoicing above Sapir’s report that in Southern Paiute the “rhetorical emphasis” which rescues final vowels from devoicing actually results in their lengthening, sometimes with the addition of a glottal stop. Finally, in the Isthmus Veracruz Nahuat dialect of Nahuatl, a “junctural glottal stop” is inserted before phrase boundaries (Wolgemuth 1969). This glottal stop after short vowels is distinguishable from the phonemic glottal stop in being very brief and lacking complete closure. When the final short vowel is stressed, however, this junctural glottal is “lengthened and strengthened”, so that it comes close to being indistinguishable from its underlying counterpart. Obviously, all of this is a problem for the above account of phrase-final glottalization as a consequence of lowering subglottal pressure and gradual vocal cord abduction at the end of the prosodic domain. It would make perfect sense, however, if the glottalization in question on these final vowels were actually true glottalization, the tight glottal adduction and tense laryngeal musculature generally described for distinctive creaky voice or glottal stop.

Here as with devoicing, the solution to the quandary is in the nature of phonologization. In its embryonic form, phrase-final glottalization is a phonetically-
contingent gradient effect, dependent entirely on the convergence of the laryngeal and alveolar states listed above. Glottalization is not an independently targeted feature of the articulation. The acoustic features produced by the irregular pulsing of the glottis this “automatic glottalization” causes, however, are extremely similar to those associated with true creaky phonation or glottalization (i.e. irregular pulsing, relative F0 lowering, lower amplitude, and characteristic spectral tilt⁹⁶). If gradient glottalization becomes regular and salient enough then on final syllables, the possibility for its reinterpretation by the listener as intended or targeted rather than contingent becomes quite real. As a consequence of this reinterpretation, glottalization becomes extrinsic, independently targeted, and hence realized as glottalization proper, with the tightly adducted vocal folds we would expect of it. At this point, a rise in subglottal pressure such as that associated with stress or emphasis would come to increase the perceptual robustness of the creakiness (and glottal occlusion if there is one), yielding the patterns described above. It is important to bear in mind at this point that, as before, “phonologized” does not necessarily mean “phonemicized”. The emphatic glottal stop in Tokyo Japanese, for example is “non-distinctive” according to Vance. This seems to be the case quite

⁹⁶ Though creaky voice can of course be produced with higher amplitude (unlike phrase-final irregular phonation presumably), it has been shown that glottalization can actually be cued in some languages that employ it contrastively with lower relative amplitude alone (Baker and Gerfen 2001). The lowered amplitude of the phrase-final syllables could actually increase the likelihood of reanalysis then. Though as noted above higher F0 may enhance creakiness, the direct effect of the irregular pulsing with longer closed periods in the glottal cycle is an F0 drop relative to neighboring modal-voiced segments.
frequently here\textsuperscript{97}. If the resulting configuration is non-contrastive, and it plays no obvious role in other aspects of the lexical or categorical phonology, there is no reason to assume the workings here of anything other than the phonetics, as outlined in Chapter 1. Which is to say, this glottalization, despite being independently targeted, may nonetheless have no direct representation in the phonology beyond the cooccurrence of a word-final vowel and a phrase-boundary.

This fairly innocuous conclusion, however, has certain peculiar consequences for some models of the phonetic component. We have been assuming, with Keating (1996), that the phonetic component must be capable of representation varying degrees of specification of a given feature. A narrow specification (or target window), often misleadingly called "categorical" implementation, allows little contextual variation, while a wider specification allows the influence of contextual factors to be felt more strongly, resulting in gradient or contingent instantiations of certain features or processes. In the case of devoicing and glottalization, it has been necessary to distinguish even complete lack of a specified target or gesture from the presence of such a target, however variable in implementation. In the case of phonetic underspecification (whether we mean complete lack of a specified target, as with passive devoicing, or only having a broad target window, like gradient), the realization of the feature or features in question will be

\textsuperscript{97} And almost exclusively in the case of devoicing.
to some extent automatically interpolated, either as a function of the realization of other
features (e.g. duration), or as transitions between two specified points along a continuum
of potential targets for the same feature (i.e. pitch). Importantly though, both are tokens
of the same phonetic feature, broadly conceived. A partially or completely underspecified
realization results in a broad continuum of potential degrees of instantiation of a
particular feature, while a narrowly specified feature singles out a small portion of that
continuum of possible realizations as optimal. In the case of the phonologizations above,
however, the specified and underspecified variants are instances of two completely
distinct types from an articulatory point of view (in that the gestures needed to make
them are largely mutually incompatible). If our phonetic component contains
representations of acoustic features or properties, of course, then the problem is more
tractable; both types of glottalization can be represented in acoustic terms according to
their common properties, restoring the original relationship between specified and
underspecified as different instantiations of a single feature type. In a strictly articulatory
model of the phonetic component (such as that advocated by many practitioners of
Articulatory Phonology (Browman and Goldstein, passim), however, the problem is more
serious. Here the relationship between the specified and the underspecified breaks down,
as there is no point or window along a continuum of possible instantiations of "lowered
subglottal pressure, vocal fold slackening, and vocal fold abduction" which corresponds

225
to the narrow target “vocal fold tension and tight adduction” (in the same sense that a command to lower the jaw is not a point along a continuum of possible jaw-raising gestures, despite the fact that the same organ is involved). In this model, then, the phonologizations of devoicing and glottalization described above cannot be characterized as instances of a shift from phonetic underspecification to phonetic specification in the way I have been proposing. Instead, they must be seen as instances of another type of process altogether, one utterly distinct in kind from, e.g. the phonologization of consonant-induced F0 perturbations on following vowels discussed by Hyman. This result commits us, however, to a serious mischaracterization of these processes, and in so doing deprives us of significant insight into both the nature and working of phonologization, and the typology of final-position effects as well98.

3.6.3. Mixed Systems

It has been demonstrated above that both allophonic devoicing and glottalization can coexist within a given system in measured doses. In Russian, however, glottalization clearly has the upper hand in terms of salience, with any devoicing that takes place

---

98 The most illustrative system in this regard would be one in which glottalization had been phonologized (as, say, insertion of contrastive glottal stop) either utterance or Intonational Phrase-finally, while before lower level phrase boundaries gradient irregular phonation was still being realized. To show this conclusively an aerodynamic study of the type Slifka performed would be most desirable.
phrase-finally being durationally small and sporadic in realization. Other systems, however, have developed differing distributions of the two. Some languages, for example, appear to use one or the other in free variation. Gordon (1998) cites Apinaye as having both lengthened, devoiced and creaky allophones of phrase-final vowels, but mentions no specific distribution. Afar (Bliese 1976) in addition to its final stressed-vowel devoicing, is said to have glottalization before certain boundaries, though the distribution is again not clear. Isthmus Veracruz Nahuat has a free variant of its junctural glottal stop called “fricative”, which is a glottal stop, after which “an ‘h’-like release is heard”.

The distribution of the variants can be phonologically conditioned, as in Koyra (Hayward 1982: 217-218), an Omotic language in which utterance-final vowels (all of which are short - see below) are realized voiced but “terminating with a glottal closure” when following voiced consonants or phonemic glottal stop, while following voiceless consonants they are “extremely brief and more or less devoiced”, especially after sibilants, where they are all but inaudible. Varying laryngeal settings can even be distributed according to sociolinguistic criteria, as in the famous case of Koasati (Muskogean, Kimball 1991), in which in the speech of men, final vowels are all

---

99 This wouldn’t necessarily be the phonologization of final vowel devoicing though.
“prolonged”, while in the speech of women a final glottal stop is inserted after the vowel (with laxing of final high vowels)\textsuperscript{100}.

Often the distribution of devoicing and glottalization serves to distinguish final long vowels from final short vowels (a contrast which, as described below, is otherwise prone to collapse in final position). Hyman notes that in most languages with final glottal stop insertion, the process only applies after a short vowel (assuming a contrast, though in collapsed systems the quantity-neutralized glottalized vowel is generally short phonetically as well). The long vowel in these systems usually remains untouched, as in Hausa, where according to Carnochan it “just dies away” (Carnochan 1988)\textsuperscript{101}, or in Luiseño (Hyde 1991), where short stressed vowels have a “clipped” quality, where longs and unstressed vowels do not\textsuperscript{102}. I have not found any system in which the short vowel is glottalized and the long vowel partially (but significantly) devoiced. Such a system could presumably come into being, though it would likely be quite unstable, due to the resulting durational similarity of the voiced portions of the vowels, which could lead to the collapse of the quantity distinction. The reverse, on the other hand, is clearly attested. Hyman cites Oromo and Tepehua as instances of such a system. In these languages.

\textsuperscript{100} Men will also choose glottal stop insertion from time to time, apparently when they wish to be “precise”.

\textsuperscript{101} A small set of long-vowel-final words in Hausa is also realized with final glottal stop, though they do not neutralize durationally with the short vowels (Newman and Van Heuven 1981).

\textsuperscript{102} It seems likely that the unstressed shorts must have or have had some allophonic glottalization as well, just not even to sound “clipped”, as would, e.g. a phonologically inserted glottal stop.

228
where the short vowels are completely devoiced and the long vowels are glottalized, the presence of any significant voiced period, even if somewhat abbreviated by the glottal adduction gesture, would be sufficient to distinguish long from short, making a perceptually more stable system. Devoicing of both long and short final vowels is reasonably common (with the proviso that the long vowel is only partially devoiced). This is in fact the system described by Sapir for Southern Paiute (Sapir 1930). In Southern Paiute final long vowels are partially devoiced or shortened, and short vowels are realized voiceless or completely elided. This is essentially the pattern termed “Final Mora-Clipping” by Hayward (1998). Hayward is referring here to any system in which a final long vowel is shortened or partially weakened in some way (a bimoraic vowel becomes monomoraic), and a short vowel is severely reduced or deleted (a monomoraic vowel goes away). Southern Paiute had conserved the pattern of “Final Mora-Devoicing”, which obviously gave rise to the alternate pattern of actual “clipping” or deletion. Many languages exhibit this latter pattern, which I assume is a reflex of a devoicing pattern in most cases. Languages such as Latvian (Matthiessen 1997), Maltese (Brame 1972), and D’irayta (Hayward 1998) exhibit this pattern103. Gilbertese represents a compromise variant, in which all final long high vowels shorten, but only short high

103 In the case of full deletion, of course, and absent comparanda or historical records, such a system would be indistinguishable from a system in which the quantity distinction is neutralized, but reflexes of both vowels are retained, described below.
vowels after nasals are deleted (Blevins 1997). The remaining logical possibility, 
glottalization of both long and short vowels (with retention of the contrast) is attested, but 
in only one instance I have found so far (Isthmus Veracruz Zapotec, Wolgemuth 1969). 
Again, it is possible that such a system would be inherently unstable, if the glottal 
adduction gesture were to shorten the long vowels enough that the durational difference 
became less robust\footnote{The result of the combinations attested above does seem to be that those patterns which would best 
preserve the quantity contrast are the ones which are best attested. This does not imply teleology, however, 
or “intelligent design” as wags have put it. These systems are the best attested simply because the other 
combinations however often they may emerge, are simply less likely to remain observable over the long 
term, resulting in their comparative rarity. The logic here is evolutionary, in the spirit of Blevins (in prep), 
Blevins and Garrett (1998), and Ohala (passim).}.

In the above sections I have detailed aspects of the phonetics of phrase-final 
position which can be characterized as “weakening” or “prominence-reducing” in nature, 
at least in their unphonologized state. I have presented a typology and developmental 
analysis of the phonological patterns which arise from these phonetic trends. I have also 
argued that the presence of these phonetic weakening patterns is in no way 
implementationally antagonistic to the presence in the same system of phonetic patterns 
of phrase-final lengthening and gestural enhancement along some dimensions. The 
acoustic result of the overlay of these patterns within a system, however, may indeed 
produce instability, resulting in phonologization patterns which ultimately efface the one 
set of patterns or the other (at least to the extent that in a system in which phrase-final
devoicing has led to final vowel deletion, or final clipping, there is an obvious sense in which the effect of the devoicing pattern has in the end won out over the final lengthening pattern. I have also shown, however, that this in no way necessarily means that in the resulting synchronic system there is no process of final lengthening or strengthening. Rather, synchronic final lengthening could simply be active over a wider span of segments than was the case in the ancestral system.

3.6.4. Nasalization

Before returning to the question posed at the beginning of this section on final weakness, that of how to reconcile evidence for supralaryngeal strengthening in final position with the idea of “final fade” and descriptions of “reduced” phrase-final vowels, it is necessary to outline one additional phonetic tendency associated with final position crosslinguistically: vowel nasalization. This tendency is not as well-known as those discussed above, and, due to the subtlety of the effect in some cases, I suspect underreported as well. One recent treatment of final nasalization is that of Hock, in his 1999 paper, mentioned above as advancing the hypothesis that final position is a monolithic weakening environment. Hock mentions only one case: Vedic Sanskrit. In Vedic Sanskrit, there is a sporadic tendency for prepausal vowels to be written nasalized. Hock attributes this to a weakening of the velic closure at the end of a phrase, where the
articulatory organs are returning to the rest position. Hock does not discuss conditions on
the distribution of this nasalization, however.

It is still not easy, unfortunately, to assess the claim that phrase-final position is
generally correlated with weakening of the velic closure gesture from experimental
decline, primarily in stressed syllables, in the height of the velum for oral consonants
over the course of the utterance. Recall, however, that this is roughly parallel to the
findings of Vayra and Fowler for supralaryngeal articulations, findings which also
showed strengthening of these same articulations specifically in word-final unstressed
syllables. If this is truly a position-specific “weakening” pattern for the articulation of
vowels as Hock claims (rather than what could be seen instead as a gradual decline in the
degree of strengthening of gestures in strong positions the stressed syllables), we would
expect it to be especially obvious on unstressed vowels, all the other articulations of
which are well known to be weakened or reduced. What is more, the Krakow et al. study
concerns specifically the position of the velum in the production of oral onset obstruents.
What is measured is velum height, which of course can vary quite a bit without allowing
actual opening of the velum, and in which there are intrinsic differences in the
articulation of non-nasal segments. What was found was a declination such as that
reported in Vayra and Fowler (1992). Here, the velum height was lower at its last peak in
the sentence than it was at its first peak. The final syllable of all the test sentences was a stressed monosyllable, meaning that, were there a pattern even of local strengthening of velic closure on phrase-final unstressed syllables (analogous to that detected in numerous studies for other supralaryngeal articulations), it would not have been detected. A pattern of declination throughout the phrase is, of course, very different from a localized jump or drop on final unstressed syllables. Were this declination trend part of a general weakening tendency for phrase-final material, we might reasonably expect it to result ultimately in phonologization of the pattern in some instances. Recall that the tendency to phrase-final devoicing is phonologized quite commonly for both final consonants and vowels in the languages of the world. I am not aware of any phonological system, by contrast, in which the oral/nasal contrast for consonants is neutralized word or phrase finally in favor of the nasal. This suggests that prepausal vowel nasalization may well be the product of something other than velic weakening.

Another study, in fact, that of Vassiere (1988), finds the opposite pattern for strength of velic closure in phrase-final position. It should be noted, however, that while Keating et al. (1999) interpret this study as showing that phrase-final position is “strong” for velum height, what the study is actually measuring may be somewhat more

---

105 The reverse, however, is attested, if rare: In certain Mbam-Nkam languages of Cameroon, final nasals in various combinations denasalize to become oral stops (Hyman 1975: 257).
ambiguous: the claim in Vassiere (1988) is that an oral consonant before a major syntactic boundary undergoes less anticipatory velum lowering from a following VN sequence (in other words, in a -C##VN- string) than it does in CVN strings not interrupted by a major syntactic boundary. The claim, then, is that there is less coproduction involving the velum across such a boundary, rather than that, all things being equal, oral consonants have a higher velum position than they do, e.g. in phrase-medial word-final position. The phonetic evidence is thus somewhat equivocal on phrase-final nasalization as a natural articulatory trend, with a need for further experimental studies to show whether or not such a thing truly exists.

The case for final nasalization as a weakening process in Sanskrit itself is also not a particularly strong one. Wackernagel (1895) is not overly specific as to the conditions of phrase-final nasalization of vowels, but a study by Lubotsky examines all such forms in the Ṛgveda (Lubotsky 1993). The first and most obvious generalization is that such examples are quite few in number and pattern in no exceptionless fashion. Over the years scholars have explained it, in fact, as an editorial scheme to avoid certain confusions potentially arising in hiatus contexts. It has also been deemed a “Prakritism”, apparently being a more common phenomenon there and in Pāli (p. 197-198). That said, Vedic final nasalization does appear in one particular context more than any other: It affects only /a/, and then most often unaccented /a/ at the end of odd numbered pādas (a metrical unit
corresponding more or less to a line of verse) before accented vowels (most often /e/ or /o/, also before syllabic /r/). There are other cases as well, though Lubotsky considers the original case to be of an unaccented /a/ nasalizing in hiatus across a metrical boundary before a rising accent on the following vowel. He considers it a consequence of the lengthening of this final vowel he believes would have occurred due to an anticipatory pitch rise before the accented vowel (p. 206). One other context for nasalization of /a/, interestingly, is in the so-called pluti forms. These are forms with lengthened final vowels used in certain interrogative and vocative contexts, which Lubotsky notes is thought to be a rising intonational contour. Whatever the case ultimately here, it seems clear that lengthening is involved, and in the case of the pluti perhaps even some emphatic element. Whether we can truly speak of “weakening” in this case is simply not clear.

One thing that is interesting about the Sanskrit example, however, is that it affects only /a/. This is also the case in some of the examples below. Now, it is a well-known crosslinguistic generalization that velum height tends to vary directly with vowel height, to the extent that low vowels are often actually produced with a substantial degree of velic opening. There has been much dispute as to why exactly this should be so. Whalen and Beddor (1989) support a production-based hypothesis, whereby velum lowering is an automatic consequence of the articulation of low vowels, and not due to active control of velum height by the speaker. It is also well-known that as systems develop distinctive
nasalization, it is generally the low vowels which nasalize first (Chen 1973, Whalen and Beddor 1989), suggesting that this inherent nasalization also makes them more vulnerable to reinterpretation as intentionally nasalized. In order to explain this fact (and particularly its instantiation in Eastern Algonquian spontaneous nasalization), Whalen and Beddor cite experimental data from Henderson 1984 showing that the difference in velum height between oral and nasal /a/ in Hindi is less than the difference for other vowels. Their interpretation of these facts is that /a/, already produced with a lowered velum, will achieve nasalization with a relatively smaller additional lowering of the velum than would other vowels. Thus, if all vowels were produced with the same slight amount of velum lowering (for whatever reason), /a/ might achieve a more salient degree of nasalization than other vowels. If this is the right explanation, then we could imagine that, were there in fact a slight trend toward velum lowering on all vowels phrase-finally, the same amount of lowering might cause /a/ to be perceived as intentionally nasalized sooner than any other vowel in the system.

Whalen and Beddor also show, however, that lengthening of a vowel which is already nasalized to some degree will increase the perceptual salience of that nasalization significantly, which is to say that a slightly nasalized vowel will be perceived as nasal more readily if it is longer. Whalen and Beddor note that this alone cannot explain the susceptibility of /a/ to nasalization, since although low vowels do generally have longer
inherent durations than other vowels, these differences are not of the magnitude necessary in their study to increase the percept of nasalization. Consider this evidence, however, in light of nasalization in phrase-final position. Phrase-finally lengthening routinely increases the durations of final vowels by the amount shown to be necessary by Whalen and Beddor for an increased perception of nasalization on vowels which are already partly nasalized. The facts of Vedic Sanskrit, associating nasalization with positions of particularly dramatic vowel lengthening, agree very neatly with this line of reasoning. Furthermore, since low vowels are the only ones which would have had any degree of nasalization on them to begin with given the crosslinguistic characteristics of their production, it is possible to understand why only they would be affected in these circumstances. This account of the development of phrase-final nasalization in Sanskrit does not require any phrase-final velic weakening to occur, though of course, neither does it exclude it.

The same pattern is found in another case as well: In Neo-Aramaic there is good evidence for phrase-final lengthening. Garbell (1965) provides intriguing information on the realizations of word-final /a/ in the various Jewish dialects of Neo-Aramaic (once) spoken in Persian Azerbaijan (including parts of what is now SE Turkey). In this language /a/ is most commonly realized as [a]. In its Northern dialects, this /a/ is backed in word-final position to [ɑ]. In the speech of females in the Urmī dialect (NW Iran), it is
pronounced word-finally with rounding as [ɔ]. In the Southern dialects, however, it is
lowered, backed, and nasalized: [ɔ]\textsuperscript{106}. It is possible to understand this development in
just the same way as was proposed for Sanskrit above, with the additional possibility that
to the extent that velum lowering is an automatic consequence of the production of low
vowels, perhaps additional lowering here creates additional nasalization as well, again
enhanced perceptually by the lengthening found in final position.

In many systems, however, phrase-final nasalization affects all the vowels of the
system. This is the case in Russian, for example (Zlatoustova 1964: 44-45), where an
original assertion by Shcherba (1912) has been subjected to extensive experimental
testing and confirmed. In phrase-final position Russian vowels are nasalized for at least
part of their duration\textsuperscript{107}. This is true regardless of the preceding consonant’s identity,
though after nasals it is stronger. It is also stronger, for whatever reason, following
palatalized consonants. Stressed vowels for the most part nasalize over less of their
duration than unstressed vowels (Krakow 1993: 102-105 reports also experimental
studies on the effect of stress on velum position, in which for one speaker at least low and

\textsuperscript{106} Certainly the final vowel is undergoing some sort of enhancement of its characteristic features in final
position. The possibility of understanding all the features found here, including rounding and nasalization,
as serving the perceptual enhancement of /a/ in an environment in which a severe amplitude drop might
otherwise obscure it is intriguing. This is particularly so in light of Kingston’s (1992: 101) observation that
a small amount of nasalization will cause vowels to sound lower (though a large amount should do the
opposite). I will not, however, pursue this line of reasoning here.

\textsuperscript{107} As, perplexingly, are vowels realized between two palatalized consonants.
high vowels had a lower velum position when stressed than when unstressed. For another speaker, however, high vowels had a higher velum under stress, while for low vowels the velum was lower - suggesting local hyperarticulation of sort documented by de Jong (1995). The effect is non-neutralizing, as Russian lacks contrastive vowel nasalization.

This pattern is also attested in Cherokee (which has an underlying nasal /§/) in which all vowels are nasalized prepausally (Walker 1975, Whalen and Beddor 1989). Likewise Aikhenvald (1996: 512) lists the following languages as having the areally-common feature of “nasal pause” (word- or phrase-final vowel nasalization): Pirahã (Mura-Pirahã), Irantxe (affiliation uncertain), Rikbaktsa (Macro-Gê), Xeta (Tupí-Guarani), Assurini (Tupí-Guarani), Tapirape (Tupí-Guarani). She notes phonetic nasalization utterance-finally in Jarawara (Arawá).

These latter systems may provide evidence supporting the phrase-final phonetic weakening hypothesis, in that they affect all the vowels of the relevant system, while the lengthening-plus-intrinsic-nasalization hypothesis proposed above should logically work only for low vowels. There is, however, another possible explanation for the development of phrase-final nasalization which again needs no assumption of weakening: rhinoglottophilia (Matisoff 1975). Rhinoglottophilia, simply stated, is the propensity for spontaneous nasalization in the environment of laryngeal (or high airflow) segments. This is discussed by Ohala (1975), who attributes nasalization in the context of voiceless
glottal and pharyngeal consonants to the fact that an open velum would have very few acoustic consequences for either segment type (meaning that in the absence of the requirement of velic closure for accurate perception, none may ultimately be produced), while for the cases of /h/ and potentially other high airflow segments such as /s/, Ohala implicates the acoustic similarities between nasalization and the breathiness that may accompany these segments. Specifically relevant are increased formant bandwidths, the presence of anti-resonances in the spectrum, and a general lowering of amplitude (Ohala1975: 303).

In the previous two sections it was demonstrated at some length that phrase-final syllables are commonly host to both breathiness/devoicing and glottalization, and hence a prime location in which for spontaneous nasalization to arise. In such a scenario an effect like breathiness on a phrase-final vowel, itself there as an automatic consequence of other factors, such as partial devoicing, could be perceived as nasalization by the listener and reanalyzed as a pattern of intentional nasalization in final position. In this scenario the rise of nasalization would not depend on the state of the velum in the original utterance (though again, nor would it exclude the presence of some degree of velic lowering there). Evidence for rhinoglottophilia as a possible source for phrase-final nasalization is found in the following: Isthmus Veracruz Nahuat (Wolgemuth 1969) has already been shown in the preceding section to have two variant realizations of its prepausal glottal stop.
insertion process. The first is simply a glottal constriction of some magnitude, as in (33a) below. The second, (33b), is described as glottal closure with an ‘h’-like release. There is, however, a third variant, as in (33c) which Wolgemuth describes as voicing continuing past a glottal closure, after which a brief nasal re-articulation of the vowel is heard. This variant is apparently especially common in emphatic circumstances, such as repetition of something a linguistic fieldworker fails to hear correctly the first time.

(33) Variants of prepausal glottal stop in Isthmus Veracruz Nahuat (Wolgemuth 1969)

a. [kisaʔ] ‘he leaves’  [moːtoʔ] ‘squirrel’
b. [kisaʔh]  [moːtoʔh]
c. [kisaʔã]  [moːtoʔõ]

Even stronger evidence for the rhinoglottophiliac origin of at least some instances of phrase-final nasalization comes from Yucuna (Arawakan, Schauer and Schauer 1967). In this language final vowels are nasalized only following /h/ or hiatus (a situation commonly enough resolved with glottalization crosslinguistically, though its realization in Yucuna is not described). In this instance, then, it seems the acoustic environment conditioning reanalysis must be present both before and somewhere during/after the vowel in order for phrase-final nasalization to take place.

One further case comes from Warekena (Northern Arawak, Aikhenvald 1996), which exhibits prepausal (and optional word-final non-pre-pausal) insertion of a sequence
/-hʊ/, the vowel is a nasalized copy of the preceding word-final vowel. Aikhenvald describes this as the result of three rules: ‘h’-insertion, translaryngeal harmony, and vowel nasalization. Diachronically though it seems more likely to be a consequence of phonologization of a combination of final lengthening, breathy phonation and spontaneous nasalization.

The correct explanation for phrase-final nasalization could thus be a consequence of a weakening tendency in velic closure toward the end of a phrase, or a perceptually-motivated result of another phonetic characteristic such as lengthening or devoicing or glottalization. It could, in fact, be some combination of all of these, or perhaps differ from case to case.

Whatever its source, it has the potential to neutralize the distinction between oral and nasal vowels word- or phrase-finally in some languages, thus making final position phonologically weak for contrasts of nasality. This is also the case in South Gujarati dialects (Mistry 1997), in which vowel nasalization is contrastive everywhere but on word-final vowels (with denasalization of etymologically nasalized vowels). An environment with allophonic nasalization of all vowels would clearly be a less-than-ideal location for the perpetuation of a nasality contrast.

There also exist, however, systems in which word-final position is phonologically strong for a nasalization contrast on vowels. In these, however, the source is usually
something altogether different. Interestingly, in what is a complication for a theory of phonological strength effects which seeks to derive the licensing of contrasts from the cues that make a particular position optimal for the perception of the phonetic features in question, these strength effects have their origins in the lack of perceptual robustness characteristic of final syllables.

It has just been demonstrated that phrase-final position may actually be quite a bad place for the realization of a vowel nasalization contrast. Nonetheless, in Nanai (Tungusic, Avrorin 1959: 29-32), nasalized vowels appear only in word-final position. The opposition is contrastive. The source of the nasalized vowels is the deletion of word-final /n/ with concomitant nasalization of the preceding vowel. In unsuffixed forms there are thus abundant minimal pairs. Upon the addition of most suffixes, the [n] is restored and the vowel becomes oral. This is not, however, the case with all suffixes. The diminutive [-kān/-kōn], for example, never conditions the return of a formerly-preceding nasal consonant, and the vowel remains nasal. Several other suffixes have already come to restore the nasal only optionally. In Bare (Northern Arawak, Aikhenvald 1996), nasalization again only occurs on final vowels. Nasalized vowels are underlying in only two forms of Aikhenvald’s corpus, while all the others are derived from the truncation of forms ending in any of several suffixes of the general shape /-nV/. The deletion leads to nasalization of the word-final vowel. Interestingly, this truncation takes place only non-
prepausally, where before pause, in another instance of domain-final resistance to deletion, the suffix remains intact. In these examples, nothing about the phonetics of final position is “licensing” the appearance of a contrast between nasal and oral vowels. The phonetic characteristics (here, detrimental) have clearly caused the emergence of the contrast, but can hardly be imagined to facilitate in any active way in the present. The contrast exists here, then, only because it came to exist here, and because it has so far avoided being effaced by the perceptual weakness of the position.

3.6.5. Final vowel reduction and loss of contrasts

3.6.5.1. Lowering

I turn back now to the question which began this section on final weakness, the question as to whether domain-final positions are in some cases singled out as positions of particularly evident phonetic weakening, eventually yielding loss of vowel contrasts. The first reduction or neutralization phenomenon I will take up is final vowel lowering, a neutralization of contrasts more most likely not conditioned by any process of phonetic reduction in final position. More likely it is the consequence, in fact, of phonetic strengthening of the type Cho refers to as Sonority Enhancement in English stressed and final syllables (§3.2.3.1). As such it would be classed among PN patterns with what Smith (2002) calls Positional Augmentation (see Chapter 4 for details).
Both neutralizing and non-neutralizing vowel lowering is attested in final position in a variety of languages. Some caution is indicated here of course, as a word-final inventory of /e, o, a/ out of a word-internal five-vowel system could just as easily be the result of devoicing and loss of final high vowels as it could of lowering\textsuperscript{108}. Some examples of final lowering (or high vowel exclusion) follow:

(34) Final Lowering

a. **Bare** (Northern Arawak, Aikhenvald 1996): Bare exhibits utterance-final diphthongization of the high-vowel /i/ to [ie], particularly under emphasis. Perhaps a weak example.

b. **Ongota** (affiliation uncertain - Cushitic or Omotic, Savà and Tosco 2000): In word-final position, high vowels /i, u/ are optionally lowered to [e, o], such that only three vowels contrast in this position. Neutralizing.

c. **Dasenech** (Cushitic, Sasse 1976): In word-final position, high vowels /i,u/ are realized as [e, o]. Underlying /e, o/ are realized as [ɛ, ɔ], so the process is non-

---

\textsuperscript{108} High vowels are especially prone to such behavior due to their brevity and the potential for the closeness of their constrictions to impede the development of the pressure drop across the glottis necessary for voicing to begin (Ohala 1993).
neutralizing. As noted above, Dasenech final vowels are also subject to devoicing or deletion in phrase-internal final syllables and to partial devoicing phrase-finally.

d. **Castillian dialects** (Penny, 1986): In certain Castillian dialects of Spanish, only [e, o, a] are found in final syllables. In others all five vowels still appear.

There are several possible factors which could give rise to final lowering, some or all of which are likely responsible for the emergence of the cases above. First, one thinks of the 'sonority enhancement” found by Cho (2001) on English final unstressed vowels, manifesting itself in part as lowering. The common pattern, discussed below, in which languages avoid shorter vowels in final position also comes to mind as a source of potential extra length through lowering. Under this interpretation then final lowering is a Positional Augmentation of the sort Smith (2002) shows to be characteristic of certain otherwise phonologically-strong positions. Another possibility is that other phonetic features found in final position are lowering the vowels here, specifically, the presence of creakiness or breathiness/devoicing. Creaky phonation is certainly commonly enough associated with raising of F1 crosslinguistically (Gordon and Ladefoged 2001: 400), though breathiness seems if anything to have the reverse effect. In general the effect of /h/ on neighboring vowels is somewhat schizophrenic. In many languages it patterns with
other gutturals in lowering neighboring vowels, while in other languages in which
gutturals lower or retract vowels, /h/ has no such effect. These patterns are surveyed and
analyzed by Rose (1996). Very briefly put, Rose advances a phonological analysis of
these systems in which vowel-lowering /h/ has a Pharyngeal node in its feature-geometric
representation, while non-lowering /h/ is placeless. The pharyngeal specification is a
consequence of the presence in the vowel inventories of certain other gutturals (uvular
and pharyngeal continuants), with which the laryngeals must contrast. Cases appearing to
be exceptional are argued to involve the spread of [RTR], which /h/ lacks.

If lengthening and sonority enhancement/strengthening are responsibility for
vowel lowering, then this potential instance of phonological weakness may be a
consequence of phonetic prominence. If creakiness or devoicing are involved, though,
then it is phonetic weakening which is responsible. Again, a combination of these may
ultimately be more likely, in which case phonological weakness emerges from a
combination of phonetic strengthening and weakening affecting different aspects of the
production of final vowels.

3.6.5.2. Final Reduction

Laxing or reduction of word- or phrase-final vowels is a commonly encountered
pattern crosslinguistically. While the laxing of high or mid vowels could be considered an
instantiation of the final lowering described above, many of these systems explicitly raise
/a/ to schwa in the same environment, precluding any such explanation. Sometimes descriptions are not specific, as in Biyagó (Atlantic, Bissagos archipelago, Wilson 2000-2001), where word-final vowels are said to be "somewhat lax and can be hard to distinguish". As noted above, if final reduction is identical in both phrase-internal and phrase-final contexts, it might be possible to attribute it to the phonologization for all tokens of the relevant forms of the realizations previously arising only phrase-medially, where the final was indeed short and heavily reduced. Recall the Bulgarian and Brazilian Portuguese durational facts showing the relative weakness specifically of phrase-medial final vowels. This would be difficult to prove, however, without fairly detailed historical documentation of the language, and in any case, it fails to resolve all our problems in this regard, as instances of the weakening specifically of phrase-final vowels, either exclusively or at least to a greater degree than phrase-medial finals are in fact attested. Mentioned already were reports of an especially large amount of reduction of phrase-final vowels in Andalusian dialects (Sanders 1994, 118). In Kawaiisu (Zigmond, Booth and Munro, 1990: 13), word-final short vowels are often dropped, especially when they are utterance-final. Patterns like these are obviously inconsistent with a view of final open syllables as environments for vowel lengthening and strengthening such as that suggested by the articulatory studies and description of Final Strength detailed above. Before addressing possible explanations for these patterns, I present a short catalogue of
reduction processes, neutralizing and non-neutralizing, targeting final vowels in particular. The balance of examples in the literature is heavily skewed toward processes affecting both phrase-medial word-final vowels and phrase-final vowels alike. Though this may indeed be a valid generalization about crosslinguistic patterning of these effects, quirks of the descriptive literature could play some role in creating the appearance of this distribution: there is a strong tendency toward underdescription of effects which are restricted to phrase-final position, which could be the fault of both oversight, conscious restriction to lexical processes, or the reporting of largely phrase-final effects as "sporadic" or "optional" word-final processes instead. Caution is therefore in order here.

Instances of word-final devoicing leading to deletion were treated above in the section on devoicing, and in any case word-final vowel deletion is so common an effect crosslinguistically as to need no exemplification here beyond those examples already mentioned. Often deletion is preceded, however, by reduction of contrasts (perhaps with devoicing as well?). Several potential instances of this are discussed in Blevins and Garrett (1998), in the context of metaphony-like quality transfers from reducing and deleting finals to strong pretonics in Austronesian languages. Exemplary in this regard is the development of final vowels in Nyungar (Pama-Nyungan, Blevins and Garrett 1998: 538-539). In the southwestern dialects of Nyungar, all final vowels are neutralized to schwa (with metaphonic transfer of some features to preceding stressed vowels). In the
southern dialects of the language, this process has been taken a step further, with final schwa disappearing altogether. Other examples of final vowel reduction follow:

(35) Final-only unstressed vowel reduction

a. Highland East Cushitic (Hudson 1976: 250) - In several East Cushitic languages and dialects, particularly Burji, Hadiyya and Kambata, word-final /i/ and /a/ are subject to laxing and devoicing, apparently leading often to deletion, more often in some dialects than in others. Hudson notes that this is particularly the case when the final vowel follows a "strong" syllable, where the final becomes "almost inaudible".109


109 Indeed, Blevins and Garrett (1998: 527-528) also note a link between final vowel loss (with "compensatory metathesis") and strong penultimate stress, implicating in particular "the existence of tonic length in peripheral binary feet, where the peripheral vowel is unstressed". This is particularly interesting in light of the fairly common attestation discussed above of what is essentially the opposite pattern: the tendency in languages with fixed penultimate stress for the vowel of the weak syllable in the final foot to fail to undergo reduction patterns affecting all other unstressed syllables (as exemplified by Seediq, Yakan, numerous Murutic and Dusunic languages, and potentially Shimakonde, Niemirova Polish and even perhaps Neo-Aramaic). The prediction then should be that fixed penultimate languages in which final syllables resist reduction should not have strongly duration-cued stress (at least as concerns the relationship between the syllables of the head foot; obviously some of the languages listed above (e.g. Seediq, with optional deletion of many pretonic vowels) appear to implement a good deal of shortening of pretonic syllables. Empirical testing of this prediction would no doubt yield interesting results.
c. **Iban** (Sea Dayak) (Omar 1981): /u/ has a tendency to unround to [ʊ] in word-final syllables, where some free variation with /o/ also occurs. Potentially neutralizing.

d. **Brok-skad** (Dardic, Sharma 1998): /e/ is realized as [ɛ] word-finally. /a/ is realized as schwa especially word-finally (though also apparently elsewhere), and often elided in rapid speech, "particularly as a case-marker". Non-neutralizing

e. **Maithili** (Jha 2001): unstressed final syllables contain only [i, u, o, ə]. Short [ɛ, æ, ɔ, a] do not appear. Neutralizing.

f. **Kandahari Pashto** (Elfenbein 1997): final unstressed /e/, /o:/ → [i], [u], except, apparently, when this merger would collapse certain grammatical distinctions, such as that between the verbal suffixes pres2sg. -e and 3sg. -i. Neutralizing.

g. **Muinane** (Witotoan, Walton and Walton 1967): Word-final /a/ is sometimes realized as schwa by some speakers. Non-neutralizing.

---

110 See Crosswhite (2001) for other instances of morphological blocking of reduction processes.

251

i. Lunda (Larry Hyman, p.c.): Three vowels /i, u, a/ word-finally, five elsewhere. Neutralizing.

j. Nanai (Tungusic, Avrorin 1959, Li 1996): The Nanai vowel system is not unambiguously described. Li (1996) notes that he may be idealizing to a prior historical stage when he characterizes it as having /i, u, a, i, u, a/, in which the latter three are RTR, produced with a constricted pharynx as in some other Tungusic languages. He assumes that the RTR distinction is still operative due to the fact that the harmony system continues to operate (Li 1996: 147). Avrorin (1959) transcribes Li's /u/ as 'o', calling it high-mid. In general he treats the pharyngealization distinction as one of vowel height, even describing it in terms of jaw aperture in places. That said: In non-initial syllables /i, u/ have a tendency to raise (lessen degree of pharyngeal constriction?) toward /i, u/. In final syllables this frequently results in complete overlap. This is more consistent with /u/ than with /u/, which in some words resists reduction, particularly in shorter words in
which the preceding vowel is not long, and especially if the preceding vowel is also \([u]\).

\(/u/\) is also especially likely to reduce following a long vowel. The prominence of the
initial syllable in Nanai is likely either a relic of earlier Tungusic initial stress (though see
Chapter 4 on initial-syllable prominence in Turkish for an alternative).

3.6.5.3. The phonetics of final syllable reduction

It is necessary to note once again that in virtually all examples above the domain
of the application of reduction is said to be "word-final". At first consideration, this could
mean one of three things: 1. The language has no final lengthening or strengthening, and
final syllables are simply all phonetically weak, 2. The pattern began in phrase internal
final vowels and was generalized to phrase-final vowels in the manner proposed above
for some speakers of Brazilian Portuguese, or 3. The author of the description has noticed
word-final reduction phrase-internally and failed to notice or point out failure to reduce
phrase- or utterance-finally. This is especially likely in the gradient reduction cases like
Nanai above. Neither of the latter two hypotheses will be simple to confirm without an
extended research program.

\[\text{See the note above on the propensity for final vowels to reduce following long penults. Provocatively, however, Nanai has fixed final stress, which is not, however, duration-cued (Avrorin 1959, 62-3), apparently being manifest primarily by raised } F_0\text{ (Note discussion of final stressed vowel devoicing in Turkish under similar circumstances. On the failure to reduce following an identical vowel, see the discussion in Chapter 4 below.)}\]
The first hypothesis above, of course, is quite plausible, given the extremely small number of detailed articulatory studies of phrase-final effects and the near total concentration of these on one or two West Germanic languages. It may well be that final lengthening, however, universal in its existence, nonetheless varies not only in its domain and degree of application crosslinguistically, but in its articulatory implementation as well. Certainly some languages, such as Creek (Johnson and Martin 2001) discussed above, have final lengthening that does not impede normal phonetic vowel reduction. The presence of articulatory strengthening in the implementation of final lengthening may then be a variable set differently from language to language. Of course, to account for reports of degrees of reduction targeting specifically phrase-final syllables, we would have to assume as well that in some languages lengthening (gestural slowing) is implemented with actual gestural weakening, rather than mere absence of strengthening. This is of course also possible, and without further articulatory studies it may be impossible to resolve this question satisfactorily. Certainly if the question is "Do some languages fail to strengthen, and indeed even reduce phrase-final vowels, regardless of whatever effect phrase-final lengthening may or may not have?", then the answer is clearly yes, to the extent that reports are correct. We should be cautious, however, before we take this to mean necessarily that this language has no final lengthening, or that final lengthening in this language is non-strengthening, in light of the results of Cambier-
Langeveld's 1997 study of final lengthening in Dutch, where it was shown that when the final vowel of a word was schwa (and hence potentially resistant to lengthening), final lengthening simply took place over a larger domain extending back from the final syllable. A careful study of a language reporting particularly great reduction of phrase-final vowels could well discover the same pattern, that not only did the language have final lengthening (just retracted from the reduced final vowel), but that it was even strengthening the articulation of the material it applied to as well.

If the latter were true, however, we would legitimately wonder why the final vowel was reducing in the first place, perhaps even to the point of reducing a five vowel system to the three vowels common in the unstressed vocalism of the languages of the world. The following is a potential answer to this question: It is clear from the discussion of the phonetic properties of final syllables above that, whatever properties their supralaryngeal articulations may take, other properties are contributing to a general lack of perceptual robustness for material realized in this position. It was argued in the discussion of devoicing, above, that if the subglottal pressure drop associated with phrase-final vowels is strong enough that they are routinely being realized as all or partly voiceless, no amount of supralaryngeal "sonority enhancing" tongue-body lowering, lip opening or gestural slowing will be effective in creating the conditions for the accurate
perception of the quality of the final vowel. The following example is highly suggestive in this regard.

In Standard Malay as described by Teoh 1994, final syllables license no contrast between mid and high vowels, while elsewhere such a contrast exists. Realization of final syllable vowels, however, varies: in final open syllables, the three possibilities are /i, u, a/, while in final closed syllables, there are only [e, o, a], where [e] and [o] are actually realized somewhere between mid vowels and laxed high vowels. Certainly closed syllable laxing/lowering of high vowels is common enough in Austronesian languages of the area, but it is usually accompanied by raising of the low vowel as in, e.g. standard Indonesian (Macdonald and Darjowidjojo 1967). Here, however, in the final closed syllable the /a/ stays low. A possible, though strictly hypothetical solution is this: contrasts between mid and low vowels were neutralized in all final syllables, but with the results varying depending on syllable structure. In final closed syllables, we have an example of the type of final lowering discussed above, possibly an effect of final lengthening and sonority-enhancement. This may or may not have originally affected final open syllables as well. Final open syllables were then (either subsequently or from the outset) subject to the type of phonetic attrition discussed above in connection with devoicing.
That a combination of features affecting final vowels such as a tendency to total or partial devoicing, a drastic amplitude drop, and various instantiations of non-modal phonation types could lead to a reduction in the perception of vowel quality contrasts for final vowels (while vowels in closed final syllables might be shielded somewhat from these effects due to increased distance from the phrase-boundary) is not difficult to conceive. Some combination of various of the above should easily be enough to obscure vowel distinctions and cause reanalysis and collapse of contrasts. But why raising? Certainly breathy phonation has been associated with raising in some languages (Gordon and Ladefoged 2001: 400). Another possibility, however, one that would account for both reduction through raising and laxing of phrase-final vowels suggests itself. Take the following spectrogram, of the Russian word [belóte], 'swamp', in phrase-final position.
Again, Russian unstressed /a, o/ is realized anywhere from [e] to [o] in phrase-final position, depending on speech rate and register (i.e. on duration). In this particular instance, the vowel is significantly lengthened (particularly in comparison with a non-phrase-final vowel of this sort), measuring nearly 180 ms from onset of voicing to the last reasonable glottal pulse. The first pretonic vowel (another strong position) is only 115 ms in duration. The F1 maximum for the final vowel shows that it is in the Degree 1 reduction range, as discussed above: 595 Hz. (The first pretonic vowel here has an F1
maximum of 556 Hz., close to the mean for this speaker - see section 3.6.1. above and the
experimental results in Chapter 2). The pretonic vowel, however, does sound
impressionistically at least more 'a'-like than the final. An important reason for this could
be that, in addition to the fact that the amplitude of the pretonic and stressed vowels is
substantially higher than that of the final vowel throughout, the amplitude of the first
pretonic vowel is more or less equally high throughout its duration, including, obviously
at its center where the F1 maximum occurs.

For the final vowel, however, this is not the case. The amplitude of the final
vowel decreases steadily throughout, becoming extremely low and ultimately fading to
nothing by the end. What is more, the peak F1 for the final vowel occurs relatively far
into its course, approximately 90 ms into the vowel. By this point, of course, the
amplitude of the vowel has fallen considerably, and only gets lower. With the application
of final lengthening, the vowel has become extremely long by comparison, and assuming
that gestural stiffness is lowered (meaning slowing), but target displacement is not
reduced, this means that the time to peak displacement will have increased significantly.
What seems to be occurring, in other words, is that the "sonority enhancement" discussed
by Cho (2001) is taking place, but the higher F1 peak it produces is coming quite late
relative to the gradually decaying amplitude of the vowel. This has the effect of making
the part of the vowel with the highest F1 cooccur with the part in which perceptibility is
drastically decreased, as an already lowered amplitude continues to drop. The most
perceptible and least distorted (by non-modal phonation) part of the vowel, on the other
hand, is at the very beginning, during the lengthened CV transitions, long before the
vowel reaches its target articulation. This unfortunate phasing relationship between the
supralaryngeal target of the vowel and the point at which the laryngeal and aerodynamic
weakening of the final vowel is at its lowest could cause listeners to perceive the height
of the earlier part of the vowel as the intended target, with reinterpretation in this instance
leading to raising. For high vowels, by contrast, the same effect could lead to
centralization or laxing.

If this scenario took place in the final open syllables of Standard Malay as
described above, it could explain the raising these final vowels in particular undergo. If
this masking of the "strengthened" target articulation by phonetic weakness of another
sort is enough to lead to phonologization, the result could be similar to that described
above for Dutch: Once the percept of laxing or raising gave rise to the phonologization of
an actual target articulation of the final vowel that was laxed or raised, the durational
characteristics appropriate to the resulting vowel would be realized here as well, final
lengthening or not (just as described above for newly devoiced finals, or for the Dutch
schwa). In this new situation, phrase-final vowels might actually be shorter and more
reduced than their phrase-medial counterparts, though in the original situation, of course,
they were not. The nature of final lengthening and the potential gestural strengthening of the affected segments, however, would not necessarily have changed. As demonstrated for Dutch, if the vowel in phrase-final position is inherently short to the point of resisting lengthening, lengthening might simply be felt further back in the word as its unaltered domain span came to include more of the word due to the shortness of the final segment.

It is possible, then, that reduction patterns targeting specifically phrase-final syllables are actually the result of a phonologization caused by the perceptual weakness of those syllables, and not necessarily by any original weakening of supralaryngeal gestures in that position. If the perceptual weakening story is true, then, as with devoicing, in the pre-phonologization state of the system, there is no problem with assuming final lengthening to have applied to unstressed final syllables with articulatory strengthening, just as it has been shown to do in numerous examples laid out in section 3.2.3. Obviously only experimental evidence can ultimately resolve this question.

This idea of final reduction patterns originating in the phonologization of the perceptual weakness of the final syllable, rather than in decreased magnitude of supralaryngeal gestures, may account for another trend found in final syllables in some languages: monophthongization. In most dialects of Seediq, for example, word-final diphthongs /aw/ and /aj/ become /e/ and /o/ (Holmer 1996: 25)\(^\text{112}\). If under final...

\(^{112}\) Which are both realized as [u] when word-final today due to the vowel reduction described in Chapter 2.
lengthening the latter parts of these diphthongs were realized in such circumstances that their accurate perception were to be hindered, a listener might perceive only the coarticulated first half of the diphthong, and believe it to be the intended production of the speaker, a syllable-internal umlaut of sorts. This phonologization of a raised and fronted, or backed, raised and rounded /a/ could yield the mid vowels attested today.

3.7. Vowel Quantity in final syllables

3.7.1. Neutralization of quantity contrasts in final position

Several instances of processes affecting vowel quality have already been discussed in this chapter, particularly the instances of mora-clipping, either through devoicing or deletion, and the deployment of features like devoicing and glottalization in patterns that tend to stave off the collapse of quantity contrasts by inhibiting the phonetic processes which work against the persistence of those contrasts. For completeness sake, and due to the relevance of certain of these patterns to wider concerns in phonological theory (e.g. in metrical stress theory) I will now present a survey of cases in which quantity contrasts are in fact neutralized word-finally by one means or another. The situation turns out to be more complex that has previously been suggested in the literature.
Languages in which there is a contrast between phonologically long and short vowels in all positions save word-final are extremely common crosslinguistically. Final position, indeed, can safely be pronounced "weak" for the licensing of quantity contrasts, all things being equal.

Examples include Yukulta (Pama-Nyungan, Keen 1983), Koyra (Omotic, Hayward 1982), Dasenech (Cushitic, Sasse 1976), Biyagó (Atlantic, Wilson 2000-2001), many Dravidian languages, including Gadaba (Bhaskararao 1998), Kolami (Subrahmanyam 1998), Konda (Krishnamurti and Benham 1998) and Dhangar Kurux (Gordon 1976), Gatha Avestan (Testen 1997), and Tigre (Semitic, Raz 1983). In most cases, the lack of contrast concerns phrase-medial and final syllables alike, and the resulting vowel is said to be short, presumably varying in actual duration according to position like any other vowel. In some instances, however, the vowel is said to be long. In Gatha Avestan (Testen 1997), for example, all final vowels are written as long (though this is obviously no guarantee as to their pronunciation). In Kolami, Subrahmanyam describes final vowels as long uniformly. Still other cases speak of final vowels which are half-long. These include Dhangar Kurux (Gordon 1976) and Maltese (Borg and Azzopardi-Alexander 1997), in which latter short vowels all deleted, and remaining long vowels shortened, but are apparently realized longer than they would be word-internally.
Avoidance of length contrasts word-finally takes another form as well: the well-known propensity for vowels in final syllables to evade a process known as Iambic Lengthening (Hayes 1995, Hung 1994, Buckley 1998). Iambic lengthening is a process whereby an underlying short vowel, if parsed into the strong position of an iamb, is lengthened. Choctaw is often cited as exhibiting this pattern. Choctaw (Western Muskogean, as given in Hung 1994, Hayes 1995 and Buckley 1998) parses words into iambic feet, lengthening the vowels in the strong position of these:

(37) Final lengthening in Choctaw

<table>
<thead>
<tr>
<th>Word</th>
<th>Transcription</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>habina</td>
<td>(habi:)na</td>
<td>'s/he receives a present'</td>
</tr>
<tr>
<td>tʃi-habina-li</td>
<td>(tʃiha:)(bina:) li</td>
<td>'I receive a present from you'</td>
</tr>
<tr>
<td>pisa-li</td>
<td>(pisa:)li</td>
<td>'I see'</td>
</tr>
<tr>
<td>tʃi-pisa</td>
<td>(tʃi.pi:)sa</td>
<td>'s/he sees you'</td>
</tr>
<tr>
<td>tʃi-pisa-tʃi-li</td>
<td>(tʃi.pi:)(satʃi:):li</td>
<td>'I cause you to see'</td>
</tr>
<tr>
<td>tokwikili-tʃi-li</td>
<td>(tok)(wiki:)(litetʃi:):li</td>
<td>'I shine a light'</td>
</tr>
</tbody>
</table>

Final vowels, however, fail to lengthen in this position.

<table>
<thead>
<tr>
<th>Word</th>
<th>Transcription</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>habina-li</td>
<td>(habi:)(nali)</td>
<td>'I receive a present'</td>
</tr>
<tr>
<td>habina-tʃi</td>
<td>(habi:)(natʃi)</td>
<td>'s/he gives a present'</td>
</tr>
<tr>
<td>tʃi-habina</td>
<td>(tʃiha:)(bina)</td>
<td>'you receive a present'</td>
</tr>
<tr>
<td>tʃi-habina-tʃi-li</td>
<td>(tʃiha:)(bina:)(tʃili)</td>
<td>'I give you a present'</td>
</tr>
</tbody>
</table>

Choctaw, like the languages cited above, has no long vowels derived or otherwise in final position (Buckley 1998 citing Nicklas 1975). For Hung, this state of affairs
results from the interaction of a constraint she calls Rhythm with a constraint mandating that all heavy syllables be stressed - Weight-to-Stress (after Prince 1990) and an OT formalization of Iambic Lengthening called Iambic Quantity (the 's' side of a (w s) foot should be heavy). Rhythm is a constraint mandating that every stressed syllable be followed by any unstressed syllable. Among the situations this constraint can treat are the avoidance of stress clash and the avoidance of final stress. Hung's analysis can be summarized briefly as follows: Rhythm demands that final syllables not be stressed. Weight-to-Stress demands that heavy syllables be stressed, and Iambic Quantity demands that metrically strong syllables be heavy. Thus, parsing the penult and final of a word into an iambic foot would require lengthening of the vowel (assuming other routes to heaviness to be blocked by Faithfulness constraints) because of Iambic Quantity. Lengthening the vowel would demand that the vowel be stressed (Weight-to-Stress), but doing this would violate high-ranked Rhythm. Assuming a certain ranking of constraints on parsing, the solution is to avoid lengthening by not parsing the last two syllables into a foot. Rules of final shortening and the general absence of long vowels in final position can be treated similarly. Again depending on the rankings of Faithfulness constraints, it

---

1. Recalling Hyman's discussion (Hyman 1977) of the crosslinguistic frequency of fixed penultimate stress and the insight that metrical prominence may be in some way more perceptible if it has metrical non-prominence following.
will sometimes be optimal to shorten an underlying long vowel rather than to violate Weight-to-Stress by leaving it unstressed, or Rhythm by stressing it.

While the non-finality of stress approach can account for some of the problems found crosslinguistically with vowel length in final syllables, it cannot account for them all, and while it seems clear that independently of these phenomena there is a demonstrable crosslinguistic pattern of avoiding the parsing of final syllables for metrical purposes, this is not the whole story here. First, non-finality refers to a tendency to avoid stressing vowels in final syllables of any kind, open or closed. As Buckley 1998 argues, there is a set of phenomena connected specifically to the vowels of final open syllables, not all of which non-finality approaches can address.

In addition to the failure of Iambic Lengthening, Buckley draws attention to other cases in which lengthening processes fail to apply to word-final vowels. One of his examples is Italian, in which the vowels in stressed open syllables lengthen (closed syllables being already bimoraic), unless stress is final. Thus we have (38a), but also (38b):
When these stressed open final syllables are followed by a consonant-initial word (subject to prosodic restrictions), however, the preference for a bimoraic stressed syllable resurfaces through a process known as *raddoppiamento sintattico* (Buckley cites Nespor and Vogel 1986: 165), whereby the need for a bimoraic stressed syllable results in gemination of the initial consonant of a following word, closing the final stressed syllable. This is shown in (39).

(39) Raddoppiamento sintattico (Nespor and Vogel 1986: 165)

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>cosi buono</td>
<td>[kɔzibbwɔno]</td>
</tr>
<tr>
<td>b.</td>
<td>caffe nero</td>
<td>[kaffɛnɛro]</td>
</tr>
<tr>
<td>c.</td>
<td>resterà con me</td>
<td>[resterákkommé]</td>
</tr>
</tbody>
</table>

This is obviously a problem for the non-finality approach, since the syllable in question is in fact stressed (violating Rhythm, which was previously satisfied by avoiding final
length), but nonetheless refuses to lengthen. An example similar to the Italian is that of Kuku-Thaypan (Paman, Rigsby 1976), in which stressed vowels are lengthened in closed syllables and open medial syllables. In word-final open syllables, however, stressed vowels do not lengthen, but rather, a glottal stop is inserted after them. One more similar system is Lithuanian (Blevins 1993, Kenstowicz 1972), in which a rule lengthening /a/ and /e/ under accent fails to apply word-finally. The problem here is the same: Since the final vowel is already stressed, there is no need to avoid lengthening it. Buckley also points out the case of Luganda (Hyman and Katamba 1990, 1993), in which compensatory lengthening caused by glide formation fails to apply where it would result in a word-final long vowel. It is not clear how avoidance of final stress or metrical invisibility of any kind could be invoked here, since Luganda is not generally analyzed as having a stress, final or otherwise (see Hyman and Katamba 1993 specifically for an overview and analysis matters concerning tone and accent in Luganda).

It is Buckley's thesis, instead, that in addition to constraints mandating avoidance of stress on final syllables, languages have a more general tendency simply to avoid final long vowels. It is this tendency which motivates both the failure of Iambic Lengthening word-finally and a variety of other processes. Buckley attributes this tendency to "final lengthening at the phonetic level" (183). If a language lengthens vowels finally (in some prosodic constituents at least), this could result in the contrast between underlyingly long

268
and phonetically-lengthened short vowels becoming difficult to perceive, and thus potentially subject to reanalysis and the collapse of the contrast.

Certainly final lengthening could have the effect of making the contrast between long and short difficult to perceive, especially if other phonetic properties of final syllables identified here were simultaneously weakening the perceptibility of the latter parts of any vowels being realized there, such that both long and short vowels would sound like a lengthened vowel that fades away toward the end. That the vowels merge ending up generally as short could be seen as the result of a hypercorrective change, whereby the listener, hearing additional length in final position, attributes that length to the final lengthening (wrongly, in the case of underlying long vowels) and factors it out in his or her interpretation of the speaker's intentions, reconstructing a final short vowel in all cases.

That more must be involved in the tendency to avoid a long/short contrast in final position than merely phonetic lengthening becomes clear when we compare potential fates for quantity contrasts here with those found in stressed syllables, as discussed in 2.1. As noted, some languages collapse quantity contrasts in stressed syllables through lengthening of short vowels. This durational enhancement, however, invariably results in neutralization to a long vowel. This is of course unsurprising given the phonetic processes commonly associated with stressed syllables, the phonologization of which
often results in grammars requiring stressed syllables to be bimoraic, or to avoid low
sonority vowels like [ə] or [i] (which latter also occurs in final syllables - see below). In
short, there is nothing about stressed syllables per se which would ever make them bad
positions in which to realize a long duration vowel. If anything, the additional duration is
beneficial, serving to enhance the contrast between stressed and unstressed syllables.
There are also languages, however, in which only the stressed syllable licenses the
contrast between long and short vowels. The implication here is that while unstressed
syllables are apparently required to be too short to contain long vowels, short vowels in
stressed syllables are not lengthened so much as to endanger the long/short contrast.

In final syllables, the situation is different. That these are never the sole position
in which a quantity contrast occurs falls out from the fact that, while unstressed syllables
may be durationally compromised or restricted relative to stressed syllables in some
systems, it is never that case that along with final lengthening the durations of non-final
syllables are actually curtailed or limited. This difference must be due to the lexical
nature of the stressed/unstressed contrast, as opposed to the largely phrase-level operation
of the final lengthening process.

But why, when final lengthening does cause neutralization of long/short contrasts,
is the resulting vowel so often phonologized as short or "half-long", as some descriptions
put it? Buckley's answer is that short is the "unmarked" member of the opposition, to
which we therefore default. But this seems rather unsatisfying as explanations go. By now we clearly recognize that "markedness" of features or segment types is often not absolute, but is rather comprehensible only relative to the environments within which the features in question are being realized (Steriade 1994, 2001). In stressed syllables, obviously, the perceptual confusion caused by lengthening short vowels does not result in default to the "unmarked" short member of the opposition\(^\text{114}\). Of course, it should not, since stressed syllables are a lengthening context, in which short vowels may sound long, but nothing about long vowels is likely to sound short. For a listener to simply decide, then, that the phonetically manifestly long stressed vowels were in fact intended to be short because this option is "unmarked' is incomprehensible. But final position is also a lengthening context, and, absent any other conditioning factors, such a decision on the part of the listener seems no less unwarranted. There, must, therefore, be something else at work here.

Again, while lexically stressed syllables will express their additional duration in all tokens of the string in question, final syllables will generally acquire significant length only prepausally, where they are also characteristically weakened in some respects as well. This weakening, primarily the result of a sharp drop in subglottal pressure in

\(^{114}\) Speaking, of course, of phonetic lengthening of short vowels in stressed syllables in a system prior to the phonologization of any constraint requiring that stressed vowels or syllables be bimoraic.
phrase-final position, is not a characteristic of stressed syllables, and may help account for differences in the quantity neutralization patterns found in the two positions. The logic is this: in stressed syllables, phonetic realizations of both long and short vowels prior to neutralization may come to be, let us say for the sake of argument, in the neighborhood of 200 ms and with steady amplitude throughout. Their similarity in this respect would obscure the contrasts between long and short and cause neutralization in favor of the longs. In final syllables, by contrast, both long and short vowels might come to be 200 ms or so in duration, but with a significant amplitude drop, possibly together with non-modal phonation or devoicing toward the end. They have thus also come, as it was with stressed syllables, to be so similar that the contrast between them is in danger of collapse. But here there is a problem. If word-internal long vowels are, say, also 200 ms long, but with steady amplitude and no devoicing or the like, and the listener ordinarily factors out or corrects for some additional duration in phrase-final position as the result of domain-final lengthening, final vowels would sound less like true long vowels (which should be long and steady), and more like short vowels that had some duration added to mark the phrase boundary, but which nonetheless sound weaker in some respects than a real long vowel ought to. The full complexity of the typological picture is thus only comprehensible when the full range of phonetic effects characteristic of final position is taken into consideration.
While Buckley's notion of a general tendency to avoid length in final syllables owing to the phonetic trend of phrase-final lengthening does help to explain the failure of a variety of lengthening processes word-finally, including Iambic Lengthening, it is less clear what we should make of the more general tendency for avoidance of final stress crosslinguistically with or without Iambic Lengthening playing a role. For Hung this is the effect of the constraint Rhythm, demanding an unstressed syllable following every stressed one. A phonetic lengthening trend in final position, on the other hand, should have the opposite effect (if any). Insofar as duration is a frequent crosslinguistic cue for stress, we might imagine that final lengthening would actually attract final stress, rather than the reverse. Hock (1999: 22-24), by contrast, discusses accent retractions from final syllables (in Baltic in particular) as a function of the monolithic final weakening he proposes. Essentially, he views accent retraction as a consequence of some degree of "segmental attrition" in final syllables, in some cases leading to the deletion of final vowels (as in certain Lithuanian dialects, and Latvian as noted above). Given the above discussion concerning the phonetics of domain-final syllables (assuming the retraction trends to have originated in larger domains, and not phrase-internally), whatever segmental attrition may be giving rise to accent retractions is most likely the result of the final drop in amplitude and F0 (possibly with attendant devoicing or creak), which have been seen to take place even in the presence of substantial final lengthening. A potential
prediction arising from this characterization of final position is that languages with strongly duration-cued stress (such as Russian) should be less subject to stress retraction or final stress avoidance, these latter being more characteristic of languages in which stress is cued more saliently by, e.g. F0 contours. Crosslinguistic investigation of the correlation of accent type and non-finality (particularly in Balto-Slavic) may yield telling results in the future.

3.7.2. Final short vowel avoidance

As much as Buckley is correct in seeing a crosslinguistic tendency for languages to avoid contrastively long vowels in word-final position, and despite the fact that at least some languages (e.g. Dutch) clearly have the means of implementing final lengthening even in cases where the final vowel itself is resistant to the process, there is nonetheless an equally clear tendency for languages to avoid certain short vowels in word-final position as well. Johnson and Martin (2001) note in passing that Korean, Tigrinya and Tigre all disallow final [i], while Moroccan Arabic prohibits final [ə]. To these examples we may add the following:
(40) Avoidance of final short vowels


c. **Punjabi** (Indo-Aryan, Bahl 1957): Short vowels /i, u, ə/ do not appear before pause, where they become long.

d. **Thaadou** (Kuki-Chin, Thirumalai 1972): The mid central vowel /ə/ does not appear in word-final open syllables.

e. **Proto-Austronesian** (Blust 2000): No schwa in word-final open syllables. Among the many modern Austronesian languages exhibiting this trait are Batak (Dairi dialect, Sumatra, van der Tuuk 1971), Cotabato Manobo (Mindanao, Kerr 1988), Iban (Sarawak, Omar 1981), and Buginese (S. Sulawesi, Sirk 1983).
The pattern here is immediately evident. In all cases the short vowels disallowed are either high or mid central vowels, renowned for their brevity, or the set of "short", "lax", or "reduced" vowels in systems where the long/short contrast is one in which long or full vowels are not in fact particularly long, but are peripheral, while the short vowels are centralized and often characterized as "over-short". Thus, in Mari for example, Crosswhite cites Gruzov's 1960 instrumental study of Standard Mari (and Volga Mari) in which the average duration for the reduced vowel in standard Mari pretonic syllables is somewhere in the neighborhood of 50-60 ms, depending on context, while full vowels are generally between 95 and 125 in the same environments. What seem to be disallowed in final syllables crosslinguistically, then, are these over-short vowels. In systems where short vowels too are peripheral and long vowels are described as more or less twice the length of short vowels (canonical monomoraic-bimoraic oppositions, as in the Dravidian languages noted above), by contrast, the long/short opposition seems generally (though perhaps not exclusively) to be neutralized in favor of the short vowel.

3.8. Summary

In this chapter I have presented a survey of the phonetic and phonological patterns common to domain-final syllables and, where possible, I have explicated the relationship between the former and the latter. Final syllables were shown frequently to host patterns
both of phonological strength and phonological weakness. Strength effects in final position were of a sharply limited character; final syllables rarely if ever function as the sole center of prominence in the word domain in the way that, say, stressed syllables, frequently do. When they are strong at all, this strength is generally manifested as a pattern of Final Resistance to some other reduction or assimilation process (such as UVR) which would otherwise have targeted a syllable of the prosodic status in question. This difference between final strength and UVR was attributed to differences in the consistency and magnitude at the word level of the durational asymmetries giving rise to the patterns. Final resistance was shown to be often phrase-level only and gradient in its operation. It is also targets in the vast majority of cases only vowels in open final syllables, a fact following directly from phonetic details concerning the domain of application of final lengthening. A non-phonetically-oriented approach to PN patterns in final position has no way of accounting for any of the typological generalizations drawn here.

As for phonological weakness, neutralization patterns targeting specifically domain-final material were shown to exist, a fact which is wholly contradictory from the point of view of the foregoing discussion of final strength and phonetic enhancement. In addition to these trends, however, final position also hosts a number of phonetic patterns, the effect of which is ultimately a decrease in the perceptual robustness of domain-final
material. These patterns include drops in pitch and intensity correlated with lowered subglottal pressure, devoicing, non-modal phonation such as breathiness or creak, and even potentially nasalization. Prominent instantiations of various combinations of these tendencies can interfere with the accurate perception of vowel contrasts in domain-final position, leading to the phonologization of neutralizations specific to that position. This combination of factors contributing at once to increases in phonetic prominence and perceptual obscurity is the source of the phonological ambiguity of final position with respect to the typology of positional neutralization. Because of this, final syllables cannot be considered from a crosslinguistic perspective either monolithically strong or monolithically weak in terms of their potential to license vowel contrasts. Indeed, they may exhibit this schizophrenic licensing potential even within a single system, such as in Nanai, where final syllables are strong for the licensing of nasality, but weak for pharyngealization contrasts. This is a serious challenge to theories of positional neutralization which assume that the phonological strength or weakness of structural positions is specified in Universal Grammar. At best strength or weakness in final position is a parameter which must be set on a language-specific, inductively-determined basis. This fact about final syllables also gives us reason to question whether it is truly necessary for strength or weakness to be predetermined by UG in any other position either. Given that patterns of strength or weakness will arise in each position through the
phonologization of phonetic patterns occurring there independently of phonological licensing restrictions, it is not clear what such restrictions in Universal Grammar are contributing to the situation. Where phonetic patterns are consistently oriented toward a single perceptual profile, robust or obscure, phonological patterns in that position will be equally uniform (as in, e.g., stressed syllables). Where the phonetic picture is more equivocal, however, phonologization will yield a typological range of patterns which is equally ambiguous, as the facts of final syllables have made clear.

Furthermore, the relationship of phonetic prominence to phonological strength is not a simple one. Patterns of phonological strength may arise which bear no obvious relationship to the quality of the phonetic cues for that contrast in question in position (laxness in Pasiego Spanish, nasalization in Nanai, final dominance in Timugon Murut), and likewise contrasts may collapse due to the addition of properties such as duration, which are otherwise thought to enhance the perceptibility of the contrasts in question (Smith's Positional Augmentation - here potentially instantiated in final lowering). It is by no means obvious, however, that the synchronic status and implementation of a pattern of phonological strength that is not supported by the robustness of its characteristic cues in final position (Nanai nasalization, Timugon final-syllable super-licensing, Pasiego laxing) is in any way different or less authentic than a pattern that is
supported by such phonetic cues. I will take up the implications of these complications for theories of positional neutralization again in Chapter 5.
Chapter 4. Initial Syllables

Domain-initial syllables are famously associated in the literature with phonological strength effects. For vowels, the best-known of these effects are the harmony systems of the Uralic, Altaic, and Bantu languages. The first striking fact about initial-syllable PN patterns affecting vowels is the wide variety of featural contrasts potentially involved. Chapter 2 demonstrated that PN patterns involving stressed syllables were severely limited in the set of contrasts they potentially restrict: the overwhelming majority of cases involves vowel height, while nasalization and quantity may also be affected. In initial syllables, by contrast, PN patterns are attested involving vowel height (Bantu, Gujarati), frontness/backness (Uralic and Altaic), roundness (Uralic and Altaic), nasality (Gokana, Kurux), pharyngealization (Tuva, Tungusic), and breathiness (Gujarati). Thus, while in unstressed vowel reduction the driving force behind most attested was pressure on the articulation of non-high vowels due to the durational impoverishment of unstressed syllables, in initial syllables clearly the phenomenon leading to the phonologization of PN is something far more general in its effect.

The second striking fact about the positional neutralization of vowel contrasts in initial syllables are their comparative rarity. I show in this chapter, in fact, that the vast majority of cases cited in the literature as instances of initial syllable PN can in fact be attributed to past or present fixed initial stress in the languages in question. Additionally,
in the relatively small set of cases remaining when those are factored out, all attested PN systems involve vowel harmony, rather than, for example, the types of surface inventory reductions characteristic of UVR systems.

In this chapter I contrast two approaches to initial syllable PN effects. One of these, the phonologization approach familiar from the preceding chapters, attributes typological regularities found in PN systems to the phonologization of patterns arising originally in the phonetics. The second approach locates the force generating typological patterns squarely in Universal Grammar. In the case of initial position, it is the psycholinguistic status of initial material which is responsible for the attested patterns. It is the initialness of the material itself, in other words, rather than any concrete attributes of that material, which is responsible.

I will argue in this chapter that neither the startling rarity of initial-syllable PN affecting vowels, nor the apparent limitation of true instances of this PN to vowel harmony follow from the initialness approach. Both these patterns, however, can be understood as consequences of the phonetic characteristics commonly associated with initial syllables crosslinguistically. The chapter is structured as follows: Section 2 presents a discussion of the most important recent literature on initial-syllable effects. Section 3 introduces the phonetic processes commonly associated with initial syllables. Section 4 begins an analysis of initial-syllable vocalic strength effects in light of what we
know of the relevant phonetic facts, while section 5 presents the results of an experimental study of vowel durations in Turkish demonstrating the possibility of a non-accentual origin for additional duration on the vowels of initial syllables, and sets forth a model for how this additional duration could contribute to the phonologization of the harmonic strength effects associated with that position.

4.1. Recent work involving the phonology of the initial syllable

The most comprehensive recent treatments of phonological licensing patterns involving initial syllables are to be found in Beckman's 1998 study of Positional Faithfulness and Smith's 2002 study of Positional Augmentation. These present catalogues and important analyses of initial-syllable strength effects for both consonants and vowels. Beckman treats only Positional Faithfulness or phonological strength effects in initial syllables, while Smith is concerned with both these and phonological weakness effects. Smith refers to restrictions on the licensing of contrasts in otherwise phonologically strong positions as Positional Augmentation, since in all cases these stem from phonetic strengthening processes charged with increasing the perceptual prominence of the position in question. She contrasts Positional Augmentation effects with phonological strength effects in these same positions, which she calls Positional Neutralization effects, though in the now-generic usage derived from Steriade 1994 both
patterns are examples of the latter. To avoid confusion with terminology used in earlier chapters here, I will substitute Positional Strength for Smith’s Positional Neutralization. The issue is purely practical and should not be taken to bear theoretical significance. One example of a Positional Augmentation pattern in initial syllables is the relatively common demand for low-sonority onsets in initial or stressed syllable onsets, which has the result of causing fewer consonantal contrasts to be licensed in those positions. Another would be the demand for long or high-sonority vowels under stress, discussed in Chapter 2 and also in Chapter 3 in the context of final vowel lowering.

4.1.2. Previous proposals for the motivation of PN in stressed and initial syllables

Both Beckman 1998 and Smith 2002 divide strong positions into the phonetically strong (i.e. stressed syllables), and the psycholinguistically strong (i.e. initial syllables, roots). For Beckman little follows from this classification. Smith pursues this dichotomy further and examines the potential effect the source or nature of a position’s phonological strength may have on the typological patterning of the licensing asymmetries found therein. Smith’s prediction is essentially the following: a position which is phonetically-strong may show both Positional Strength and Positional Augmentation effects, since phonetic strengthening can have the effect of both exempting a position from lenitions or reductions (as with final resistance to VR) and also itself of effacing contrasts by, e.g.,
lengthening or lowering the sonority of vowels. Psycholinguistically-strong positions, on the other hand, will according to Smith show only a restricted inventory of Positional Augmentation effects, since the importance of these positions for processing makes it crucial that as many contrasts as possible be realized faithfully there\textsuperscript{115}.

It is Smith's contention that initial position shows precisely these characteristics crosslinguistically, such that while instances of initial syllables as sole licensers of a language's full inventory of contrasts are common, both for consonants and vowels, instances of Positional Augmentation are severely restricted. Specifically, Smith claims that the only such processes one finds are related to the presence or sonority of onset consonants in initial syllables. This is expected according to a principle Smith formulates as the Segmental Contrast Condition, which mandates (to oversimplify somewhat) that segmental contrasts not be neutralized in a way that would be detrimental to early-stage processing.

\textsuperscript{115} Smith also presents a rigorous delineation of the type of psycholinguistic prominence that should be involved in determining aspects of the typology of licensing potential. Specifically at issue are positions which have been shown to be important for early-stage word recognition (The relevant discussion is in Smith 2002: 254-306). This is a first stage hypothesized in most models of word recognition in which, in Smith's words, "phonetic/phonological information is used to identify a set of candidate lexical entries for further examination" and which is followed by "a later stage, in which the selected set is narrowed down (often on the basis of more than just phonetic or phonological information) until the best-matching lexical entry is identified" (Smith 2002: 257). There is an abundance of strong evidence, surveyed by Smith, demonstrating the importance of word-initial material in precisely this aspect of speech processing. Smith cites further evidence showing that stressed syllables do not exhibit the same importance in this stage of processing. Recalling the discussion of the psycholinguistic status of the final syllable in Chapter 3, if it is indeed solely significance to early-stage word recognition which has the potential to affect the status of a position as a licenser of contrasts, then we might look to this as an explanation of the seeming irrelevance of such information in determining the typology of licensing asymmetries in final position.
word recognition, unless those neutralizations themselves contribute to segmentation of
the speech stream (e.g. by providing cues for the presence of word boundaries). Thus, the
requirement for low-sonority word-initial onset consonants fulfills this latter goal, and
hence is allowed to neutralize, e.g., the voicing contrast for obstruents, even though doing
so is unhelpful from the point of view of word recognition. According to Smith, these are
the only contrast-neutralizing effects found in initial syllables, with the types of
Positional Augmentation phenomena affecting, for example, the vowels of stressed
syllables being conspicuously absent.

The typological regularities Smith seeks to capture for initial position using the
psycholinguistic- vs. phonetic-strength dichotomy are, to summarize, these: Word-initial
onset consonants are seen both to undergo neutralizations involving the lowering of their
sonority, and also to preferentially license more contrasts than the onsets of other
syllables within the word. For vowels, on the other hand, initial syllables frequently
license more contrasts than other positions, but no neutralizations resulting from phonetic
augmentation for the sake of perceptual prominence are seen to occur. This follows from
the classification of initial syllables as psycholinguistically, rather than phonetically
strong. If we assume with Smith (though incorrectly, as it turns out; see below) that initial
syllables are not subject to phonetic strengthening, then any licensing asymmetries
observed in initial syllables must owe their existence to something else, such as
psycholinguistic prominence. This prominence, with the restrictions on licensing asymmetries it carries, is what Smith uses to account for the typological regularities observed in the licensing potential of this position.

By contrast, stressed syllables, which are phonetically-strong, are seen by Smith both to behave as strong licensors of vowel contrasts (as in VR systems), and often to undergo Positional Augmentation of the vowel, resulting in the neutralization of contrasts. The onsets of stressed syllables are also seen to undergo Positional Augmentation of the type found in initial syllables (e.g. mandatory low sonority onset), but are said not to function as strong licensors of contrasts for onset consonants in the way initial syllables are. This is said to be comprehensible under the assumption that stressed syllables are not psycholinguistically prominent in the same way the initial syllables are, such that there is no processing-related mandate to conserve contrasts in those syllables. On the contrary, the most important factor is that the stressed syllable be perceptually prominent, and often in the service of achieving this additional prominence, both onset and vowel contrasts are lost. In sum, Smith’s account of initial syllable PN processes relies on psycholinguistic properties resulting from the fact that the syllable is initial to derive the typological patterning of affects.
4.1.3. Agenda for this chapter

As noted above, one of the main contributions of this chapter is a reexamination of the typology of vocalic licensing asymmetries in initial position, a typology which as it turns out has been characterized only incompletely in all previous work. On the basis of this reexamination, I will argue that all patterns of phonological licensing asymmetries attested in initial position are entirely predictable on the basis of the documented phonetic characteristics of those syllables, and, furthermore, that the predictive fit achieved on this basis is a closer one than that of a model assuming this typology to flow from the psycholinguistic status of the initial syllable. This argument is very specifically devoted to the question of how best to account for the typological regularities observed in patterns of positional licensing asymmetries, and here again I argue for the phonologization model over accounts based on the putative structure of Universal Grammar. In arguing that psycholinguistic status is not a directly operative factor in determining the typological regularities involving initial syllables, I am not in any way disputing that psycholinguistic status itself, which has been clearly demonstrated irrespective of any hypothesized role in constraining patterns of positional neutralization. I am also not disputing the potential need for separate formal mechanisms in the synchronic grammar to deal with Positional Strength and Positional Augmentation, as proposed by Smith (2002). As discussed in Chapter 6, I am also in agreement with Smith that the patterns of asymmetrical
phonological licensing in question are best handled in synchrony through the
manipulation of abstract categorical symbols in the phonology. These patterns should be
held distinct, however, in their implementation from the phonetic patterns out of which,
through phonologization, they arise. It is the latter which will concern me here.

4.2. Licensing asymmetries in initial position

This section provides an overview of the typological generalizations emerging
from the database of several hundred languages I have constructed in connection with
this project. In my survey, the licensing asymmetries involving the onsets of initial
syllables described by Beckman and Smith are clearly attested. Other regularities,
however, not identified in previous studies, emerge as well. For cases in which initial
syllable onsets license more contrasts than the onsets of other syllables, Beckman cites as
examples, among others attested, Doyayo (implosives, labiovelars), Shilluk (secondary
articulations), and !Xóó (clicks). Among languages with Positional Augmentation of
initial-syllable onsets, Smith cites Arapaho and Guhang Ifugao as requiring initial
syllables to have onsets, and Mongolian, Mbabaram and Campidanian Sardinian as
requiring that those onset consonants be of lower sonority. Again, that these contrast-
neutralizing constraints should be allowed in this psycholinguistically-strong position is
justified by the fact that the presence of a low-sonority onset aids in the segmentation of
the speech stream by marking the beginning of the word with a perceptually-robust low-
sonority to high-sonority CV-transition. Smith presents evidence that robust modulation of the signal in this way increases the perceptual salience of a syllable.

While it is true, however, that this "strengthening" of the onset of the word could assist in demarcating a word-boundary, and thus be permitted under Smith's Segmental Contrast Condition, it is not at all clear how the requirement that constraints neutralizing contrasts in initial syllables must assist in the segmentation of speech stream necessarily limits the Positional Augmentation constraints referencing that position to those affecting the presence or sonority of onsets alone. If it is the case, as Smith explicitly argues, that the perceptual robustness of initial material can assist in the segmentation of the speech stream, and if it is likewise true that a steep rise in sonority at the beginning of a syllable increases its perceptual prominence (Smith 2002: 78-80), then there is no reason that raising the sonority of the nucleus of the initial syllable would not be just as effective a measure in the service of boundary demarcation as would be lowering the sonority of its onset. The latter is quite well attested, of course, while the former is not known to occur\textsuperscript{116}. The psycholinguistic explanation of the typology here, however, provides no

\textsuperscript{116} There is the case of Luganda, in which absolute-initial vowels are restricted to /e, a, o/ out of a standard five-vowel inventory, and these not contrasting in quantity (they have been described as obligatorily long, though Hubbard's experimental results are somewhat equivocal on this matter), but this restriction does not concern initial syllables with onsets (Hubbard: 161). See below for further discussion.
reason to favor the one strategy over the other (in fact, if the information-bearing status of the contrasts in question should be a factor as to which neutralization would be more detrimental to lexical access, the consonantal effects might well be disfavored, rather than attested to the exclusion of anything else). Implementation of both might conceivably neutralize too many contrasts for the system to bear, though then we need a metric for evaluation of licit numbers of neutralizations in a given system, rather than merely guidelines as to which types are permitted and which are not. For that matter, since it is argued also that a bimoraic vowel increases the perceptual prominence of a stressed syllable, and also that in some systems (such as English) the prominence of the stressed syllable is used to facilitate the segmentation of speech, it is not clear why, in a language without duration-cued stress, all initial-syllables should not be required to be bimoraic. Such a neutralization would contribute enormously to the perceptual prominence of the initial syllable, and hence to the demarcation of the word boundary, while presumably not detracting nearly as much from word recognition as the neutralization of multiple consonantal contrasts would. And yet such a thing seems to be utterly unattested in the languages of the world. Smith's formulation of the Segmental Contrast Condition defines legitimate exceptions to bans on neutralizations in psycholinguistically-strong positions as those in which the satisfaction of the Positional Augmentation constraint "serves to demarcate the left edge of σ," (Smith 2002: 296), which perhaps implies a priori
limitation to the onset consonant, yet the sonority rise from C to V implicated in boundary-demarcation by Smith is surely as much a property of V as C. Additionally, even if the constraint affected only the vowel, by mandating that it be long, for example, surely the fact that the vowel is not always in absolute-initial position cannot mean it is incapable of assisting the listener in the process of segmentation of the speech stream. Indeed, it is a truism of functionalist speculation that systemic properties like fixed stress (initial or otherwise) and even word-bounded vowel harmony ultimately serve some demarcative purpose, and yet do not necessarily involve directly boundary-adjacent material. The purely psycholinguistic approach to licensing asymmetries in initial syllables is therefore by no means as restrictive as it would need to be to account for the typology of Positional Augmentation effects in that position.

Turning to vocalic contrasts, as noted, Positional Augmentation effects appear not to occur. As for Positional Strength, perhaps the most striking thing about initial position is, despite numerous claims in the literature, how very few such licensing asymmetries are in fact attested. Indeed, even of those known and accepted, the majority are either ambiguous in status from the outset, or ultimately attributable to phonetic factors rather than to the "initiality" of the syllable per se, or any psycholinguistic prominence bestowed on the initial syllable thereby.
Before turning to the attested cases of licensing asymmetries for vowels in initial syllables, it should be noted that, by contrast with an entirely processing-based account, the typological patterning of licensing asymmetries in initial position described above is entirely predictable from the well-attested phonetic characteristics of initial syllables alone. The idea that initial syllables are not the locus of phonetic prominence enhancement is just not accurate. In fact, a phenomenon known as “initial strengthening” in the phonetic literature is attested in a variety of languages, where it has been seen to affect only the onset consonants of domain-initial syllables (Byrd 2000, Cho and Jun (2000), Dilley, Shattuck-Hufnagel and Ostendorf 1996, Fougeron 1999, Fougeron and Keating 1996, Keating, Cho, Fougeron, and Hsu 1999, Oller 1973 inter alia). The following section introduces the details of initial-strengthening phenomena.

4.3. The Phonetics of Initial Position

Domain-initial strengthening refers to a set of processes that have been shown to enhance various aspects of the articulations of domain-initial consonants in a fairly wide variety of languages, including to date English, French, Taiwanese, and Korean. Experiments such as those cited just above have documented an increase in the magnitude of the supralaryngeal gestures associated with initial consonants, assessed according to the amount of linguopalatal contact involved in the constrictions (measured
using electropalatography). In addition to gestural magnitude, closure durations of the relevant initial consonants were also found to be greater than those of their word-internal counterparts\textsuperscript{17}. Laryngeal gestures have also shown significant enhancement in domain-initial position. Specifically, glottal opening gestures (and likewise VOT of aspirated stops) have been shown to increase in magnitude in English (Pierrehumbert and Talkin 1992), as does VOT in Korean (Keating, Cho, Fougeron and Hsu 1999). Moreover, Cho and Jun (2000) actually demonstrate for Korean increased glottal airflow and VOT for the lenis and aspirated series of stops, and decreased glottal airflow and VOT for the fortis stops, suggesting a general enhancement of articulations producing cues important to the perception of specific contrasts. A number of the initial strengthening studies have also recorded an effect similar to that discussed in Chapter 3 in connection with the implementation of final lengthening, to wit, that the degree of strengthening of domain-initial articulations is greater at the boundary of higher-level prosodic domains (e.g. U, IP, AP) and smaller at the left edge of lower-level constituents. Cho and Jun (2000) go on, intriguingly, to characterize certain patterns of initial strengthening as functioning specifically to enhance the syntagmatic contrast of the relevant segments. Thus, what they characterize as an increase in the "consonantality" (Cho and Jun 2000: 58) of the

\textsuperscript{17} While both magnitude and duration were found to increase in a number of the relevant experiments, correlation tests in Fougeron and Keating 1996 show that the direction or even existence of a causal relationship between these two parameters is questionable.

294
initial consonant serves to enhance the contrast between it and the following vowel. Additionally, building on work by Hsu and Jun (1998) on Taiwanese, they hypothesize (and demonstrate for Korean stops as noted above) that other strengthening patterns serve to enhance paradigmatic contrasts (in the Korean case, those between lenis, fortis and aspirated stops). Cho and Jun (2000: 67) themselves suggest directly that these phonetic strengthening processes may serve to enhance the segmentation of the uninterrupted speech stream into smaller units, in addition to potentially facilitating lexical access. The parallels here should be abundantly clear between phonetic syntagmatic strengthening of the initial CV contrast and phonological Positional Augmentation of word-initial onsets on the one hand, and phonetic paradigmatic strengthening of initial consonants and phonological positional neutralization of consonantal contrasts in non-initial positions on the other hand.

Domain-initial strengthening of vowels of initial syllables with onsets, however, has not been documented in any of the studies cited above. Cho and Jun (2000) note that domain-initial strengthening seems to target exclusively that first segment of the word of domain-initial position (in contrast with final lengthening, which targets the entire final rhyme, and often more). Byrd (2000) discusses this pattern together with final lengthening in the context of an experiment assessing the lengthening of gestures on both sides of a phrase boundary. As noted in Chapter 3, she attributes the lengthening of both
domain final and initial material to a single articulatory phenomenon, a local decrease in
gestural stiffness resulting in increased duration for the segmental material involved. This
decrease in gestural stiffness she attributes to the activation of what she calls a π-
(prosodic-) gesture, a non-constriction-based variable in the task dynamic model of
Articulatory Phonology that can alter the stiffness parameter for any gestures taking place
during the scope of the π-gesture. The π-gesture in question here straddles the phrase
boundary, affecting the segmental material on either side, and the material closest to the
boundary more dramatically than that further in on either side. This accounts nicely for
the gradual decrease in the effect of final lengthening as distance from the boundary
increases. It also predicts that initial consonants should be more strongly affected than the
vowels following them, though perhaps not the total lack of effect seen on this vowels in
the languages tested. Byrd speculates that the articulatory strengthening documented on
boundary-adjacent segments might also be a function of the presence of this π-gesture,
though the specifics of this interaction are still not entirely clear.

Now if the pattern of phonetic initial-strengthening described here were ultimately
to form the basis of the positional licensing asymmetries found in initial syllables, we
would expect exactly the effects described by Beckman and Smith for initial onset
consonants, while expecting the absence of either Positional Augmentation or Positional
Strength effects involving the vowels of initial syllables. Smith documents the non-
existence of the former, and I will turn momentarily to the documentation, despite all expectations (including my own), of the latter. I should note here, however, an important point about the potential relationship between psycholinguistic prominence and phonetic strengthening. Inasmuch as syntagmatic contrast enhancement involving initial onset consonants can be said to facilitate segmentation of the speech stream, and insofar as paradigmatic contrast enhancement affecting that same can be imagined to enhance perceptibility of the contrasts in initial position so crucial to early-stage word recognition, it is perfectly logical to hypothesize (though I will refrain from doing so) that these phonetic effects themselves are derived from precisely the concerns invoked by Smith in her characterization of the psycholinguistic status of the initial syllable. This may indeed be so, but it in no way changes the substance of my argument here. For even if psycholinguistic concerns of one sort or another are ultimately the raison d'être of the set of phonetic patterns associated with domain-initial segmental material, it is still the precise characteristics of those phonetic patterns which determine the shape and distribution of the phonological licensing asymmetries to which they may ultimately give rise through phonologization. Psycholinguistic status, in other words, may ultimately explain why anything gets phonologized in initial position, but it is the phonetic patterns alone which determine what will be phonologized where. Smith, by contrast, places the psycholinguistic concerns discussed here directly into the phonology, as filters on the
types of position-specific constraints which can be formulated by the grammar. The typological patterns we wish to account for here flow directly from the categorical phonological grammar, and as Smith argues, since the constraints in question have the entire initial syllable as their domain, should in principle hold for all elements within that domain equally. This is not the case, which the psycholinguistically-based phonology model fails to predict. The phonetically-based phonologization model, on the other hand, predicts precisely the attested asymmetry, whether the phonetic patterns in question ultimately have psycholinguistic motivations (as is potentially the case with initial strengthening) or are general properties of motor systems linguistic and non-linguistic (as with final lengthening).

Smith discusses domain-initial strengthening in the context of determining the precise domain to be referred to by the phonological constraints producing positional licensing asymmetries. She decides ultimately on the morphological word rather than the prosodic word for a number of reasons. The first and most important of these is that initial-strengthening is phonetic, and hence "cannot be directly related to true initial-syllable augmentation effects, which are phonological" (Smith 2002: 312). The gradience and interspeaker variability in implementation of initial strengthening distinguish it from, for example, categorical prohibitions on onsets above a given sonority level. Likewise, the fact that initial strengthening is stronger in the higher prosodic domains is troubling,
since the majority of phonological initial-syllable effects take place at the lexical level.

Smith notes that initial strengthening may ultimately "bolster" the pressure to demarcate
the left edge of the word, but that it is not by itself enough to make initial syllables
phonologically strong positions.

From the point of view of the synchronic modeling of individual phonological
systems, this is entirely correct. Synchronically, initial strengthening is hardly the same
thing as phonological strength. Indeed, given the apparent generality of attestation of the
former, it is likely that languages with specific phonological constraints on word-initial
onset sonority also implement phonetic initial strengthening of prosodic-word or phrase
initial material, such that, for example, a phrase-initial voiceless obstruent satisfying the
phonological constraints will have greater linguopalatal contact, closure duration, and
VOT than its phonological equivalent phrase-internally. Obviously, the grammar of a
speaker of this language must account for the phonological distribution and the phonetic
pattern separately, as distinct regularities, and within that grammar there would be little
reason to imagine that the implementation of the former be necessarily contingent upon
the presence or particularities of the latter.

But this only addresses part of the question. Indeed, where the focus of our
inquiry is on how to account for typological regularities in attested patterns of licensing
asymmetries in initial position, then, as in the cases of stressed and final syllables
surveyed above, it is precisely this phonetic information which should interest us. The phonologization approach to typology makes the relationship between the phonetic pattern of initial strengthening and initial-syllable phonological licensing effects, however hazy in synchrony, both central and clear to our accounting. As in other cases, the gradient phonetic process, despite its variability and stronger implementation in higher prosodic domains nonetheless creates sufficient individual strengthened tokens of the forms in question to bring about the ultimate phonologization of that strength, be it as resistance to neutralizations operating internally, or as neutralization along the lines of Positional Augmentation. In this sense the initial syllable is no different from the stressed syllable. Both positions undergo phonetic strengthening. What is different, however, is the nature of the phonetic strengthening involved. Phonetic initial strengthening targets primarily the first segment of the word or phrase, thus making it perfectly clear why both Positional Augmentation and Positional Strength processes should single out onset consonants.

It also explains the increased duration of the absolute word-initial vowels which leads to phonological effects never observed on initial-syllable vowels following onset consonants. Some evidence for such durational increase was presented in Chapter 2; recall that Balasubramanian 1981 demonstrates this longer duration experimentally for Tamil, the same duration was implicated in the resistance of absolute initial /a/ to
reduction to schwa in Russian and in the resistance of absolute front vowels to centralization in Nawuri (Casali 1995). Increased syllable duration may also potentially be implicated in the reduction of the five vowel inventory of Luganda to three vowels ([e, a, o]) morpheme-initially (Hubbard 1994). There is also evidence of this process from the literature on initial strengthening. Absolute word-initial vowels are realized somewhat longer than word-internal vowels in French and English (Fougeron 1999, Turk and Shattuck-Hufnagel 2000).

Furthermore, contrary to the expectations engendered by the synchronic psycholinguistic approach, there are indeed instances of the phonologization of initial syllable effects only in higher prosodic domains: In some dialects of Runyambo, for example, word-initial /i, u/ are lowered to [e, o] after pause (Larry Hyman, p.c.). The origin of this pattern in the phonetic durational increase and sonority enhancement characteristic of initial strengthening should be abundantly clear. It is certainly true that phonologized versions of initial strengthening effects tend to apply at the lexical, rather than the phrasal level, making patterns such as that of Runyambo comparatively rare. But is this not also true of other phonologization patterns originating in phrase-level phonetics, such word-final obstruent devoicing? It is generally accepted that final devoicing has phonetic motivation at the phrase level, while at the word level this motivation is at least weak if not absent (Gordon 1998, Hock 1999). Yet, phonological (neutralizing) final devoicing processes are generally recognized at the level of the word, and not the phrase. None of this is problematic from the point of view of phonologization.
The nature and attestation of initial-segment effects (C and V) are attributable to the phonologization of phonetic initial strengthening patterns. The phonologization approach predicts precisely the effects attested, while making the non-existence of unattested patterns (such as postconsonantal initial-syllable vowel augmentation) comprehensible as well. This result was obtained in chapter 2 and 3 as well, and it is a central thesis of this study that regularities in phonological typology are in general best accounted for from the point of view of phonologization. It may be that these phonetic effects, sometimes or in part, serve a function related to the psycholinguistic status of initial material, but it is the specific characteristics of the phonetic processes thus engendered, and not the existence of the psycholinguistic status per se, which determine the shape and attestation of the phonologized patterns. An approach relying solely on restrictions embodied in the synchronic grammar fails to make the connections between the phonetic patterns and the phonological typology, and so provides both less comprehensive and less restrictive predictions as to the nature of that typology. Most current work in generative phonology (and in particular with respect to the centrality of the factorial typology to work in Optimality Theory) makes the entirely aprioristic assumption that in providing a complete account of the representation of a specific phonological process in the grammatical competence of a speaker of a given language, we will also necessarily achieve a comprehensive account of the range of typological regularities observed crosslinguistically in the patterning and attestation of all related phonological processes. This postulate, I am arguing, is both fallacious and counterproductive, as it places logically out of bounds precisely the information necessary for a full accounting of the crosslinguistic facts. By contrast, the empirical
coverage of the phonologization approach (often based on information either irrelevant or unavailable to the individual speaker in synchrony) in accounting for phonological typology underscores this point. As evidenced throughout this study, there is simply no reason to assume that the same body of information necessarily explains both why set X and only set X variants of a phonological process exists, and also how a given speaker of a language acquires and implements a single member of that set X. Put more directly, a speaker certainly needs to know that his or her language contains a given pattern, but the assumption that he or she must also know why his or her language contains that pattern (in functionalist terms) is not at all obviously supported.

4.4. Initial-syllable strength effects

The foregoing account of the development of phonological initial strength effects from phonetic strengthening of initial material makes a prediction which flies in the face of apparent typological fact: insofar as significant phonetic initial strengthening has been detected only on word- or phrase-initial segments, and not on segments deeper inside word-initial syllables, we would expect not to find phonological licensing asymmetries involving those segments. This is indeed the case as concerns Positional Augmentation effects, and yet the literature contains numerous reports of vocalic Positional Strength patterns in which the initial syllable is the strongest licenser.

In this section I will argue that in none of the cases reported in the literature is it actually the "initiality" of the syllable per se which is in fact responsible for the licensing
asymmetries. Rather, in each case either phonetic factors (various types of additional duration) or morphological factors (the initial syllable is also the root) can account for the phonologization of the patterns in question. Other factors, such as processing concerns, need not be directly invoked to reach an understanding of attested typological regularities. Aside from this, a third group of patterns, to be discussed first, has simply been misanalyzed in earlier accounts of initial syllable typology. Section 4.4.1 analyzes this group of putative initial-syllable PN effects which are better attributed to initial stress. Section 4.4.2 introduces the possibility of durational asymmetries arising between initial and non-initial vowels from sources other than stress. This discussion leads into 4.5, which presents experimental evidence of initial-syllable vowel lengthening from initial strengthening in Turkish, and details a hypothesis as to how this small durational asymmetry could ultimately lead to the phonologization of vowel harmony.

4.4.1. Initial syllables in languages with fixed initial stress

Beckman 1998 notes that initial-syllable phonological strength effects are robustly attested in the harmony systems of Turkic, Tungusic, Mongolian, Finno-Ugric, and Bantu (Beckman 1998: 54), to which she adds also the surface-inventory reducing patterns of Tamil and Dhangar-Kurux. I will assume for the sake of argument that the first four of these families showing initial-syllable strength effects do not derive from a
common ancestor in which the strength effects were already present. Others, such as Poppe (1960), would disagree, reconstructing some form of harmony in the proto-language from which all this families are claimed to be derived. I will also assume that their geographic contiguity is irrelevant to the explanation of the phonological patterns. Steriade 1994 also lists asymmetries in initial syllable/non-initial syllable licensing in a variety of Uralic and Altaic languages as well as Tiv and Gokana. The licensing asymmetries in all of these are clear enough in the descriptions. The problem, however, in all these cases, excluding of course the Niger-Congo examples, is that in each of these cases at the time of development of the relevant licensing asymmetries the languages or language families in question had stress fixed on the initial syllable.

Proto-Finno-Ugric is generally thought to have had both initial stress and a form of palatal harmony, and indeed the overwhelming majority of modern Finno-Ugric languages have either initial stress or some obvious derivative synchronically as well (Abondolo 1998, Sammallahti 1988). To the extent that we wish to reconstruct Proto-Altaic at all, this is generally done with fixed initial stress. Poppe reconstructs an "expiratory accent" on Altaic word-initial syllables, generalized from the reconstruction of the same in the Mongolian, Tungusic and Turkic parent languages (Poppe 1960:143-118). The initial stress in certain Uralic languages, Ob-Ugrian in particular, in which earlier harmony systems have ceased to operate, creates the appearance of the UVR systems of the otherwise unattested type in which front/back or round contrasts are neutralized in non-initial syllables. See chapter two for discussion.
147). Li (1996: 20-21) notes that Poppe's characterization of Tungusic stress as fixed initial is too categorical as concerns the modern languages, though the reconstruction seems solid enough.

Most Dravidian languages are also characterized by initial stress or some obvious derivative thereof (e.g. default initial with attraction to heavy syllables further right), and fixed initial stress is assumed for the proto-language as well (Zvelebil 1970: 40). This accent, like that assumed for Mongolian, Tungusic, Turkic, and Uralic, was probably not strongly duration-cued like that of Russian or English, but this does not mean that a durational asymmetry was absent altogether. The resistance of Dravidian initial syllables is especially apparent in the massive syncopations of non-initial vowels found throughout the language family, and most dramatically attested in Toda and Kota (Zvelebil 1990: 3)119. Dhangar Kurux, cited by Beckman as contrasting nasal and oral vowels in initial syllables only lost original vowels, while reducing most other non-initial short vowels to an ultra-short schwa, later restored to "syllabicity" as short copies of the initial (Gordon 1976). Bosch (1991) notes that Tamil, another language with a restricted inventory of vowel contrasts outside the initial syllable, also has initial stress (albeit not strongly cued by durational). Andronov argues that Classical Tamil must have had initial stress

119 Resulting in forms such as Kota anţṛēčěgvőlk, 'because of the fact that (someone) will cause (someone) to terrify (someone)'.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
(Andronov 1975: 6) (while arguing that in must have shifted rightward in Colloquial Tamil to account for central vowel deletions from etymological initial syllables). Balasubramanian (1980) reports native speaker intuition of stress on initial syllables of words in isolation and lengthening of various elements thereof when the relevant word “has to be said with special emphasis” (Balasubramanian 1980: 456).

A similar state of affairs may obtain in Gujarati, where the stress pattern is somewhat less clear. Cardona (1965) describes it as a close derivative of the accent system of Classical Sanskrit (stress a heavy penult, else stress the antepenult - Bubenik 1996), which is to say not fixed initial. There are, however, licensing asymmetries in which the initial syllable is strong. Specifically, in Gujarati there is a contrast between plain and breathy-voice vowels, but there is a strong tendency for this opposition to be realized in initial syllables only (Silverman 1997:115)\textsuperscript{120}. Importantly, Gujarati breathy vowels are distinguished from clear vowels by, among either things, characteristic longer duration (Fischer-Jørgensen 1967: 94). Additionally, while Gujarati contrasts tense and lax mid vowels /e, o/ and /e, a/, this contrast is only available in the initial syllable (Pandit 1955). The origin of this licensing asymmetry in durational differences between initial and non-initial syllables becomes apparent when we consider the source this

\textsuperscript{120} Derived from the loss of earlier voiced aspirates or following /h/ (Ohala 1991). Its development is probably best seen as a case of perceptual metathesis à la Blevins and Garrett 2002.
pattern. Initial syllable /e/ and /ɔ/ arise through the monophthongization of the earlier diphthongs /aj/ and /aw/. Prior to this the difference between [e, o] and [e, ɔ] in initial syllables was purely allophonic, a function of segmental context and syllable structure. Crucially, monophthongization of /aj/ and /aw/ took place in non-initial syllables as well, only here the outputs of this sound change merged with preexisting /e/ and /o/ (Pandit 1961: 62-63). In other words, the diphthongs have two monophthongal realizations: a lower, presumably longer vowel in initial syllables, and a higher, presumably shorter vowel in non-initial syllables. The change in initial syllables causes the distribution of the low mid vowels to become unpredictable, and hence contrastive. Clearly though, the different outputs of the monophthongization must be the result of different phonetic characteristics, most likely duration, of the syllables in question. Other accounts of stress in Gujarati suggest a source for that additional duration in initial syllables: Pandit (1958), for example, reports that all syllables have “even stress”, except for those which are postjunctural. Initial syllables after juncture carry a stronger stress than others (Pandit 1958: 216). It is unclear whether this should be understood to imply a phrase-initial stress (as in Bengali), or the type of initial strengthening described below for Turkish. Clearly, however, additional duration in the initial syllable can be implicated in the phonologization of both licensing asymmetries. Dave (1967: 2) in fact claims that murmured vowels are only possible in stressed syllables, but does not elaborate in that
particular study. The precise phonetic characteristics of the domain-initial syllable in Gujarati and the status of these with respect to accentuation awaits experimental verification.

In any case, the problem with all the above licensing patterns is clear enough: If we are attempting to distinguish the set of phonological licensing asymmetries found in initial syllables from the set of those found in unstressed syllables, we can gain no insight by enumerating as exemplars of a particular initial-syllable strength pattern languages in which the word-initial syllable is also always the stressed syllable as well. This confound in the data makes it impossible to distinguish the effect of the initiality of the syllable from that of its stressedness. Fixed stress languages tend not to have strongly duration-cued stress, which explains why most of the languages in question have not developed UVR systems of the type discussed in chapter 2. It will be argued below, on the other hand, that the smaller durational asymmetries characteristic of languages with fixed stress can actually give rise to vowel harmony, precisely the form initial-syllable strength effects almost always take (Uralic, Turkic, Tungusic, Mongolian, Niger-Congo). That not all of the relevant languages have initial stress synchronically is irrelevant to the phonologization approach to the typology of positional neutralization, assuming initial stress in the parent language at the point at which the licensing asymmetry first arose. As demonstrated in numerous cases in chapters 2 and 3, the phonetic effects associated with
the syllables in question are relevant only at the very point at which phonologization takes place. That such effects were present then (e.g. that Proto-Uralic or Proto-Turkic had both initial stress and palatal harmony) is enough to account for the existence of the pattern. Once phonologized, of course, the licensing asymmetries cease to be dependent on the phonetics for their existence, and may continue to exist even in the event of radical changes to the phonetics of the language (e.g. shifts of stress away from the initial syllable. This is in fact the standard account of the development of restrictions on vowel licensing outside initial syllables in Latin). Such cases receive their “accounting” as part of the typology of initial-syllable effects from a stage of the language long since ceasing to exist, an accounting which the largely-irrelevant synchronic state serves only to obscure.

4.4.2. Other sources of duration in initial syllables

Remarkably, the exclusion of languages with fixed initial stress in synchrony or diachrony immediately removes from our typology the overwhelming majority of vocalic initial-strength effects remarked upon in the literature. Still, a few cases, such as Bantu, cannot be explained in this manner, as there is no evidence that anything like an initial stress accent ever existed in these languages. But is lexical stress the only possible source of additional duration in initial syllable vowels? Certainly the literature on initial
strengthening to date has found (to the extent that it has sought) additional duration on vowels only when these are absolute initial in the word, not preceded by an onset consonant. Otherwise, no across-the-board lengthening of initial syllable vowels has been identified experimentally. The nearest thing is found in the Korean experiments of Cho and Keating 1999, where two of four speakers show increased durations of vowels in initial #CV sequences, but only after consonants, whose VOT does not increase as a function of initial strengthening ([n] and tense t, which they transcribe as [t*]). Cho and Keating speculate on this basis that there is a compensatory relationship in the implementation of initial strengthening between duration of the initial-syllable vowel and the VOT of the preceding consonant. As in most of the initial-strengthening studies, unfortunately, here comparisons are only among initial-syllable vowels at the left edge of prosodic domains of differing levels. No comparison is made between initial-syllable vowels (phrase-initially or phrase-medially) and comparable vowels in word-internal syllables. Certainly for predictions concerning lexical-level licensing asymmetries this would be among the most important measures to be taken.

To assess the possibility that in certain languages initial strengthening does in fact affect initial-syllable vowels in a way that could ultimately lead to the phonologization of initial-syllable licensing asymmetries such as those found throughout Central Eurasia and sub-Saharan Africa, I constructed an experiment investigating the durations of initial
syllable vowels in Turkish and English. English was selected as a control for the experiment, as previous work (Fougeron and Keating 1996, Byrd 2000) had found no significant lengthening of initial syllable vowels in that language. Turkish was selected due to persistent reports in the literature and impressionistic observations in my own fieldwork that, despite fixed final stress in most of the native lexicon, something on the order of durational enhancement was nonetheless afoot in the Turkish initial syllable. There is, of course, a potential confound here, since as noted above, Proto-Turkic is generally reconstructed with fixed initial stress, so that any additional duration found there could in principle be a phonetic footprint (as it were) of the earlier prosodic system. I will not dispute the reconstruction of initial stress for Proto-Turkic, though it should be noted that of the putatively Altaic languages discussed by Poppe, by far the weakest case for an earlier initial stress can be made for Turkic. Johanson (1998:111) notes that an assumption of initial stress would help to explain the development of vowel harmony, early reduction and loss of final vowels, and alliteration patterns in Old Turkic. It is the purpose of the following to show that there is reason to believe that even without initial stress enough additional duration is contributed to initial syllable vowels in Modern Turkish by initial strengthening to give rise. Substantial initial strengthening of consonants could also have been responsible for giving sufficient prominence to these that alliteration might stand out as an attractive literary device. As for the reduction of
final vowels, is was noted in chapter 3 that stressed final high vowels in Modern Turkish are subject to devoicing, and that /a/ is raised toward schwa in Modern Turkish, again irrespective of the placement of stress. Furthermore, chapter three also cites the abundant cases in the diachronic literature of other language families (originally) with pitch-cued accent (e.g. Baltic and Slavic) in which vowels which were indisputably stressed at one point ultimately see stress retraction and deletion. For none of these problems, then, is an initial stress the sole (or even an adequate) explanation. The default initial stress in Chuvash in words containing only reduced vowels does suggest an earlier initial stress, which was later attracted to heavy syllables further to the right (though see Dobrovolsky 1999 and following this Gordon 2000 for phonetic data suggesting that the initial pitch peak in Chuvash words is better analyzed as an intonational rather than a stress prominence). Be this as it may, some scholars continue to maintain, along with Turkologists like Poppe, that Modern Anatolian Turkish continues to have both a pitch accent on the final syllable and a "dynamic stress" generally, though not always on the initial syllable (Csató and Johanson 1998: 207). It is my contention that what these scholars are in fact hearing is initial strengthening of onset consonants and vowels in a language which otherwise has a non-duration-cued fixed final stress. The following sections discuss my experimental results and the implications thereof for the development of palatal harmony systems of the Eurasian type.
4.5. Vowel harmony and initial syllables

The wide attestation of progressive vowel harmonies proceeding (at least historically) from the initial syllable in many languages without initial stress (such as Turkic or Bantu) demands an explanation. In such systems, only the initial syllable of the word realizes the language's full set of contrasts, while in non-initial syllables certain features are predictable from the specification of the vowel in the initial syllable. In languages with progressive palatal vowel harmony, such as Turkish, the frontness or backness of non-initial vowels is generally determined by the frontness or backness specification of the initial vowel. This is illustrated in (41) using several monosyllabic roots and suffixes.

(41) gözle'rim  dostla'rum  dʒanla'rum  dʒeple'rim  
    eye-pl-1sg  friend-pl-1sg  soul-pl-1sg  pocket-pl-1sg  
    'my eyes'   'my friends'   'my souls'   'my pockets'

Given the lack of initial stress in Modern Turkish, if we are to assume that this pattern is ultimately the result of the phonologization of some phonetic characteristics of Turkish initial syllables, we must seek those phonetic characteristics in some other aspect of Turkish prosody, such as the implementation of initial strengthening.
4.5.1. Initial syllables in English

The first experiment to be presented here was designed as a control to see that the results of the methods employed here matched those of Fougeron and Keating 1996 and Byrd 2000, who demonstrated a lack of initial strengthening for initial-syllable vowels in English. Keating and Fougeron 1996 demonstrated that in English vowel duration is strongly correlated with degree of jaw opening. In my experiments here I analyze only the durations of initial-syllable vs. second-syllable vowels. No articulatory measures were collected.

4.5.1.1. Methods

Stimuli were 24 more-or-less actual words of English. All stimuli contained an open syllable with an /æ/ nucleus under secondary stress. The vowel /æ/ was chosen for its long inherent duration, on the assumption that any systematic temporal variation would be more readily detectable in a longer stimulus than in a shorter one. In one set of stimuli, the syllable containing the target vowel was initial in the word. In the other set, the syllable in question was second in the word. The target vowels all receive secondary stress in their words. The reason for choosing specifically this level of prosodic prominence was as follows: Since unstressed vowels in English are heavily reduced and extremely short, they would make poor candidates for the detection of small-scale
durational variations. Choosing primary stresses, however, would in effect confound two distinct loci of potential phonetic enhancement. This would be particularly detrimental in comparing vowels in a number of different prosodic constituents, since accentual lengthening is known to behave much as initial strengthening of consonants in this regard. Adjacent segments were also controlled to avoid perturbations of vowel duration stemming from these. Following segments were in all cases voiceless obstruents, while voiceless stops were avoided as preceding segments because of their long positive VOT in some environments in English. Several tokens contained syllable-initial voiceless fricatives, which in retrospect may have been an error, since these too are known to impact negatively the durations of following vowels. The number of such tokens, however, was relatively small, and evenly distributed between initial-syllable and second syllable tokens. All target vowels occurred in open syllables, again with an eye to reducing non-position-dependent durational variation. (42) shows a pair of tokens illustrating the two types of stimuli.

(42) Syllable 1 vs. Syllable 2
màcerabilité anàphrodísiaç

Each token was situated in three different frame sentences selected to place the target word in initial position in a variety of prosodic domains à la Fougeron and Keating 1996.
The relevant domains were Utterance, Phonological Phrase, and Phonological Word. This is shown in (43).

(43) Prosodic Environments
a. Utterance-initial: U[Phr[X is an interesting topic.
b. Phonological Phrase-initial: U[Phr[I think]Phr[X is an interesting topic.
c. Word-initial: U[Phr[Y X compound] is an interesting topic.
   e.g. fish macerability, frog anaphrodisiacs, toe lacerability, plan irrationality

Participants were two native speakers of North American English raised in the northeastern United States, one male and one female. Speakers read the test sentences aloud from a randomized list. Sentences were uncovered one at a time by the author to insert a short pause after each sentence. Audio recordings of these sessions were digitized at 22.5 KHz., and vowel durations were measured from spectrograms and waveforms created using the Praat 3.9.5 speech analysis software (Copyright 1992-2000 by Paul Boersma and David Weenink).

4.5.1.2. Results

Mean vowel durations for both classes of stimuli are shown for each speaker in (44). T-tests revealed no significant differences between the vowels in initial and non-initial syllables. Additionally, no lengthening of the target vowels was observed in
higher-level prosodic constituents either. In other words, the results of this experiment are in agreement with those of previous investigations: English vowels in initial syllables are not subject to domain-initial strengthening. Nothing new emerges from this experiment.

(44) Mean vowel durations for English syllables 1 and 2

4.5.2. Initial syllables in Turkish

Turkish shows clear strengthening of domain-initial consonants. While a comprehensive study remains to be done, my preliminary observations over a large corpus of Turkish words recorded in isolation under laboratory conditions by a single speaker (created by the Turkish Electronic Living Lexicon project of Sharon Inkelas,
University of California, Berkeley) suggest that at least VOT of voiceless stops, prevoicing of voiced stops and duration and energy of voiceless fricatives are significantly enhanced domain-initially. Such strengthening is clearly visible in the spectrogram in (45).

(45)  [pʰaʰatja] chamomile

Here the strong aspiration of the initial labial stop is clear. It is in fact substantially stronger than that found in the onset of the stressed second syllable[121]. The longer duration of the second vowel is due to syllable structure, closed syllables hosting longer vowels than open syllables in Turkish (Kopkalli-Yavuz 2001, Barnes 2001b. I suspect that it is this initial-consonant strengthening, combined with the vocalic effect presented below, with is being heard and interpreted as an initial "dynamic stress" in Modern

[121] In Turkish as in English, both word-initial and stressed-syllable-initial voiceless stops receive strong aspiration.
Turkish by some scholars, as detailed above. The fact that it continues to occur in words like [pʰaʼpʰatja], where the adjacent penult is unequivocally stressed (or receives what Johanson would call the pitch accent and possibly “dynamic stress” as well122), makes it unlikely that an accent or secondary stress is somehow involved. Given the generality of initial strengthening in languages in which there can be no question of any fixed initial stress (e.g. English), there is little reason to assume an analysis of an altogether different nature to account for similar (though not identical) facts in Turkish.

This experiment investigates the phonetic durations of vowels in Turkish initial and non-initial syllables in a manner analogous to that used the English experiment described in the previous section. While both that experiment and two previous ones confirm the absence of strengthening of initial-syllable vowels in English, Turkish, I will show, presents a different picture.

4.5.2.1. Methods

Stimuli were 85 actual trisyllabic nouns or adjectives of Turkish. In one class of stimuli, an initial closed syllable contained the vowel /a/, while in another class, a closed second syllable contained that same vowel, chosen again for its inherent duration.

122 See Konrot (1981) for discussion of the possibility that while Turkish final stress is cued solely by $F_0$, the non-final stress of borrowings and toponyms may actually be cued by both pitch and amplitude (though not duration).
Turkish stress does not affect vowel durations (Konrot 1981), and unstressed vowels do not undergo reduction, making the problems in this connection in English irrelevant here. Instead, all stimuli had final stress, placing target vowels in either the first or second unstressed syllable. Surrounding consonantal environment was again controlled. Codas in target syllables were (in equal numbers for each class) voiceless stops and nasals. Onsets were voiced stops or sonorants. An initial version of the experiment included forms with voiceless fricative onsets, which can negatively influence following vowel duration. The majority of these were later replaced. Example stimuli are given in (46), with target vowels in boldface.

(46) Syllable 2 vs. Syllable 1
CV(C).CaC.CV(C) vs. CaC.CV(C).CV(C)
[kaj.mak.'tʃui] vs. [mak.buz.'dʒu]
cream-agt vs. receipt-agt
crea\n\nmaker/seller/enthusiast vs. receipt maker/seller/enthusiast

As in English, each stimulus was placed in three frame sentences, such that it would appear initially in three different prosodic constituents. The level intermediate here between Utterance and Phonological Word is in all likelihood an Intonational Phrase, though to my knowledge no comprehensive study exists of prosodic phrasing in Turkish. The three prosodic environments selected are shown in (47). In each environment the target word is the initial (non-head) element in a compound. The word-initial
environment makes that compound the head of an NP. These particular constructions are quite natural sounding, many street names in Turkey taking this form.

(47) Turkish Frame Sentences

Utterance-initial: U[Phr[X sokau tʃok gyzel bir jerdır
street very nice one place-pred
X street is a very nice place.

Phrase-initial: U[Phr[bana sorarsan] Phr[X sokau tʃok gyzel bir jerdır
I-dat ask-cond street very nice one place-pred
If you ask me, X Street is a very nice place.

Word-initial: U[Phr[istanbuldaki W[X]sokau tʃok gyzel bir jerdır
istanbul-loc-pcp street very nice one place-pred
The X street in Istanbul is a very nice place.

Four native speakers of Istanbul Turkish were recorded reading the stimulus sentences from a randomized list. Again, the sentences were revealed to the speakers one by one, such that a short pause was induced following each sentence. Recording sessions took place at UC Berkeley and the Bosphorus University in Istanbul. Recordings were digitized at 22.5 KHz. and vowel durations were measured from waveforms and spectrograms using the Praat 3.9.5 speech analysis software (Copyright 1992-2000 by Paul Boersma and David Weenink).
4.5.2.2. Results

The results of this experiment are shown in (48). It becomes clear immediately that for each speaker, mean durations of initial-syllable vowels are significantly longer than those of the vowels of second syllables. This conclusion was supported by the results of two-tailed t-tests in which analysis of the durational difference between initial and non-initial syllables for each environment and every speaker showed significance at \( p < .05 \).

(48)  Mean vowel durations for Turkish syllables 1 and 2
Mean vowel duration differences between syllables 1 and 2 were uniformly greater for syllables closed by nasals than for those closed by voiceless stops. This is presumably due to the fact that in Turkish vowels are longer in general before nasals than before voiceless stops, and that all things being equal longer entities tend to exhibit more durational variation than shorter entities. In this connection it is also worth noting that Speaker 2 has substantially smaller differences between vowels in initial and non-initial syllables than the other speakers. This is most likely explained by the fact that his overall vowel durations are also shorter as a result of the rapid tempo at which he read the stimuli presented to him.

The strengthening effect detected, however, was not seen to increase at the boundaries of higher-level prosodic constituents. While vowels in utterance-initial syllables were consistently shorter than others, no significant patterns emerge between phrase- and word-initial syllables. In fact, vowels in phrase-initial syllables often turned
out to be shorter than their word-initial counterparts, for some speakers even with statistical significance. The same was in fact true for one of the English speakers as well. This reversal of the expected pattern might be explained by the fact that in both experiments the target words in phrase-initial contexts were both in longer sentences than the other stimuli, and farther from the beginnings of those sentences as well. These facts may have conspired to shorten the overall durations of the stimulus words at that level of the Prosodic Hierarchy, and hence of the vowels being measured as well.

While it may also be that initial strengthening in Turkish is simply not sensitive to these distinctions, it is nonetheless worth mentioning that even in the earlier studies which did detect increases at higher-level boundaries in other languages, differences were not always found for all speakers, and speakers differed frequently in their choice of boundaries playing a role. It is conceivable that a larger study of Turkish would have uncovered such an effect. It is also possible that the phrase-boundary selected for comparison with the word-boundary was insufficiently high in the hierarchy to trigger the increased strengthening effect. Further work in this direction would do well to vary the sentential contexts used, in order to control for possible problems of this type. One last problem may also have been in the choice of real-word stimuli, which, although carefully controlled, may nonetheless have introduced sufficient variation into the results to make detection of this level of durational asymmetry difficult. The studies that detected the
hierarchical strengthening patterns all used single repeated stimuli and nonsense words representing each phonetic environment under investigation. Such a strategy could prove more effective in Turkish as well.

We have seen so far that Turkish, but not English, exhibits a pattern of domain-initial strengthening affecting the vowels of initial syllables. Minimally then, we can say that domain-initial strengthening is implemented to differing degrees in different circumstances on a language-specific basis. This was shown in the earlier experiments as well (i.e. Keating, Cho, Fougeron and Hsu 1999), in service of the point that insofar as domain-initial strengthening varies on a language-specific basis, it cannot be relegated to the level of "universal phonetic implementation", and must receive some representation in the grammars of the languages in question.

Concerning the language-specific nature of domain-initial strengthening, I believe it is not by chance that it is Turkish, and not English, which exhibits initial-syllable vowel lengthening. One of the primary cues for stress placement in English is additional vowel duration in the stressed syllable. It is therefore unsurprising that English would avoid simultaneous implementation of other positionally-determined vowel-lengthening patterns, insofar as doing so would have the potential to seriously confound accurate perception of the placement of stress. This is hardly a novel argument. Indeed, something similar is hypothesized in Keating, Cho, Fougeron and Hsu 1999 concerning the
deployment of boundary signals in English, as opposed to French and Korean, which have substantially different types of prosodic systems. They in turn cite Lehiste (1964) making a similar argument concerning vowel length and boundary signals. Continuing this line of reasoning, then, in Turkish stress is not even weakly correlated with vowel duration (Konrot 1981), leaving that phonetic resource available for use in signaling word-boundaries, as the experiment presented here shows. The implicational force of this analysis is, to be sure, only negative, predicting which languages should not show initial-syllable vowel lengthening. Only a more extensive experimental survey of languages with non-duration-dependent accentual systems will provide the information necessary to make any further predictions on this matter. Mollaev 1980 demonstrates a strong tendency for the vowels of syllable 1 to be longer in duration than comparable vowels in syllable 2 of disyllables in Turkmen, but the closeness of the relationship of the latter to Modern Turkish does not allow generalizations to be expanded on this point.

It might be objected at this point that English does in fact use vowel duration to signal things other than placement of stress. Final lengthening is a well-known example. However, English final lengthening is consistent and robust only in the higher

---

[^12]: I do not consider durational variations induced by local segmental or syllabic environment in this connection, as these lack the culminativity associated with stress placement. In such cases duration is not being deployed in the language as the primary cue for a uniquely-identifiable prosodic position (such as primary stress or edge). That additional duration supplied by local segmental environment can nonetheless dramatically impact the structure of a stress system is of course clear, as the attraction of otherwise-edge-based stress to internal heavy syllables in many languages clearly demonstrates.
prosodic domains. At the word-level, it is sporadic and weak (cf. Beckman and Edwards 1990), which would tend to minimize its effect on perception of lexical stress placement (The phonetic differences between final lengthening and stress are discussed at length in chapter 3). Turkish initial strengthening, unlike the final lengthening found in English, is realized consistently in both lower and higher prosodic domains, and would therefore be expected to be more disruptive to the perception of stress placement than English final lengthening. Final lengthening is also frequently realized alongside a severe amplitude and F0 drop, both of which would minimize the likelihood of additional duration on the final syllable being mistaken for lexical stress. The crosslinguistic propensity for stress to avoid final syllables is also discussed in chapter 3.

4.5.3. Consequences of domain-initial strengthening: The Emergence of Turkic Palatal Harmony

The experimental results given above demonstrate the application of domain-initial strengthening to the vowels of initial syllables in Turkish. This process represents a source of additional phonetic duration in initial syllables independent of the initial stresses discussed above for Uralic, Dravidian, and at least some of the languages grouped as Altaic. Since the fixed initial stresses in those languages are not known to be strongly duration-cued, any additional duration contributed by stress to those initial
syllables might well be on the order of that documented here in the initial unstressed syllables of Turkish. If strengthening of initial-syllable vowels ultimately turns out to be more widely attested in languages with non-duration-cued stress, this process may represent a second source for the initial-syllable phonological strength effects attested in the literature. Still, it is not at all clear how a small-scale durational asymmetry of the type documented above could develop into word-bounded phonological vowel harmony of the Turkic variety. The following presents a possible solution to this problem.

4.5.3.1. Does vowel harmony come from vowel reduction?

It is often observed that vowel harmony and vowel reduction share many common features, most notably that they both involve the positional neutralization of vowel quality contrasts. Indeed, for Beckman (1998), the difference between the two is essentially just whether the language allows multiple linking of vocalic features. It seems quite plausible that the one type of system might be linked in some way with the other developmentally, and specifically, that the chain of assimilations imagined to give rise to systems of vowel harmony would be phonetically quite a bit more plausible were it to take place across a string of vowels whose quality had already been neutralized by some other process (as was suggested above for the shift from vowel harmony to vowel inventory reduction in Ob-Ugrian dialects). If weak positions (e.g. non-initial syllables)
already licensed the appearance of fewer contrasts, and the vowels which did surface there were of significantly diminished duration, we could imagine they would be more susceptible to coarticulatory effects from neighboring strong vowels. Rhodes (1999), for example, observes English reduced vowels undergoing low-level gradient assimilation in roundness to neighboring back rounded vowels. Certain East Slavic dialects with robust systems of vowel reduction also display patterns of dissimilation or assimilation involving the stressed vowel and the first pretonic vowel (see Crosswhite 2001 for an OT-based analysis some of these systems). Languages with some sort of harmony system already in place are also known to add new assimilations to pre-existing ones (e.g. the gradual extension of rounding harmony in the Turkic languages).

All these cases, while not in fact instances of the development of full-blown word-domain vowel harmony systems from earlier non-harmonizing reduction systems, nonetheless suggest there is something to the reduction-then-assimilation hypothesis. There are a number of reasons, however, why in the cases at hand, such an origin for harmony is implausible. First, Proto-Turkic is generally reconstructed with some system of palatal harmony already in place. Assuming that it is correct to reconstruct initial stress (rather than just phonetic initial strengthening) for Proto-Turkic, it is necessary to bear in mind that fixed-stress systems are not generally strongly duration-cued and non-duration-cued stress-systems rarely generate robust patterns of phonological vowel reduction. A
glance across the prosodic systems of the Slavic languages illustrates this point well on a smaller scale: Czech, Polish, Serbo-Croatian, Standard Macedonian with no duration-cued lexical stress and no phonological vowel reduction, alongside Russian, Belorussian, Bulgarian and Macedonian dialects with duration-cued phonemic stress and phonological vowel reduction, to name only the best known cases. For the reduction-first generalization to be correct for Turkic then, Pre-Proto-Turkic must have been a counterexample to this generalization. Specifically, it would have had to develop a vowel reduction system sufficiently sweeping in nature to give rise to the attested harmony system, and to do so in a system with fixed, non-duration-cued stress. Alternatively, one might imagine that Turkic fixed stress was at one point duration-cued and later changed. If the initial-stress reconstruction is wrong, on the other hand, and Proto-Turkic had only some degree of phonetic initial strengthening (like Turkish today), then we are still dealing with a relatively small durational difference between the vowels of initial and non-initial syllables. We would not expect such a system to yield the necessary contrast neutralizations either. There were in fact reductions affecting vowels in non-initial syllables in Turkic. In Proto-Turkic initial syllables, a complete inventory of long and short vowels were contrasted (Róna-Tas 1998: 69-70, Johanson 1998: 90-91). In non-initial syllables of the root short vowels contrast with “reduced” vowels, these latter derived from earlier short vowels and ultimately subject to syncopation or reduction to
schwa depending on the syllabic environment. There is no reason, however, to think that unreduced vowels in non-initial syllables were especially durationally impoverished, but they were subject to palatal harmony nonetheless.

More damning though for the reduction-first hypothesis is the following: Recall that Turkic vowel harmony involves the neutralization of frontness/backness distinctions, which is to say, quality contrasts along the F2 dimension. In order to derive this state of affairs from an earlier system of vowel reduction, then, we must imagine that Pre-Proto-Turkic had a reduction system allowing a full range of contrasts in initial position, but neutralizing all F2 contrasts in non-initial syllables while retaining three contrastive heights (assuming the non-initial syllable inventory of Róna-Tas 1988: 70). As elaborated in chapter 2, however, among the unstressed vowel reduction systems of the world, such a system seems completely unattested, and with good reason, given the phonetic motivations behind the development of UVR systems. Were the immediate predecessor to the Turkic palatal harmony system one in which the soon-to-be-alternating qualities did not contrast due to reduction, that reduction system would have had to be virtually unique.

Furthermore, while some featural agreement of stressed and unstressed vowels can arise in UVR systems, it tends always to be harmony of the same type: either an unstressed vowel retains its earlier quality when the stressed vowel is of the same quality,
or the unstressed vowel takes on the quality of the stressed vowel entirely, showing complete agreement\textsuperscript{124}. Thus, in Yakan, pretonic /a/ does not reduce to [e] if the stressed vowel (along with any intervening pretonics) is also [a] (Behrens 1975). In Timugon Murut pretonic non-high vowels are [o] if the stressed vowel is [o], and otherwise they are [a] (Prentice 1979). In Russian dialects exhibiting what is known as “assimilative” vowel reduction, there are systems in which, for example, a first pretonic /o/ will fail to reduce if the tonic vowel is also /o/, but otherwise will merge with /a/, or in other dialects, first pretonic /o/ remains [o] unless the stressed vowel is [a], in which case it merges with /a/ (Chekmonas 1987: 342). It is important to note that in Murut and Russian, where unstressed /o/ retains its rounding due to what we might call support for this from the stressed syllable, it is only under conditions of identity with the stressed vowel that this occurs. The rounding of stressed /u/, for example, is insufficient to allow the harmonization that prevents the loss of rounding in the stressed vowel\textsuperscript{125}. This support provided by the stressed vowel to unstressed vowels of certain qualities could be understood in some cases as a strategy for coping with durational pressures on the

\textsuperscript{124} This are also, of course, stress-dependent harmony systems, such as those treated by Majors (1998)

\textsuperscript{125} It is not clear, however, whether this is due to a demand for complete identity of the vowels involved, or only a demand that they be identical in height. The only other rounded vowel in both systems is /u/, and the facts here do correspond to the generalizations about rounding harmony made by Kaun (1993) (specifically even that mid vowels are more likely triggers of rounding harmony than high vowels. Unfortunately in these systems the high vowels do not reduce, so it is unclear what height restrictions might turn out to affect potential targets).
realization of certain gestures. Thus, if an unstressed vowel is too short for its target articulation to be reached, one solution is to undershoot or even alter the phonological target. Another solution, possible in cases of identity of the relevant features would be to realize a single token of the problem gesture over two syllables, thus providing the necessary time to reach target. This is essentially Kaun’s (1995) notion of Uniformity, a phonetic multiple-linking which Boyce (1990) shows to be implemented in Turkish sequences of harmonizing round vowels (but not in similar sequences in English).

Supporting this interpretation is an intriguing note on Russian vowel reduction from Avanesov (1968). In this work Avanesov is concerned with the norms of Contemporary Standard Russian. In addition to describing these norms, along the way he provides advice for teachers of non-standard dialect speakers on how best to get their students to conform to that description in, for example, the system of UVR they employ in their speech. In this connection, on the topic of causing speakers of dialects that realize [o] faithfully in unstressed syllables, to begin merging /o/ and /a/ when unstressed as described in chapter 2, Avanesov notes that speakers of these dialects usually pass through a number of intermediate stages of pronunciation before completely acquiring the norm: often the first instances of /o/ reducing merge with /a/ (as [a] or [e]) are those in pretonic syllables preceding a stressed [a]. Additionally, often speakers learn to reduce

---

126 The mostly northern, so-called ‘okajushchie’ or ‘o-saying’ dialects.
nearly all unstressed /o/’s to [a], but nonetheless retain thereafter unstressed [o] in pretonic syllables immediately preceding stressed /o/ (Avanesov 1968: 52). This discussion, entirely in the context of the linguistic development of the individual speaker of a completely non-/o/-reducing dialect nonetheless describes precisely the characteristics of the two “assimilative” UVR dialects described above, dialects which, not accidentally, are located in the borderlands between the faithfully ‘o-saying’ and the solidly ‘a-saying’. In general this is a phenomenon worthy of serious further phonetic and dialectological study.

So while there clearly are systems in which there is a connection between systems of UVR and some kinds of harmonization, what we apparently never find are systems in which one set of contrasts is effaced by UVR, upon which a categorical system of, for example, palatal harmony, springs up opportunistically on the weakened vowels of the unstressed syllables177. The development of palatal vowel harmony in Turkic will not be understood through an appeal to UVR. If it is not the case that Proto-Turkic palatal harmony arose from an earlier phonological vowel reduction system, to what then can its emergence be attributed? I will argue that it emerges from precisely the type of small-scale durational asymmetry between initial and non-initial syllables as this paper shows is still found in the Anatolian Turkish of today.
A widespread and intuitively plausible conception of the development of vowel harmony patterns maintains that harmony arises diachronically through the phonologization of vowel-to-vowel coarticulation. This view is defended in e.g. Ohala 1993 and 1994. Flemming (1997) derives vowel-to-vowel assimilation patterns from coarticulation synchronically as well. Majors in her 1999 dissertation shows that in many languages vowel-to-vowel coarticulation is most robust from stressed vowels to unstressed, a fact which she argues could lead to the development of harmony patterns with stressed vowels as their triggers. Unfortunately, this information alone does not provide us with an analysis of the emergence of harmony in Turkic.

Inkelas et al. (2001), building on work by Beddor and Yavuz 1995, demonstrate experimentally that in Modern Turkish, anticipatory vowel-to-vowel coarticulation is stronger than carryover. Furthermore, this is true regardless of the position of stress (again, usually but not always final in Turkish). These findings present a serious challenge for coarticulation-based theories of vowel harmony for the following reason: If vowel harmony is driven synchronically by vowel-to-vowel coarticulation, then stronger anticipatory coarticulation should yield right-to-left harmony patterns, rather than the left-to-right system attested in Turkish. In Turkish, coarticulation and harmony run in opposite directions. For diachrony, of course, it would be possible to conclude, with

\[\text{[336] Rounding harmony, though, famously parasitic in any case, appears to do just this, as in Timugon}\]
Beddor and Yavuz, that all this simply means that prosody was radically different in Proto-Turkic times. It is conceivable that at the time of the emergence of palatal harmony in Proto-Turkic, carryover coarticulation was stronger than anticipatory, and that at some point this situation reversed itself. If Proto-Turkic did in fact have fixed initial stress as some scholars hypothesize, we could imagine that, unlike in Modern Turkish, Proto-Turkic coarticulation flowed most strongly from the stressed syllable à la Majors 1999, yielding the desired direction of assimilation. This hypothesis, however, has a number of serious flaws: If Proto-Turkic did in fact have fixed initial stress, then this stress pattern obviously changed somewhere along the way to Modern Turkish, such that the language developed a fixed final stress. Now, if early fixed initial stress was in fact duration- and/or amplitude-cued, then the assumption of Proto-Turkic carryover coarticulation would acquire a certain phonetic naturalness (again, as per Majors 1999). The idea of a subsequent shift to stronger anticipatory coarticulation, however, in the face of the uninterrupted phonetic prominence of the initial syllable, now becomes difficult to countenance. Certainly the innovation of the FO-cued accent of later Turkic is a poor

---

128 The alternative potential source for the modern initial syllable vowel-lengthening mentioned in note 4 above. The relationship between amplitude and stress in Modern Turkish is not straightforward, though there may be a correlation to some extent for non-final stresses (Konrot 1981). Numerous confounding factors in Konrot’s experiment make this difficult to ascertain. As noted before, duration is not correlated with stress at all.

129 Even assuming that only later did stress and coarticulation become dissociated, as they are today in Turkish.
candidate for the driving force behind that change. A shift in the direction of vowel-to-vowel coarticulation in these circumstances seems capricious and unmotivated. On the other hand, if the earlier fixed initial stress had prosodic characteristics similar to those of the stress in Modern Turkish, then even the motivation for assuming an earlier carryover coarticulation in the first place becomes obscure.

The preferable option, of course, would be to produce an analysis that could save the coarticulation theory without recourse to the historical assumptions discussed above. To this end I propose the following: Anticipatory coarticulation, all things being equal, may well be stronger than carryover in Modern Turkish, and lacking any reason to assume otherwise, I take this to have been the case in Proto-Turkic as well. As demonstrated above, however, initial-syllable strengthening in Modern Turkish produces, independent of the position of stress, a durational asymmetry between the vowels of initial and non-initial syllables. Assuming domain-initial strengthening to have been active in the past as it is today, the same would be true of Turkic at the time of the phonologization of vowel harmony. This durational asymmetry seems not to affect the dominant direction of coarticulation in Modern Turkish (Inkelas et al. 2001). It could, however, produce significant changes in the perception thereof.
Turkish vowel-to-vowel coarticulation patterns received absolute measurements in Inkelas et al. These were arrived at through comparison of mean formant values at vowel onsets or offsets in a coarticulated context (adjacent vowel different) with the corresponding values found in a baseline context (adjacent vowels identical). In a relative sense, however, an absolute coarticulatory effect of a given magnitude would nonetheless occupy a smaller portion of the total duration of a longer vowel than it would of a shorter vowel. This coarticulatory effect, however strong in the absolute sense, could then prove perceptually less salient on the longer initial-syllable vowel than on a shorter one. In Turkish this would mean that the overall stronger effect of anticipatory coarticulation might nonetheless fail to be perceptually robust on the lengthened vowels of the initial syllable. By the same token, carryover coarticulation, however weak overall, would receive additional salience perceptually on the shorter vowels of non-initial syllables. This durational skewing effect allows us to understand why, stronger direction of coarticulation aside, Vowel 1 of a Turkic word might still be less likely to assimilate to Vowel 2 than vice-versa.

All the foregoing, however, buys us no more than a single sound change: Vowel 2 assimilates to Vowel 1 in frontness/backness in Pre-Proto-Turkic. But this alone cannot be the full story. I am also less than sanguine about the plausibility of an analysis in which word-domain harmony is brought about gradually by the methodical creep of
palatality across from left margin to right in the word. Rather, the sound change described here must account for only the first step in the rise of Turkic vowel harmony. The remainder of the process would then have to be analogical in nature.

This idea receives support from the fact that the overwhelming majority of roots reconstructed for Proto-Turkic are either one or two syllables in length. Trisyllabic roots are shadily attested at best (Johanson 1998). Assuming also the possibility of the addition of suffixes to the root in questions, we must bear in mind crucially the following word-forms in any discussion of the emergence of vowel harmony:

(49) Some Crucial Proto-Turkic Word Shapes
a. \[[CV.CV]_{root}\]
b. \[[CV]_{root}[CV]_{suffix}\]
c. \[[CV.CV]_{root}[CV]_{suffix}\]

Assuming now that Vowels 1 and 2 in (49a-c) disagree in frontness/backness, the application of a sound change assimilating Vowel 2 to Vowel 1 in this respect yields the results schematized in (50):

(50) Assimilation of V2 to V1
a. \[[CV_aCV_{-a}]_{root}\] > \[[CV_aCV_a]_{root}\]

---

130 Calculations done by Kemal Oflazer in fact suggest that the mean number of syllables per word in running text in Modern Turkish may not be any greater than this (Sharon Inkelas. p.c.).

131 These representations are highly schematic and not meant to imply consideration of open syllables only.
The sound change in (a) produces a disyllabic root with palatal harmony, meaning essentially that the overwhelming majority of roots in the language are now harmonic.

The output of the change in (b) is a disyllable in which the suffix takes on the palatality specification of the root, creating another word without disharmonic vowels. Example (d) merely shows that if a suffix added to a disharmonic disyllabic root has the same [back] specification as the first syllable of the root, the output of the sound change will be a completely harmonizing trisyllabic word. The only problem among these examples, in fact, is form (c). Here the output of the sound change is a harmonized disyllabic root with a non-harmonizing suffix. The crucial analogical step occurs here.

The addition of the same suffix to [+back] and [-back] monosyllabic roots means that the sound change shown in (50) has the effect of creating an alternation whereby the choice of suffix vowel is dependent on the identity of the root vowel. The fact that no alternation occurs when the root is disyllabic creates irregularity in the choice of suffix vocalism: Sometimes the suffix vowel harmonizes, and sometimes it does not (the addition of two suffixes to a monosyllabic root potentially creates the same irregularity). This irregularity is removed from the system simply by the generalization of the
harmonizing allomorph to all forms, as illustrated in (51) where the plural suffix -lAr is taken as representative.

(51) Generalization of Suffix Alternation Pattern: Proportional Analogy

\[ CV_{[a]}C : CV_{[a]}CIA_{[a]}r :: CV_{[a]}CV_{[a]}C : X \]
\[ X = CV_{[a]}CV_{[a]}CIA_{[a]}r \]

At this point we have a system with all trisyllables, regardless of morphological structure, displaying palatal harmony. The extension of the alternation pattern to longer strings of suffixes is not difficult to conceive. It is worth noting that just as the generalization of the alternating allomorph is possible, we might equally well expect the reverse, viz., instances of generalization of the invariant allomorph to all forms. The fact that some suffixes are reconstructed for Proto-Turkic as non-harmonizing could be an indication that this in fact took place.

This account derives palatal vowel harmony in Turkic from a small-scale durational difference between initial and non-initial syllables, a difference which, while not enough to cause articulatory difficulties of the type implicated in chapter 2 in the development of UVR, is nonetheless sufficient to cause a skewing in the perception of vowel-to-vowel coarticulation, such that a non-initial vowel might be reinterpreted as agreeing with a preceding initial-syllable vowels in some featural dimension. Note also
however the importance of the role played in this account by root-structure and morphology. It is significant indeed in light of this that in fact the Altaic, Uralic and Bantoid languages with initial-syllable strength manifest as vowel harmony all have primarily mono- and disyllabic roots, suffixing morphology, and stress (if any) not strongly cued by duration. It is possible that a similar account could be viable for others of these cases as well.

4.5.4. Bantu

Turning now to a case of initial syllable strength for which past or present initial stress is clearly not a potential explanation, it is possible, given the lack of duration-cued stress in most Bantu languages (and certainly in the proto-language), that the same pattern of initial strengthening observed in Turkish is or was active in Bantu, and could be implicated in the development of Vowel Height Harmony in those languages. Unfortunately, I know of no experimental results providing evidence one way or the other. Researchers such as Hyman (1987) and Goldsmith (1982) have certainly proposed an abstract “accent” on the stem-initial syllables of some Bantu languages. Hyman 1987, on the basis of work by Paulian (1974) proposes an analysis of Kukuya placing an accent on the initial mora of the stem. The accent in question is not in any way meant to indicate

132 Any prefixes being less closely integrated into the stem phonologically than the suffixes.
that Kukuya is a "stress language"; rather, accent here is a slightly more abstract property affecting the licensing potential of consonants and vowels and affecting tone as well. The licensing restrictions are perhaps suggestive of phonetic strengthening, but without explicit evidence thereof no conclusions can be drawn. One fact about Kukuya does suggest a phonetic strengthening of accented material similar to that typically found in stress languages: There is a postlexical variable rule (which is to say, most likely, a phonetic process) of vowel reduction operative in Kukuya affecting material linked to an immediately postaccentual unaccented mora (providing the following mora is also unaccented), such that /a/ in this position is realized as schwa (Hyman 1987: 331). Phonetic investigation into the nature of this accent could yield important results for an understanding of the development of licensing asymmetries in Bantu.

Otherwise, it is certainly the case that initial segments in many Bantu languages are subject to initial strengthening of some kind. I have already discussed domain-initial vowels in Luganda and Runyambo. Hubbard (1994) provides evidence from a number of Bantu languages showing significantly greater durations for stem-initial consonants than for comparable consonants stem-externally. Languages showing this effect included CiYao and Kikerewe among others (Hubbard 1994: 95, 102). Certain Bantu languages

---

133 'Stem' here in the Bantuist sense of root plus suffixes.
134 Hubbard notes that the effect is less consistent here. Specifically, while it is strong for voiceless obstruents, it is less so for voiced stops, and actually reversed for sonorants /m/ and /l/. For nasals at least
of the southern zones, such as Chichewa, have regular lengthening of vowels in penultimate syllables. In some languages, such as Makonde, this has even been analyzed as a penultimate stress with concomitant UVR in unstressed syllables (Liphola 2001). What is fascinating here are the attendant changes in the results of Hubbard’s durational measures for consonants. Specifically, while a first set of measurements makes it appear that in Chichewa as elsewhere root-initial consonants are significantly longer than root-internal, measures taken comparing situations in which the root-initial consonant belongs to either the penult or the antepenult show that in Chichewa consonant strengthening takes place not in all root-initial syllables, but rather in the onset of the penult, where the vowel also undergoes significant lengthening. When in the antepenult, the root-initial consonant is significantly shorter (Hubbard 1994: 148). Now if strengthening of the penult in the relevant languages, an innovation in the Bantu context, is in fact the reanalysis of an earlier prominence on stem-initial syllables (an attested process in the histories of some stress languages, e.g. West Slavic), as is suggested by the transfer of consonantal strengthening from the initial to the penult, we could hypothesize that the new vocalic prominence on the penult may have had a correlate on initial-syllable vowels as well.

_____________________________________
this is the case in Turkish, and also for Korean (Sun-Ah Jun, p.c.), and is probably explicable along with other effects increasing the “consonantality” of initial sonorants.
Hyman (1999) proposes an analysis of the origins of Vowel Height Harmony in Bantu which adds some credence to this hypothesis. It has traditionally been thought that, for example, in the verbal system, harmonizing suffixes originally contained high vowels, but that when those suffixes were added to mid vowel roots, they eventually came to assimilate to the root in height, themselves becoming mid, and producing root-to-suffix height harmony. The issues are complex and cannot receive justice in the context of this study, but the essence of Hyman's proposal is this: There are compelling reasons to believe that the suffixes in question may actually have originated with mid vowels. What followed then was a process of raising of mid vowels in non-initial syllables. Such a process, of course, bears a striking resemblance some systems of UVR discussed in chapter 2. Even more striking is what would then give rise to the roots of the harmonies to develop later: mid vowels would fail to raise (reduce) only in case the root vowel was also mid. This support provided by the stem-initial syllable mid vowels for their non-initial counterparts is clearly reminiscent of those similar processes discussed above in the UVR systems of Yakan, Timugon Murut, and Russian dialects. A phonetic durational asymmetry between vowels in stem-initial syllables and other vowels provides an obvious motivation for such an account of the rise of Bantu VHH. Further evidence comes from a group of Bantu languages, such as Punu, which today lack VHH. In Punu mid vowels are excluded from non-stem-initial syllables altogether (Kwenzi Mickala 346)
In these syllables only /i, u, a/ of the five-vowel inventory may be realized. Crucially, in Punu non-stem-initial syllables /a/ is realized as schwa, and phonetic process which can only bespeak a concomitant reduction of phonetic duration.

Punu, then, can be seen as the natural endpoint of the chain of developments of which VHH in various incarnations represent middle stages, a process of the imposition of gradually stricter limitations on the realization of non-high vowels outside stem-initial syllables. In the ancestor to phonologized systems of VHH (which of course develop subsequently according to a logic constrained not by phonetic limitations), mid vowels can occur in non-initial syllables only when supported by a mid vowel in the initial. Subsequent weakening then leads to a situation such as that found in Punu, where mid vowels are banned altogether from non-prominent positions, much in the manner of a UVR system (though of course, without assuming an initial stress per se in the Bantu cases).

Another consideration in the Benue-Congo cases of initial-syllable strength effects is the morphological status of the initial syllable. In the cases of the Bantoid Tiv (Pulleyblank 1988) and the Ogoni Gokana (Hyman 1982) cited by Steriade 1994, the initial syllable which determines the specifications of [high], [low], and [round] (Tiv) and [nasal] (Gokana) of the vowels of following syllables is the stem-initial syllable of the verb. This syllable is also historically the root in both cases, with following syllables
being suffixes etymologically at least, though probably not analyzably so in all cases in synchrony (Hyman, p.c.). It is not clear whether it is the psycholinguistic status of the root which allows it to license more contrasts than do affixes, or if this too can be ascribed to phonetics (through the weakening of segmental material in affixes as they become increasingly bound prosodically to roots or stems). In this case it may well be a combination of both, or the phonetic effect may be a direct consequence of the psycholinguistic facts. Whatever the reason, it quite possible that the status of the initial syllable in these cases as the etymological verb root has more to do with the development of the relevant licensing asymmetries than does the initialness of syllables in question per se.

4.6. Summary

In this chapter I have presented evidence from a broad typological survey that phonological strength patterns involving the vowels of initial syllables are substantially rarer than previous treatments of the topic suggest. The vast majority of cases, in fact, can be attributed to a present or past fixed initial stress in the languages in question. Furthermore, while in this group of languages patterns both of vowel harmony and surface inventory reduction as in canonical UVR patterns are attested, in the small set of
languages for which initial stress cannot be documented or reconstructed, only harmony seems to be found.

I have argued that these typological facts are readily accounted for by the phonologization approach, in that they follow directly from the phonetic profile of initial syllables crosslinguistically. Robust operation of a process of phonetic initial strengthening is well-attested on the initial segments within domains, but unlike final lengthening, its effects in most cases which have so far been investigated do not extent further into the domain than that initial syllable. Thus, while PN effects involving initial consonants and absolute word-initial vowels are both quite common crosslinguistically, PN patterns affecting all initial-syllable vowels at not at all common.

Toward an explanation of why initial-syllable strength effects should appear at all the, I have demonstrated on the basis of experimental work with Turkish that in a language with the right type of prosodic system some phonetic initial strengthening can have a measurable effect on the duration of all initial syllable vowels. Furthermore, I have presented a scenario explaining how the small-scale durational asymmetry this causes between initial and non-initial syllables can lead to the development of a system of vowel harmony such as those attested in the initial-syllable strength languages left over when the initial-stress group has been removed from consideration. The fact that such patterns are as rare as they seem to be follow from several factors about the
developmental process I propose: first, only some subset of the world's languages (of unclear proportions) have prosodic systems (with non-duration-cued stress) which might allow the expression of postboundary strengthening on the vowels of initial syllables. Second, in addition to the sound change phonologizing patterns of vowel-to-vowel coarticulation which is the first step in my account, the development of a harmony system of the kind found in the Turkic languages relied crucially on subsequent morphological changes rooted in paradigm uniformity to extend harmonization across-the-board. Third, these two steps are likely to take place creating word-bounded vowel harmony only in languages of a particular morphological profile; for the process to work, the language must have almost exclusively mono- or disyllabic roots and robust (if not exclusive) transparently agglutinative suffixing morphology. The need for the convergence of all these factors for a true initial-syllable strength effect to arise for vowels explains both why we find such systems in Turkic and Niger-Congo, and why they are so rare otherwise.

The UG-based approach to PN in initial syllables, relying as it does on psycholinguistic properties of initial material arising as a consequence of the very fact of that material’s initialness, makes none of the typological predictions recorded above. In fact, it makes concrete predictions which are seen not to be born out in attested typological patterns. Were it the case that licensing asymmetries spring from the fact that
it is equally important to the grammar for processing reasons to conserve all segmental
contrasts for every segment in the initial syllable, we would expect equal attestation of
positional strength effects involving consonants and vowels. This is not the case.
Furthermore, if the need to preserve vowel contrasts in initial syllables were truly
enshrined universally in phonological grammar, we might expect, for example, the
existence of systems of unstressed vowel reduction in which all unstressed vowels save
that of the initial syllable undergo neutralizations of contrasts, while the vowel of the
initial syllable, because of its psycholinguistic importance, resists the reduction process.
We know there are systems in which absolute initial vowels resist reduction for phonetic
reasons (Russian). We have seen numerous instances of phonetic final lengthening giving
rise to patterns of final vowel resistance to reduction, both phonetic and phonological.
Crosswhite (2001) provides an analysis of several systems in which the vowels of certain
affixes fail to undergo reduction, presumably due to the need to preserve morphological
contrasts. But I know of no system in which vowels of all initial syllables, onset or no,
are exempt from a process of unstressed vowel reduction that otherwise targets all
unstressed vowels. This gap in the typology is expected given the predictions made above
as to which languages should exhibit initial-syllable vowel strengthening and which
should not. It was predicted that languages like English, in which duration is cued
strongly or primarily by stress, should not implement phonetic initial-syllable vowel
strengthening. Since, however, it is precisely these languages in which we expect to find systems of unstressed vowel reduction (and not in those, like Turkish, with small or nonexistent durational differences between stressed and unstressed syllables), the fact that the two phenomena should not be found to cooccur is unsurprising. In sum, the phonologization approach to initial-syllable strength effects provides an account which is far more accurate than that provided by other approaches, and which allows us to understand both the differences and similarities between PN effects found in initial syllables and those affecting the other positions surveyed in this study.
Chapter 5. Conclusions

5.1. Phonologization and Universal Grammar in phonological typology

I have argued throughout this study that typological regularities observed in patterns of positional neutralization are best accounted for using the phonologization approach presented here: Phonetics influences phonology by providing the inputs out of which categorical patterns are created by phonologization. Once phonologized, however, these patterns cease to be dependent on phonetic factors for their existence. Another class of approaches to typological patterns involving the phonetic content of positional neutralization systems seeks to derive the range of attested patterns from restrictions on the possible forms of such patterns either stated directly in Universal Grammar or arising fortuitously from the interaction of various independent statements within Universal Grammar. In this type of approach the phonology itself restricts the contents of the patterns it implements, making sure they are “grounded” or “natural” from the point of view of phonetics. In the phonologization approach, phonology is natural because the phonetic inputs from which most phonological regularities are derived are natural. The phonetics, in other words, determines the shape of future phonological patterns according to its own logic, one not involving abstract categories and structural positions. This view of phonologization presented here has its roots in the listener-oriented approach to sound change developed by Ohala (1981, 1993a), and Hyman’s (1977) work on phonologization. Recent work in a similar vein includes Blevins and Garrett (1998), Garrett and Blevins (to appear), and the work of Blevins (in prep.) under the rubric of Evolutionary Phonology. Chapter 1 of this dissertation sketches how the interaction of
phonetics and phonology through phonologization takes place, while Chapters 2 through 4 demonstrate the operation of phonetic principles in phonologization.

Chapter 2 demonstrated that the vast majority of unstressed vowel reduction patterns are based on the neutralization of vowel height contrasts due to perceptual difficulties caused by the shrinking of the vowel space produced by duration-dependent undershoot. This very specific and highly consistent, indeed nearly exceptionless pattern makes perfect sense in the phonologization model of typological patterning. UG-based approaches such as that of Crosswhite (2001), on the hand, predict a wide variety of completely unattested patterns, while failing to provide sufficient rationale for the restricted nature of the attested patterns.

Chapter Three dealt with the behavior of the final syllable in phonology, showing at once a widespread tendency for final syllables to resist various neutralization processes which might otherwise apply to them, but also in some languages themselves to be the target of neutralization processes not affecting other prosodically and segmentally comparable targets. The ambiguous nature of the final syllable, I argued, is due to the complexity of the phonetic patterns characteristic of vowels in this position, and thus perfectly understandable in the phonologization approach. Final lengthening and supralaryngeal gestural strengthening make final syllables phonetically strong; low subglottal pressure, however, together with an attendant drop in F0, lower intensity, and partial or complete devoicing or the onset of non-modal phonation can also negatively impact the perceptual robustness of contrasts realized in this position. Which tendency ultimately wins out seems to be a matter decided by the particular configuration of the language-specific phonetics of a given system. Furthermore, even when one of these
phonetic tendencies is phonologized, however, creating either phonological weakness or strength, this by no means entails the effacement of the opposing phonetic tendencies, since these two systems function largely independently. From a UG-based perspective, on the other hand, there is no particular reason to expect such waffling from the final syllables of the languages of the world. If phonological prominence is in fact a matter of Universal Grammar specifying, for whatever reason, whether a given position is to be strong or weak as a licensor of contrasts, the best we can do let final position belong to both lists. The precise nature of the patterns arising in cases of either behavior, however, remain mysterious.

Chapter Four contrasted the phonologization approach with a UG-based approach deriving PN in initial syllables from psychological characteristics connected with the initialness of the syllable in question. Contrary to what a UG approach would predict, positional neutralization patterns affecting all vowels in initial syllables were shown to be few and far between in the languages of the world (in contrast to the well-attested patterns involving initial consonants, which the phonologization model derives from the well-attested phonetic strengthening of precisely these segments, as well as word-initial vowels). In fact, most of the putative instances of PN involving the vowels of initial syllables take place in languages which either have fixed initial stress today, or had fixed initial stress at the time of the development of the patterns in question. There are a few other cases, such as Gujarati, in which the placement of stress is less clear, but in which the patterns of PN are clearly the result of phonologizations affected by durational asymmetries. Remaining cases of initial-syllable vowel PN, in which there can be no question of initial stress at any stage (such as the Benue-Congo examples and potentially
also Turkic, though the accentual facts are more ambiguous there) are remarkable in that all show systems of vowel harmony (or derivations thereof) rather than inventory reductions of the UVR-type. This regularity among the cases left over when initial-stress languages are removed from the picture is unexpected in the initialness model, but I present evidence that in certain languages small-scale durational asymmetries may arise through the application of initial strengthening to the vowels of initial syllables. I also present a scenario for how such an asymmetry could lead to the development of vowel harmony, as in Proto-Turkic. UVR-like patterns would be unexpected, on the other hand, as the durational asymmetries in question would be too small to produce the articulatory pressures which lead to patterns of this sort.

In each case - stressed, final, and initial syllable positions - the phonologization model was shown to account better for the full range of typological regularities in PN systems than competing approaches. I have argued that since attested typologies fall out naturally from the phonologization of phonetic patterns in diachrony, there is little reason to imagine that the phonology needs to cover this same ground again, "justifying" within the competence of the individual speaker the functionalist wholesomeness of phonological patterns which are already in essence a fait accompli for the language. If it is necessary at all that the phonology place restrictions on the phonetic content of the patterns it implements, it should be possible to show that there is some aspect of these patterns which cannot be accounted for through phonologization. We should not simply assume a priori that the phonological grammar contains such information because it is possible to model it in such a way. It must be clear additionally that it is necessary for us to do so.
The central problem here is the aprioristic assumption that the phonological grammar must be capable both of implementing the phonological patterns of any given human language and accounting restrictively and exhaustively for the range of variation in those patterns crosslinguistically. This assumption (though prevalent in generative phonology, and especially Optimality Theory) is simply unwarranted. It burdens the phonological grammar unnecessarily with the machinery needed to derive phonetically-grounded typological regularities, while failing to account for those regularities as accurately as an approach appealing to the phonetics itself.

There is still, of course, a long way to go to the complete resolution of these issues. In addition to the search for evidence of phonetic content in phonology in the behavior of individual speakers, more typological investigation may serve to further elucidate the issues at hand. Smith (2002) treats licensing asymmetries between roots and affixes, arguing that these are a result of the psycholinguistic status of the former. Here I think this argument is quite plausible. A competing hypothesis, though, might seek to derive such asymmetries from processes of phonetic weakening visited on affixal material as it suffers prosodic demotion to form a single prosodic word with the root. Typological surveys may serve to disambiguate further. Other potential loci for the phonologization of positional neutralization patterns, such as closed and open syllables, could also provide further empirical tests for the phonologization approach.

---

135 Here again it is conceivable that phonetic weakening is the result of the psycholinguistic status of the entity in question. Still, it is the phonetic patterns which would be responsible for the shape of the resulting patterns of PN. The role of psycholinguistic factors is indirect.
The success of the phonologization model in using phonetics to account for typological regularities in phonological patterns suggests the possibility that a synchronic model of phonetics-phonology incorporating phonetic information directly into the phonological grammar might be equally successful. In the next section I will discuss this option briefly, arguing that there are problems with this approach that the phonologization model avoids.

5.2. On the integration of phonetics and phonology

A very different approach to modeling the relationship between the phonological patterns of positional neutralization presented in this study and their phonetic antecedents is to allow the phonology direct reference to phonetic information, creating in effect an integrated model of phonetics-phonology. Such models include Steriade’s Licensing-by-Cue approach (Steriade 1994, 1997) and Flemming’s integrated model of phonetics and phonology (passim) introduced in Chapter 1. A model which derives phonological licensing patterns directly in synchrony from the phonetic patterns which have engendered them diachronically would succeed, as Steriade has argued, in accounting for the typological patterning of neutralization processes with far greater accuracy than a phonetics-free approach to PN such as that of Beckman (1998). Direct derivation of categorical neutralization patterns from phonetic cues or pressures, however, makes strong predictions characterizing the relationship between phonetic sources and phonological consequences as far more intimate a connection than in fact it is. The split model of phonetics-phonology, together with the phonologization approach to phonological typology, allows the same precise accounting for crosslinguistic
regularities, without forcing those patterns in synchrony to be something which they are not.

I will take categorical unstressed vowel reduction as a test case, as a treatment of it has already been worked out in a direct-phonetics model of the phonetics-phonology interface (Flemming 2001). An approach deriving UVR from phonetic factors would need to focus, as described in Chapter 2, on the relationship between the time needed to achieve the relatively more open target articulations for the upper vocal tract characteristic of non-high vowels and the duration allotted in UVR languages to vowels in unstressed syllables. Phonetic UVR such as, say, raising of /a/ toward [ə] or mid vowels /e, o/ toward [i, u] will can be modeled as the interaction of durational targets for vowels in unstressed syllables and constraints mandating minimization of articulatory effect where possible. Shortening the duration of an unstressed low vowel would increase the amount of articulatory effort necessary to reach the low vowel’s articulatory targets in the time allotted, since without altering either the duration of the vowel or the gestural targets the only fix is to increase the speed of the gestures in question. Assuming lower ranking of constraints specifying gestural targets, however, and higher ranking of both the constraints fixing syllable durations and those forbidding expenditure of the amount of effort necessary to reach a low vowel target in an unstressed syllable, the correct solution is to undershoot the vowel’s openness targets (raising the non-high vowels phonetically), such that gestural maxima are reached in a timely fashion with an acceptable expenditure of articulatory effort.

This describes gradient vowel raising in UVR systems such as that found for unstressed /o/ ([o]–[u]) in Bulgarian or /a/ ([a]–[ə]) in Russian. Formalized correctly, the
system would generate F1 values for the vowels in question as a function of the durations allotted to those vowels in a given unstressed syllable. Where unstressed syllables were shorter (fast speech, immediately posttonic non-final syllables) more reduction would take place, and where durations were longer (absolute word-initial or absolute phrase-final position) less reduction would take place.

To produce neutralization of contrasts, Flemming’s model (1999, 2001) uses the interaction of constraints mandating the maintenance of a certain number of contrasts and constraints demanding that there be a minimum phonetic distance between entities in contrast. Put simply, if constraints mandating that the system retain the contrasts between /o/ and /u/ or /e/ and /i/ outrank constraints specifying the minimum phonetic distance along the height dimension that [o] must be from [u] or [e] must be from [i], then no neutralization will take place, and the contrast will be maintained even in such case as the mid vowels are raised perilously close to the high vowels. If, however, Minimum Distance constraints outrank the relevant Maintain Contrasts constraints, and durational targets and constraints against articulatory effort are demanding realizations of [e] and [i] or [o] and [u] closer to one another than Minimum Distance would allow, the only viable solution is the neutralization of contrasts. By merging /e/ - /i/ as /i/ and /o/ - /u/ as /u/, we will violate the Maintain Contrasts constraint, but Minimum Distance will be satisfied, as will the duration and effort constraints. This approach, in essence, sees neutralization as the consequence of two contrasting entities being forced so close together in phonetic space by realizational constraints that in the end the contrast must be abandoned rather than realized in so perceptually non-robust a fashion.
Now here the phonologized versions of UVR processes present a fairly obvious problem common to all direct-phonetics models, but one that is by no means insurmountable. The problem generally is in the variability of vowel durations. If it is necessary to specify a particular durational minimum below which a contrast is best left undeployed, there will always be some subset of all the tokens of a target string which for whatever reason (segmental context, speech rate, phrasal position, etc.) exceed that minimum, and also some subset of all the tokens of non-target (e.g. stressed) strings which fall below it. Turning specifically to Flemming’s model, what the scenario sketched above predicts is a distribution for the realization of mid vowels in which, as duration decreases, the vowels are raised proportionally along with it, until a certain point (specified by Minimum Distance constraints), after which there is a break in the continuum following which neutralization must occur, and for all durations below a certain boundary mid vowels will simply be realized as high vowels. We picture an /e/ which becomes gradually more and more [i]-like down to X ms., after which realizations jump from [i] minus Y Hz at X ms straight to [i] at X + 1 ms. But no such systems seem to exist. Rather, systems are either neutralizing or non-neutralizing, phonologized or unphonologized, as detailed in Chapter 2. Russian /o/ does not become gradually more [a]-like in unstressed syllables until it reaches the threshold of contrastibility and leaps to become identical to /a/. It is simply realized in the same distribution as /a/ in all unstressed syllables regardless of duration. I argued in Chapter 2 that this is because after phonologization a neutralization process is nothing more than the substitution of one abstract symbol for another in a given structurally-definable position, completely
divorced from whatever phonetic trends or pressures may ultimately have given rise to
the phonologization in question.

What all this means is that when we speak of direct reference to phonetics in the
phonology, we cannot mean this literally at the level of the individual token in speech
production. Some relativizing or generalizing process must be able to intervene, such that
the phonetic representations and values the grammar evaluates are actually more like
statistical generalizations over distributions of concrete tokens, rather than values for the
individual tokens themselves. This result of this generalization (I will call it for simplicity
of exposition the mean realization, though in practice it might not be formalized as such)
allows the grammar to judge that, for example, while a given posttonic /e/ in Brazilian
Portuguese in phrase-final position might exceed even the stressed vowel in duration,
nonetheless for a speaker for whom posttonic raising is phonologized, the fact that the
mean realization calculated over all tokens of posttonic /e/ is below a certain threshold
means that neutralization must apply, final lengthening or no.

Still though, the relation between phonetic duration and neutralization assumed
here is too constraining. For even if generalizing over tokens to find a short enough mean
allows UVR to apply despite individual tokens with durations well sufficient to
countenance realization of whatever contrasts may be necessary, it still must be the case
that a high enough proportion of stored tokens of a vowel in a given position be short
enough the mean realization in fact comes out below the threshold. But this may not
always be the case. Take for example the Russian first pre-tonic syllable described in
Chapter 2. Here /a/ and /о/ are realized as something like [a] or [e], depending on the
dialect, the speaker, the speech style, and so forth. In the speech of many Muscovites, the
vowel in question is famously a robust [a], with a long duration frequently exceedingly
even that of the stressed syllable. I have no statistical analyses demonstrating the mean
duration for /a/-/o/ in this position, but it will certainly be longer than the 90 ms given by
Matusevich (1976) for this vowel in Standard Russian. We might well wonder at this
point just how long a vowel this would need to be to realize the contrast between /a/ and
/ö/ robustly. It is by no means obvious that even as mean duration of 90 ms is terribly
impoverished in this respect. Additionally, given that a clear [a] is realized here
consistently, there can be no question of duration pressures causing undershoot of any
kind, so that it not clear why any minimum distance constraints would be violated in the
first place. The problem is that whatever phonetic circumstances gave rise to this
particular neutralization were relevant and operational only hundreds of years ago, when
the neutralization was phonologized. Synchronically, the neutralization simply takes
place when the relevant segments occur in the specified structural positions, regardless of
the phonetic characteristics of vowels pronounced there in synchrony.

136 Other dialect speakers often find this feature of Muscovite speech grating, and mock it as a “drawled”
pronunciation.
137 It might be possible in the Russian case to derive this result by assuming that the mean realization values
for the duration of unstressed /a/ were calculated over all unstressed syllables, rather than just over first
pretonics. Given that first pretonic and non-first pretonic syllables must have separate durational targets
specified though (the former being substantially longer than the latter), it is not clear why this would be so.
138 Recall that it is still not clear whether the neutralization of /a/ and /ö/ in unstressed syllables was actually
a merger historically, or if the two (derived originally form long and short /a/ respectively) rather simply
failed to split everywhere but the stressed syllable. Either way the conditions in question are not relevant to
the synchronic implementation of UVR in Russian.
Consider also in this connection the case of Seediq. In Chapter Two I presented an analysis of Seediq UVR, where in posttonic syllable unstressed /e/, /o/, /u/ merge as /u/. I demonstrated that the source of this pattern is that fact that Seediq stressed /e/ developed from an earlier schwa, such that in stressed syllables schwa become /e/, while in unstressed it merged with /u/ (whatever the details of this latter merger may be). To model this correspondence in terms of phonetic pressures would simply misrepresent the situation, since there is clearly no sense in which synchronically Seediq /e/ is being reduced in unstressed syllables ever closer to [u] in unstressed syllables until finally some threshold for minimum distance is crossed and the contrast must be abandoned. Indeed, this never took place even diachronically, as it turns out. Yet the fact that Seediq UVR cannot be modeled in this way implies that synchronically the correspondence between Seediq stressed /e/ and posttonic unstressed [u] is of a completely different order than the relationship between Russian stressed /o/ and unstressed [a] or Brazilian stressed /o/ and posttonic [u]. The latter two cases would be instances of real UVR, while the former would have to receive some completely different formal implementation, making it an instance of some other, different process altogether. This may have been the case in diachrony, but in synchrony I see no reason to treat any of these cases as anything other than the phonetically arbitrary substitution of one abstract symbol for another in the same
structural position. All three cases are thus equally examples of a type of categorical positional neutralization known as unstressed vowel reduction.

The all-phonetic approach makes diachronic predictions as well which seem not to be supported by the evidence. As discussed above, the application of UVR in unstressed syllables in a given language implies that at some level of statistical generalization the durational value for unstressed vowels is such that they are forced by constraint ranking either to violate Minimum Distance constraints or to submit to neutralizations. Again, this means that in spite of the phonologization of the pattern of neutralization, the connection of the process to phonetic reality is alive and functional. With this in mind, consider a scenario in which the language to underwent some change to its prosodic system such that the durational asymmetry between stressed and unstressed syllables were to even out (e.g., if the language were to switch from strongly-duration-cued to non-duration-cued or less-duration-cued stress)\textsuperscript{139}. We would now predict that once the mean realization of the durations of unstressed vowels came to exceed whatever threshold would allow Minimum Distance constraints to be satisfied properly again, the contrasts between unstressed vowels which were previously neutralized should emerge again unscathed, as if nothing ever happened. In other words,

\textsuperscript{139} As in fact I argue to have taken place in Timugon Murut in chapter three, where in TM pretonic syllables /a/ is not realized as schwa, while in related neighboring languages it is (and apparently was in TM at an earlier period as well).
were Russian suddenly to adopt the durational targets for unstressed syllables of Polish, 

*akanje* should disappear without a trace, unstressed */o/* restored to its fully contrastive 
realization. I know of no such attested development.

5.3. Does “anything go” in phonology?

What I am arguing here, then, is that while categorical patterns of positional 
neutralization are best understood in their typological dimension through the 
phonologization approach applied throughout this study, in synchrony the best approach 
to the implementation of phonologized PN is one that accomplishes the substitution of 
one abstract category for another without reference to or concern for phonetic reality. 
Phonetic factors in synchrony could of course influence patterns of PN by leading to 
further phonologizations, perhaps disrupting patterns in some way. It is not, however, 
necessary for the phonetics to continue to provide formal justification in synchrony for 
the existence of the patterns it creates.

This view prompts an obvious question: since phonology contains no restrictions 
on the phonetic content of the patterns it implements (whether these be present up front in 
the constraint system or set back in grounding conditions), is it the case that as far as the 
phonology is concerned, anything goes? The fact that the vast majority of phonological 
patterns appear to be phonetically “natural”, or “grounded” or “motivated” seems to 
argue otherwise. As I have argued throughout this study, however, this on its own means 
nothing. In fact, this would be the case with or without phonetic restrictions operating in
the phonological grammar. Phonological patterns appear to be “natural” because the primary source of phonological patterns in diachrony literally is natural. Phonological patterns such as those treated in this study arise through the phonologization of phonetic patterns, which are natural by definition. Barring any further developments then, the overwhelming majority phonological pattern will obey phonetic “restrictions” as an automatic consequence of their diachronic origin. Patterns of morphological change, subsequent sound changes, or even instances such as Seediq UVR in which two unrelated sound changes apply to the same segment in different environments can ultimately disrupt the uniformity of the naturalness pattern by creating correspondences in synchronic PN systems which no longer operate in accord with their original phonetic logic. But such cases are the exception rather than the norm, involving as they do phonetically motivated sound changes plus some number of additional complicating developments.

It is not therefore necessary to assume that the categorical phonology restricts the phonetic content of the patterns it implements in order to account for the fact that the vast majority of sound patterns appear to be phonetically grounded or natural. Such an assumption, in fact, makes the undesirable prediction that the phonology should contain only sound patterns which are phonetically grounded, or at the very least that processes such as those described in the preceding paragraph should be strongly selected against by the phonology. We might expect, for example, that phonetically unmotivated alternations, where they succeeded in arising at all, would be pressured out of the phonology over time. This, however, seems not to be the case. Garrett and Blevins (to appear) in fact demonstrate in several languages the free analogical extension of phonetically
unmotivated alternations such that they come actually to play a greater role in the grammars of the languages in question rather than a lesser one. Similar patterns are reported in Anderson (1981) and Bach and Harms (1972). In addition to research into morphophonological change, investigation of the behavior of natural and unnatural sound patterns in child language acquisition might also yield important evidence bearing on these questions. Is it the case, for example, that all of the systematic phonological patterns displayed by children in the early stages of acquisition are phonetically natural?

Phonology no doubt places its own restrictions on the implementation of sound patterns, involving perhaps concerns of predictability and symmetry rather than phonetic naturalness. It is the task of the phonologist to determine what these restrictions are and how they function to mold the inputs provided to the phonologization process by the phonetics. It is by concentrating on concerns such as these that we will ultimately locate the contribution the phonology itself makes to the shapes of the processes it implements.
References


371

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


378

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Doctoral Dissertation, University of California, Berkeley.


391


Pandit, P. B. 1955. e and o in Gujarati. *Indian Linguistics* 15: 15-44.


Penny, Ralph J. 1969. Vowel harmony in the speech of the Montes de Pas (Santander). *Orbis* 18: 148-166


401

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


